

DISS. ETH NO. 24498

**Carbohydrate-based vaccines against
the sheep parasite *Haemonchus contortus***

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

SUSANNA MARIA FLEURKENS

MSc. in Biology, ETH Zurich, Switzerland

born on 22.10.1983

citizen of the Netherlands and Spain

accepted on the recommendation of

Prof. Dr. Markus Aebi
Prof. Dr. Annette Oxenius
Prof. Dr. Peter Geldhof

2017

Acknowledgements

I gratefully thank Prof. Dr. Markus Aebi for giving me the opportunity to work in his group and for his continuous support, guidance and advice during all stages of this thesis. His door was always open to discuss new ideas or challenging problems. His passion and enthusiasm for science has been very inspiring.

I am also truly thankful to my PhD committee, Prof. Dr. Annette Oxenius, Prof. Dr. Peter Geldhof and Prof. Dr. Markus Künzler for the fruitful discussions and stimulating ideas during our yearly meetings.

Likewise, my sincere gratitude goes to Dr. Christine Neupert who has been a great mentor. Her fresh ideas and input, as well as the scientific discussions and helpful advice she gave me during these past years were invaluable.

The work presented here would not have been possible without many collaborations. I thank Dr. Patrick Brunner for his support and contribution to the polymorphism project and patience to answer my many questions. I am also very grateful for the many constructive discussions and the helpful suggestions to Dr. Bruno Oesch, PD Dr. Irene Schiller, Dr. Lino Camponovo and Dr. Karin Ilg from Malcisbo AG, as well as to PD Dr. med. vet. Hubertus Hertzberg and Prof. Dr. Peter Deplazes from the Parasitology Institute of the University of Zürich.

I also owe my deepest gratitude to Jasmin Frey for her hard and excellent work during her semester thesis. It was a pleasure to work with her during her first scientific project and sharing moments of passion and dedication.

Big thanks go to the Aebi group for the pleasant and very collaborative atmosphere in the lab. Thank you Chia-wei, Robert, Marie-Estelle and Tim for creating an interesting, diverse and very open scientific environment. In particular, I want to thank Francesca for all the help, the countless inspiring discussions and advice I got from her during these years. There was always time for a chat and good laugh. I would also like to thank Ilaria for her continuous support and encouragement, especially during the last months when I needed it the most. Many thanks also go to my friends Jillianne, Corina, Corinna, Ilaria and Francesca, with whom I spend the most joyful non-scientific moments in these four years.

I thank the management and the staff of the Institute of Microbiology. Especially, I want to thank Susanne Surber for her help with non-scientific problems and many funny coffee breaks.

Very special thanks goes to my boyfriend David for his precious support, patience and encouragement not only during my PhD, but every day.

Above all, I want to thank my friends and family, especially my parents Julia and Hans and my sister Carmen for being such a great loving and supporting family. This thesis is dedicated to my parents.

Table of contents

Summary	i
Zusammenfassung.....	iii
Chapter 1.....	1
Introduction	
Chapter 2.....	37
A glycoengineering approach for production of <i>Haemonchus contortus</i> vaccine candidates in insect cells	
Chapter 3.....	73
The extensive surface variability of H11 aminopeptidase glycoproteins from <i>Haemonchus contortus</i>	
Chapter 4.....	121
Analysis of site-specific <i>N</i> -glycan remodelling in the ER and the Golgi	
Chapter 5.....	171
Concluding remarks and future perspectives	
Curriculum vitae	179

Summary

Helminths are complex multicellular organisms that infect billions of humans and livestock worldwide causing an immense economic impact. The globally most important parasite of small ruminants is the blood-feeding parasitic nematode *Haemonchus contortus*. This parasite is currently controlled by anthelmintic drugs, but its increasing resistance indicates the need for new strategies. The development of an effective vaccine against this parasite would be the best alternative. Native vaccines have proven to be protective, but every attempt of producing recombinant versions has failed so far.

Chapter 1 summarizes the current status of vaccine development and provides insights into recent progress in the components affecting the future progression of producing a potent recombinant vaccine against this parasite. A special focus is placed on the unusual and structurally distinct *N*-glycans of parasitic nematodes as compared to the glycan structures of their host.

In Chapter 2, we evaluate the potential of a recombinant glycoprotein-based vaccine for sheep against *H. contortus*. We produced various vaccine candidates in glycoengineered insect cells and confirmed by HPLC and MS analysis the presence of nematode-like *N*-glycans on these proteins. Furthermore, comparison of native and recombinantly produced proteins revealed glycosite-specific microheterogeneity. Unfortunately, no protection was observed in a sheep vaccination trial, despite a significant immune response stimulated in immunized animals. We postulated that the lack of other proteins is mainly responsible for the unprotective recombinant vaccines.

Our studies of the protectivity of the novel produced antigens led to new insights on one of the most promising vaccine candidates. We identified a high rate of polymorphism specifically in the aminopeptidase family H11 of *H. contortus*. Chapter 3 focuses on the further analysis of this newly discovered heterogeneity on both protein and *N*-glycans. A massive surface variability of H11 was detected due to amino acid alterations on the protein surface, different positioning of *N*-glycans and a site-specific variability in the *N*-glycan structures. Furthermore, using maximum-likelihood model

positive diversifying selection was acting on the H11 genes could be shown. Which selection pressures might be responsible for this high variability was further discussed and a different view on the requirements for vaccine-induced protection is presented. Our findings suggest that high genetic variations in *H. contortus* populations are one of the main reasons for failures in the recombinant vaccine development.

In Chapter 4, we examined in more detail how *N*-glycan microheterogeneity is generated in the Golgi apparatus of insect cells and what are the 'driving forces' of this diversity. The model protein, the yeast protein disulfide isomerase 1 (PDI1) was therefore produced in the insect cell expression system. *In vivo* and *in vitro* studies of *N*-glycan processing events showed that the underlying protein sequence had a defining role in the *N*-glycan processing. The protein conformation determined the final glycan structure by interacting with the covalently linked *N*-glycan.

The final Chapter 5 summarizes the main findings and conclusions of the thesis and provides potential strategies for future research.

Zusammenfassung

Helminthen sind komplexe, multizelluläre Organismen, welche auf der ganzen Welt Milliarden von Menschen und Vieh infizieren und damit einen immensen ökonomischen Schaden verursachen. Der weltweit häufigste Parasit in kleinen Wiederkäuern ist der blutsaugende Nematodenwurm *Haemonchus contortus*. Dieser Parasit wird derzeit mit Anthelmintika kontrolliert, jedoch zeigt seine zunehmende Resistenz dagegen die Notwendigkeit zur Erarbeitung neuer Strategien auf. Die Entwicklung eines wirksamen Impfstoffs gegen diesen Parasiten wäre die bestmögliche Alternative. Impfstoffe die aus nativem Wurmmaterial extrahiert werden, haben sich zwar als schützend erwiesen jedoch sind jegliche Versuche diese Proteine rekombinant herzustellen bisher gescheitert.

Kapitel 1 fasst den aktuellen Stand der Impfstoffentwicklung zusammen und gibt Einblicke in die jüngsten Fortschritte im Zusammenhang mit den Aspekten, welche die zukünftige Entwicklung eines potenten rekombinanten Impfstoffs gegen diesen Parasiten beeinflussen könnten. Ein besonderer Schwerpunkt dieses Kapitels liegt auf den *N*-Glykanen der parasitären Nematoden, welche im Vergleich zu den Glykanstrukturen ihres Wirtes strukturell unterschiedlich und sehr ungewöhnlich sind.

In Kapitel 2 untersuchten wir das Potential eines rekombinanten glykoprotein-basierten Impfstoffs gegen den Parasiten *H. contortus*, der zum Schutz von Schafen eingesetzt werden soll. Dabei wurde ein neuartiges Expressionssystem mit glykanoptimierten Insektenzellen entwickelt um Impfstoffantigene mit nematoden-spezifischen *N*-Glykanen herzustellen, welche denen des Pathogens sehr gleichen. Mit dieser Methode haben wir verschiedene Darmproteasen des Parasiten rekombinant produziert. Hochleistungsflüssigkeitschromatographie und zwei massenspektrometrie-basierte Analysen verifizierten die nematoden-spezifischen *N*-Glykanstrukturen auf den rekombinant produzierten Antigenen und zeigten eine unterschiedliche Prozessierung dieser Strukturen (Mikroheterogenität), welche abhängig von der Lage der *N*-Glykanbindung sind. Obwohl eine signifikante Immunantwort bei geimpften Tieren festgestellt wurde, konnte gleichwohl kein Schutz gegen den Parasiten erzielt

werden. Unserer Meinung nach war das Fehlen anderer relevanter Proteine hauptsächlich dafür verantwortlich, dass die rekombinanten Impfstoffe nicht ihre erhoffte Wirkung erzielten.

Unsere Untersuchungen über die Wirksamkeit der rekombinanten Antigene führten gleichzeitig zu wichtigen, neuen Erkenntnissen im Zusammenhang mit einem dieser aussichtsreichen Impfstoffkandidaten. Wir entdeckten insbesondere in der Aminopeptidase Familie H11 von *H. contortus* eine hohe Polymorphismusrate. In Kapitel 3 konzentrierten wir uns auf die weitere Analyse dieser neu entdeckten Variabilität. Diese erfolgte sowohl auf Protein- als auch auf *N*-Glykanebene. Dabei wurde insbesondere aufgrund der Aminosäureveränderungen auf der Oberfläche, der unterschiedlichen Positionierungen der *N*-Glykane und einer ortsspezifischen Variabilität in den *N*-Glykanstrukturen, eine massive Oberflächenvariabilität der H11-Proteine nachgewiesen. Durch die Verwendung von Maximal-Wahrscheinlichkeits-Modellen entdeckten wir, dass ein positiver Selektionsdruck auf die H11 Gene wirkt. Des Weiteren werden die Selektionsdrücke, welche für diese hohe Variabilität verantwortlich sein könnten, diskutiert. Aufgrund der präsentierten Resultate wird eine andere Auffassung über Anforderungen an den impfstoff-induzierten Schutz vorgestellt. Unsere Ergebnisse deuten darauf hin, dass hohe genetische Variationen in *H. contortus* Populationen einen der Hauptgründe für das Ausbleiben einer erfolgreichen rekombinanten Impfstoffproduktion darstellt.

In Kapitel 4 untersuchten wir genauer, wie die *N*-Glykan-Mikroheterogenität im Golgi Kompartiment der Insektenzellen erzeugt wird und was die verantwortlichen Komponenten dieser Glykanvielfalt sind. Hierzu wurde das Modellprotein, die Hefe Proteindisulfid-Isomerase 1 (PDI1), im Insektenzellexpressionssystem hergestellt. *In vivo* und *in vitro* Studien von *N*-Glykanprozessierungen zeigten, dass die zugrundeliegende Proteinsequenz eine entscheidende Rolle bei der *N*-Glykanprozessierung einnimmt. Die Proteinkonformation bestimmte dabei die endgültige Struktur des Glykans durch ihre Wechselwirkungen mit dem kovalent gebundenen Glykan.

Das abschliessende Kapitel 5 fasst die wichtigsten Ergebnisse und Schlussfolgerungen der vorliegenden Dissertation zusammen und zeigt mögliche Strategien für künftige Forschungsarbeiten auf diesem Gebiet auf.

Chapter 1

Introduction

1 Introduction

More than 1.5 billion of the world's poorest people are infected or at risk for infections with parasitic helminths [1]. This astronomical number means that 24% of the current human population is affected. Moreover, parasitic diseases are highly prevalent in livestock causing an immense economic impact worldwide.

Helminths are complex multicellular organisms and compared to viruses and bacteria relatively large animals, which cannot simply be removed by phagocytosis. Therefore, many parasitic worms have the capability to survive for decades in the host, also through their ability to suppress the host's immune system. Commonly, these parasitic infections are clinically asymptomatic and do not cause any harm to the host. Sometimes, they are even beneficial for the host by reducing inflammatory diseases. However, in some cases, parasitism can cause severe immunopathological complications in infected individuals, such as granulomatous disease and organ failure [2].

Despite many years of research and knowledge of the immune response being capable of expelling parasites upon vaccination in model systems, no vaccines are currently available for human use [3]. Only a handful of recombinant vaccines against animal parasites are available to date, including vaccines against cestode infections in livestock and against hookworm in dog. More often, native antigens have proven to be protective, while recombinant versions failed to induce the same levels of protection. This applies also to the globally important parasite of small ruminants, the blood-feeding nematode *Haemonchus contortus*. The intent of this chapter is to summarize the current status of vaccine development and recent progress in each of the components affecting the future progression of producing a potent recombinant vaccine against *H. contortus*.

1.1 Introduction to *Haemonchus contortus*

Haemonchus contortus (barber's pole worm) is a gastrointestinal, blood-feeding parasitic nematode infecting millions of domestic and wildlife ruminants in almost all regions of the globe. This parasite belongs to the superfamily Trichostrongyloidea in Clade V in the phylum Nematoda [4]. The roundworm Nematoda and the flatworm Platyhelminthes (Trematodes and Cestodes) together comprise the Helminths. The main clinical signs of haemonchosis are anemia (due to average blood losses of 49 ml per day [5] in hosts infected with 1000 worms), oedema, lethargy, decreased live-weight gain, impaired wool/milk production and decreased reproductive performance [6]. Often the infections lead to death in young animals and in severely affected animals with heavy worm burdens [7].

From a global perspective, *H. contortus* is the most important blood-feeding parasite in sheep and goats. It causes a high economic impact worldwide with production losses estimated at tens of billions of dollars per annum [8]. Two other gastrointestinal nematodes of economic importance in sheep are *Teladorsagia circumcincta* and *Trichostrongylus* spp [7].

H. contortus is sexually dioecious with obligate sexual reproduction [9] and has a direct and rapid life cycle (Figure 1) [10]. Adult worms reside and mate in the host abomasum, the fourth and final stomach compartment in ruminants, where each female worm produces up to 2000 eggs per day [11], which are excreted with the feces. The eggs embryonate, develop as first stage larvae (L1) and hatch in the grass. The larvae molt, within approximately one week, through the second stage larvae (L2) to become third stage larvae (L3), which migrate onto pasture. Infective L3 larvae are ingested by the grazing ruminant host and pass through the forestomachs to the abomasum, where they exsheath and develop through the fourth larval stage (L4) to become adults in two to three weeks [10]. The last two stages both feed on blood. Due to the rapid life cycle and high infection rates in a flock, *H. contortus* has an extremely high fecundity. In a flock with hundreds of sheep, each harboring hundreds of egg-laying females, billions of new progeny are passed onto pasture every few days [11].

Currently, the disease is treated with broad-spectrum chemotherapeutics or anthelmintic drugs, such as benzimidazoles, levamisole or ivermectin [12]. However, *H. contortus* resistance to all major anthelmintic drug classes and the rapid spread of drug-resistant strains indicates the need for new strategies [7, 13]. Therefore, numerous attempts to either breed genetically resistant sheep or to develop a safe and effective vaccine against *H. contortus* have been made over the last decades.

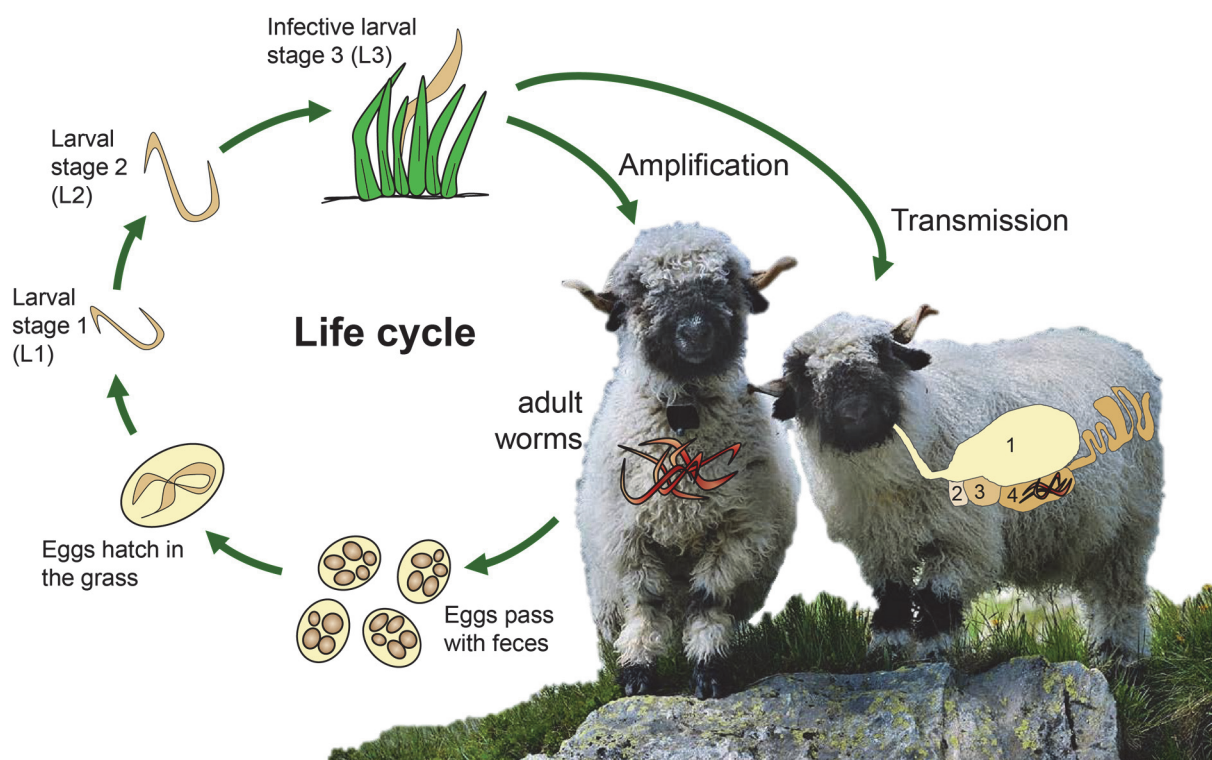


Figure 1: Life cycle of *Haemonchus contortus*

Hundreds of adult worms sexually reproduce in the ruminant abomasum (the fourth stomach) and each female thereby produces 2000 eggs per day. Eggs are excreted with the feces, then they embryonate, develop as first stage larvae (L1) and hatch in the grass. The larvae molt within approximately one week through the second stage larvae (L2) to become third stage infective larvae (L3), which migrate onto pasture. Grazing hosts ingest the larvae reinfesting themselves (amplification) or other grazing ruminants (transmission). Ingested L3 larvae travel through the four chambered stomach (1. Rumen, 2. Reticulum, 3. Omasum, 4. Abomasum). After arrival in the abomasum, they exsheath and develop through the fourth larval stage (L4) to become adults in two to three weeks. The last two stages both feed on blood. Figure adapted from [14, 15].

1.2 Genetic Diversity of *Haemonchus contortus*

The need to combat parasitic diseases has been the primary stimulus for the majority of the parasitological research, but some aspects of parasite biology, e.g. genetics, were neglected for many years [16]. Genetic variation in *H. contortus* was evident before studying the underlying genetic basis due to several features of the parasite. For example, the adaptation of *H. contortus* to a broad range of different host species, including domestic and wild animals, and the adaptation to different climatic zones, spreading from its origin in sub-Saharan Africa to almost all regions of the globe [17], suggests underlying genetic diversity. In addition, the different vulva morph types of *H. contortus*, smooth, knobbed, linguiform A-right, linguiform A-left or linguiform C-type, are another example of genetically determined variation [16, 18]. Furthermore, a well-known hallmark of this parasite is its remarkably high potential of developing anthelmintic resistance, even after only a few years of drug treatment [11, 19]. This adaptive capacity is largely due to a high genetic variation in the parasite population [19].

Many studies on Trichostrongyloidea nematodes, including *H. contortus*, showed that this superfamily is characterized by high genetic variability and large effective population sizes [20, 21]. A first study indicated extremely high levels of within-population genetic diversity (96 – 99 %) in five species of this superfamily from three different hosts (*Ostertagia ostertagi* and *H. placei* from cattle, *H. contortus* and *Teladorsagia circumcincta* from sheep, and *Mazamastrongylus odocoilei* from white-tailed deer) [21]. In addition, *H. contortus* was shown to have high genetic diversity not only within, but as well between populations and within and among continents [22]. Moreover, these high levels of genetic diversity are retained in laboratory isolates passaged for many years [12, 23, 24]. These reports were mainly analyzing the nicotine amide dehydrogenase (NAD4) locus of the mitochondrial DNA, which has a higher rate of substitution than in nuclear DNA, making it possible to differentiate the most closely related species pairs [25].

Other studies confirmed the high genetic diversity in *H. contortus* populations by analysis of isozymes [26], analysis of mRNA [27] and analysis of DNA by amplified fragment length polymorphism (AFLP) [12, 22], microsatellites [23, 28, 29] and transposon associated markers [30]. These high levels of sequence polymorphism in

addition to the relatively large genome size (~300 Mb) caused a major challenge to the draft genome assembly of *H. contortus* [8, 31].

Sequence polymorphisms have generally been ascribed to result from high mutation rates, large effective population sizes and migration rates [19]. *H. contortus* has indeed a mutation rate that is up to ten times higher than the one of vertebrates [21]. However, the high levels of genetic diversity in *H. contortus* are predominantly due to its extremely large effective population sizes [32]. In fact, billions of new genotypes are passed onto pasture every few days in a flock with hundreds of sheep [11]. Assuming that *H. contortus* has a similar mutation rate as *C. elegans* (2.1×10^{-8} mutations per site per generation [33]), enough progeny is created to mutate every single nucleotide position in the genome each day [11]. Consequently, this parasite has a remarkable ability to generate new mutations that provides a high adaptive capacity to respond to selective pressures, such as environmental factors, host immune responses and chemical treatments [32]. During the parasite's life cycle its different stages are subject to distinct selection pressures. While in the free-living stages, temperature and humidity will mainly affect population size, host immune responses and drug treatments are applying strong selection pressures to the parasitic stages [9].

Nucleotide diversity might lead to non-neutral modifications on the protein level, either through non-synonymous mutations (amino acid change), frameshift mutations or stop codons leading to non-functional proteins. A number of studies have analyzed the diversity of nematode proteins and especially the variability of vaccine candidates [34], as this could have considerable implications for the production of a globally-effective protein vaccine. Among the vaccine candidates studied there are the gut-expressed cysteine proteases of *H. contortus*, which are predicted to be encoded by 63 genes [31]. In this protein group an extreme variability was reported by using ESTs (expressed sequence tags) or PCR-SSCP (Single Strand Conformation Polymorphism) for analysis [27, 35, 36]. Substantial functional diversity was suggested for enzymatic properties of the highly variable cysteine proteases [37]. However, they were shown to hydrolyze similar di- and tetra-peptide substrates under neutral and acidic pH conditions [38]. In contrast, a relatively low level of diversity was found in the hookworm vaccine antigen, the *Ancylostoma* secreted protein termed ASP-1 (10/424 amino acid variation) [39].

In summary, many studies have shown a high genetic diversity in Trichostrongyloidea nematodes and there is some evidence of variation in vaccine candidates, but more research is needed to understand the underlying mechanisms and potential consequences for future vaccine development.

1.3 *N*-glycans of *Haemonchus contortus*

1.3.1 Introduction to *N*-glycosylation

Asparagine-linked glycosylation (*N*-glycosylation) is one of the most abundant post-translational modifications on proteins in Eukaryotes, Bacteria and Archaea [40]. It is important for protein folding, oligomerization, quality control, sorting and transport of secretory and membrane proteins. The *N*-glycans not only promote protein folding directly by stabilizing polypeptides, but also indirectly by serving as recognition “tags” in the interaction with a variety of lectins (carbohydrate-binding proteins), glycosidases, and glycosyltransferases [41, 42]

During the synthesis of *N*-glycans in the Endoplasmic Reticulum, a preassembled core oligosaccharide unit (Glc₃Man₉GlcNAc₂, where Glc is glucose, Man is mannose and GlcNAc is *N*-acetylglucosamine) is transferred from the dolicholpyrophosphate lipid carrier to an asparagine residues of a growing, nascent polypeptide chain by the oligosaccharyltransferase (Figure 2) [43]. The acceptor sequon for *N*-linked glycosylation in eukaryotes is N-X-S/T (whereas X can be any amino acid except for proline) [42].

The early stages of the classic *N*-glycan processing pathway in the endoplasmic reticulum show an extreme degree of conservation across all domains of life [44]. However, some parasitic protists can transfer truncated forms of lipid-linked oligosaccharides, sometimes shorter “high mannose” or even just the core GlcNAc₂ sequence [45].

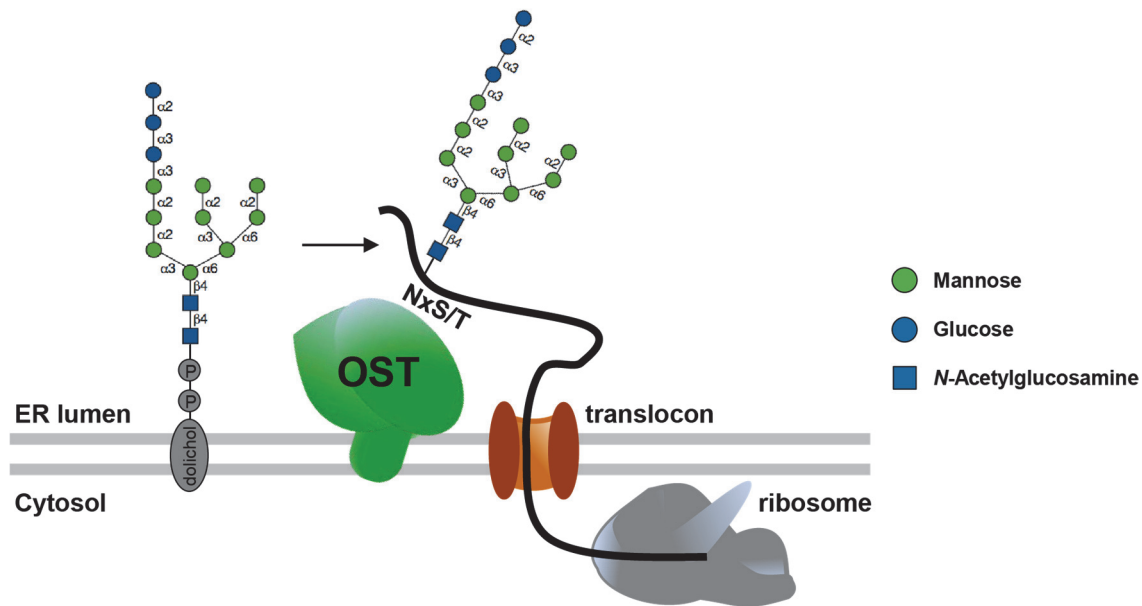


Figure 2: N-linked glycosylation in the endoplasmic reticulum (ER)

The preassembled oligosaccharide is transferred from the dolicholpyrophosphate lipid carrier to asparagine residues of a growing, nascent polypeptide by the OST. Figure modified from [46].

Once the glycoprotein has folded and reached the Golgi complex, the glycan chains undergo further trimming or extension, and acquire more complex structures through a series of nonuniform modifications carried out by the Golgi processing machineries [41]. As these modifications vary from species to species, the resulting final *N*-glycan structures on secreted proteins are highly different among distinct eukaryotic taxa (Figure 3) [44]. Yeasts do not appear to trim their mannose residues, but often they further extend them into large high-mannose glycans (mannans). In insects, mannoses are trimmed like in mammals, leading to simple final structures with only the three core mannose residues (paucimannose structures) and the addition of core difucosylation. Plants follow a similar pathway but then often add a bisecting β 1-2-linked xylose residue on the β -linked mannose residue. This glycan epitope, as well as the α 1-3-linked core fucose, are also present in some parasitic helminths (e.g. *Schistosoma mansoni*) and are immunogenic in vertebrates. In contrast, vertebrates further decorate their glycans by producing multi-antennary complex type glycans with terminal galactose and sialic acid residues [44]. As a result, this heterogeneity of the final *N*-glycan structures adds an additional level of information content to the underlying polypeptide structure [47]. Finally, *N*-glycans contribute to numerous biological functions, in particular at the cell surface, where they do not only have many protective,

stabilizing, organizational, and barrier functions, but also influence the interactions with the extracellular environment by providing ligands for cell adhesion, macromolecule interactions and pathogen invasion [47, 48].

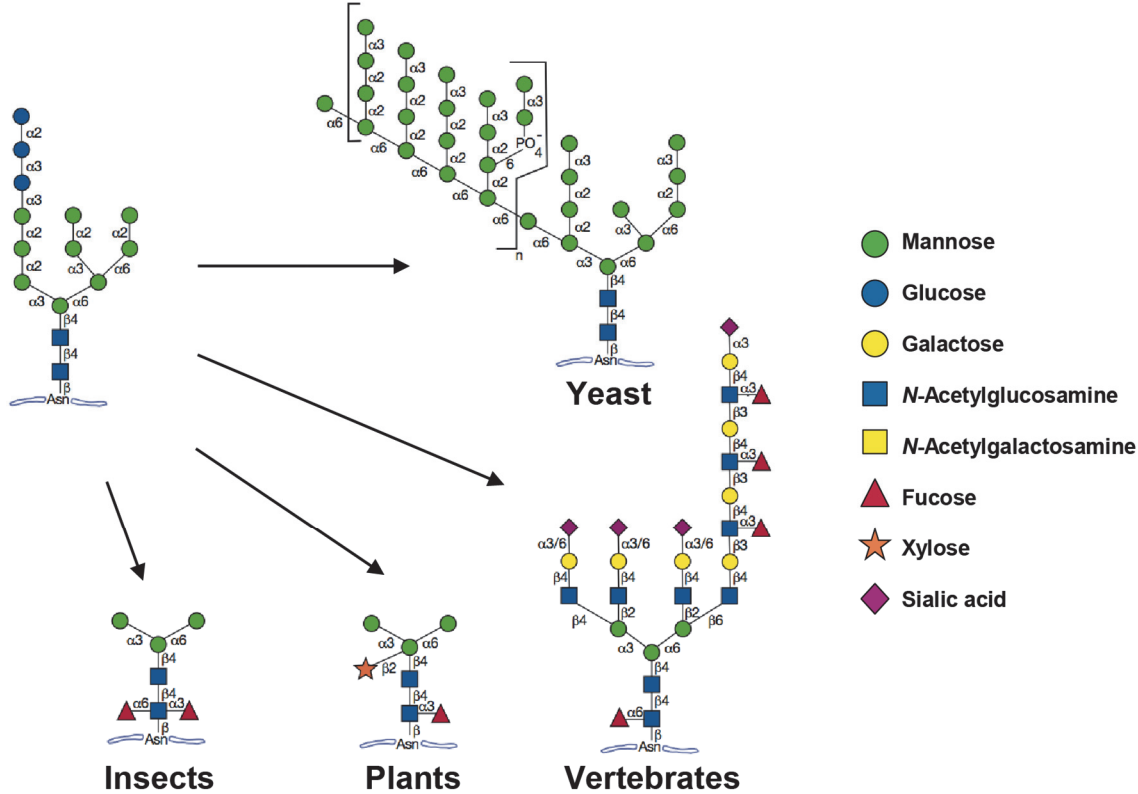


Figure 3: N-glycan modifications in the Golgi

Heterogeneity of final N-glycan structures among different eukaryotic taxa. Typical N-glycans are shown produced by yeast (mannan), insects (paucimannose), plants, and vertebrates (multi-antennary complex type glycans). Figure modified from [44].

1.3.1 *N*-glycans of *Haemonchus contortus*

Many parasitic glycans are unique to helminths or to a particular worm species, they are highly abundant on the worm surfaces or secretions and most of them are highly immunogenic to vertebrates [49, 50].

In *Haemonchus contortus*, an early study from the 1990s revealed unusual carbohydrate modifications with up to three core fucose residues by FAB-mass spectrometry on the native H11 antigen [51]. The three fucoses are attached to the chitobiose core, two of them are found to be α 1,3- and α 1,6-linked to the proximal GlcNAc and one fucose is α 1,3-linked to the distal GlcNAc (Figure 4). The latter substitution is unique in *N*-glycans and stage-specifically expressed in adult worms [52]. The α 1,3-fucosyltransferase transferring the third fucose to the distal GlcNAc has been identified in *C. elegans* [53]. The core α 1,3 fucose epitope is a common antigen synthesized by many parasitic helminths and is highly immunogenic to vertebrates when present on plant or insect glycoproteins [54, 55], whereas the core α 1,6 fucose epitope is commonly found on vertebrate *N*-glycans (Figure 3).

In nematodes, α 1,3 and α 1,6 core fucoses are often additionally decorated with galactose residues, creating another nematode-antigen, the 'GalFuc' epitope [56, 57]. The enzyme GALT-1, a core α 1,6-fucoside β 1,4-galactosyltransferase-1 is generating the GalFuc epitope in *C. elegans* [58]. This glycosyltransferase was identified with a forward genetics-based target-search screening of *C. elegans* strains resistant towards the nematotoxic fungal galectin CGL2 (*Coprinosia* galectin 2) [59], which binds to the 'GalFuc' epitope. The CGL2 lectin, as well as CCL2 (*Coprinosia cinerea* lectin 2 [60]), which recognizes the α 1,3 Fuc epitope by binding GlcNAc- β 1,4-(Fuc- α 1,3-)GlcNAc, were shown to be toxic towards *H. contortus* in an *in vitro* larval development assay and bind the digestive tract of the larvae, as well as the brush border of the adult parasite gut [61]. In addition, the lectins AAL (Fucose [62, 63]) and MOA (Gal α 1,3Gal/GalNAc [64, 65]) were toxic against *H. contortus* in the *in vitro* larval development assay, whereas CGL3 (GlcNAc β 1-4GlcNAc and GalNAc β 1-4GlcNAc [66]) and Tec2 (2-O-Me-Fucose and 3-O-Me-Mannose [67]) were not. This study thereby delivered valuable indirect evidence on the potential absence or presence of *N*-glycan epitopes in *H. contortus*.

The immunogenicity of the glycans and the sensitivity of nematodes towards specific lectins (carbohydrate-binding proteins) lead to the hypothesis that lectins or anti-glycan antibodies might be effective anthelmintics [59, 61]. Moreover, there is evidence that a significant part of the sheep immune response targets carbohydrate epitopes of the parasite [68]. For example, IgG antibodies from lambs vaccinated with a native vaccine (H-gal-GP complex) recognize the LDNF epitope (fucosylated GalNAc β 1-4GlcNAc) [69, 70] and IgE antibodies from *H. contortus* infected sheep bind the core α 1,3 fucose and may contribute to an induction of a Th2 response [55].

The identification of a thrombospondin-like glycoprotein (GP300) in *Dictyocaulus viviparus*, lead to the analysis of GP300 orthologues in *H. contortus* and revealed that these proteins contained phosphorylcholine-substituted *N*-glycans, as was shown by immunoblotting with the PC-specific antibody TEPC-15 [71].

A recent study reappraised the *H. contortus* *N*-glycome by using HPLC and MALDI-TOF MS/MS in combination with selected digestions of *N*-glycans [56]. Many previously predicted or identified *N*-glycan modifications were verified in the study, trifucosylated glycans, LDNF, GalFuc epitope and phosphorylcholine [56]. In contrast to earlier studies the trifucosylated glycans were not found to be attached to trimannosyl cores, but to *N*-glycans with two mannoses attached to the core and where the α 1,6 mannose is missing. This finding was consistent with the *in vitro* requirements of the α 1,3-fucosyltransferase FUT-6 [53]. The additional hexose was found to be a galactose, which is α 1,2-linked to the distal α 1,3-fucose [56]. Therefore this study identified a new unusual *N*-glycan modification in *H. contortus*.

1.3.2 *N*-glycans of other parasitic nematodes

Parasitic nematodes are characterized by the production of glycoproteins containing many unusual and structurally distinct *N*-glycans as compared to the *N*-glycan structures of their host (Figure 3 and 4). Typically, *N*-glycans of vertebrates terminate in sialic acid, whereas nematodes do not synthesize this monosaccharide. Parasitic nematodes commonly decorate their structures with terminal β -linked GalNAc on a GlcNAc, leading to the LacdiNAc (LDN) motif [72]. This epitope is found on *N*-glycans in the sheep nematode *H. contortus*, in the cattle lungworm *Dictyocaulus viviparus* [73] and in the wide-spread nematode *Trichinella spiralis* [74] that infects rodents, pigs, horses, bears, and humans. The LDN motif can further be decorated on the GalNAc by the unusual tyvelose sugar, found in *N*-glycans of *T. spiralis* [75, 76], or by fucose residues on both GlcNAc and GalNAc residues in the LDN motif, giving rise to FLDN or LDNF, found in *H. contortus* [69] and *Trichuris suis* [77].

Another unusual modification of glycans, such as the phosphorylcholine (PC), is a more common feature and is generated in several parasitic nematodes on the terminal GlcNAc in: *Ascaris suum* [78], *H. contortus* [56], *Onchocerca volvulus* [79], *Oesophagostomum dentatum* [57] and *T. spiralis* [74]. There is indirect evidence by TEPC-15 immunoblotting that also *D. viviparus* proteins contain phosphorylcholine-substituted *N*-glycans [71]. In addition, PC is also detected on GalNAc in LDN and on GlcNAc in LDNF on *N*-glycans of *T. suis* [77].

Some glycoforms or epitopes are not uniquely found on *N*-glycans of parasitic nematodes, like the Lewis X antigen, the core α 1,3-fucose and high mannose glycans. The Lewis X antigen is commonly expressed on human milk oligosaccharides [80], but very limited in helminths, it can only be found in the nematode *D. viviparus* [73] and in the trematode *S. mansoni* [81]. The core α 1,3-fucose epitope is present on plant or insect glycoproteins and a highly immunogenic motif [54]. It can be found in *H. contortus* [51], *T. suis* [77] and many other parasitic helminths [55]. High mannose glycans (GlcNAc₂Man₅₋₉) are known to play a crucial intracellular role in the control of protein folding in the ER. They are commonly found in invertebrates and parasitic nematode glycomes, where they might (in part) derive from intracellular proteins, but they are also found on surface exposed and ES glycoproteins [81]. For example, on an ES protein of *Cooperia onchophora*, one of the most common parasitic nematodes in cattle, GlcNAc₂Man₆ was found to be the major glycan form [82].

Some nematodes modify their core fucose residues with galactose, either β 1,4-linked to the core α 1,6-fucose leading to the 'GalFuc' epitope or α 1,2-linked to the distal α 1,3-fucose. The latter has been recently discovered [57]. Both of these unusual modifications represent motifs particular to a subset of nematodes, occurring in the model nematode *C. elegans*, and in some parasitic worms including *H. contortus* [56], *A. suum* [57], and *O. dentatum* [57], but also in the free-living nematode *Pristionchus pacificus* [83]. The substituted or unsubstituted GalFuc epitope is not uniquely found in nematodes, but also in planaria and some molluscs [84].

Mass spectrometry based glycomics has provided insights in the unusual glycan repertoires for many but not for all parasitic nematodes. Future research will unravel the specific roles that these distinct *N*-glycans might play in the parasite's biology and during parasite-host interactions.

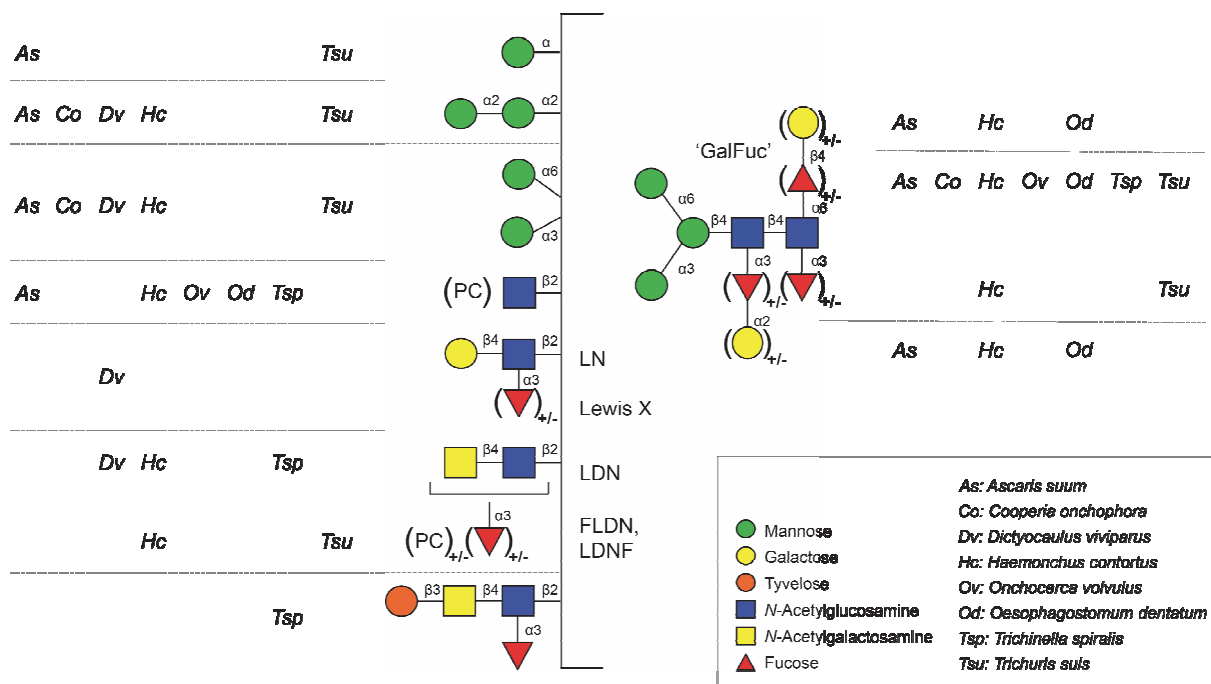


Figure 4: Overview of representative *N*-glycan structures and modifications of the parasitic nematodes where they were identified.

1.4 Vaccine Development against *Haemonchus contortus*

Anthelmintic drugs currently control *H. contortus* infections. However, there are *H. contortus* strains that are resistant to all major anthelmintic drug classes and they are rapidly spreading, which indicates the need for new strategies [7, 13]. The development of a safe and effective vaccine against this sheep parasite would be the best alternative to effectively control the disease and to combat the anthelmintic resistance problem. Over the last decades numerous protective native antigens have been identified but, so far, the remaining substantial challenge is to produce recombinant versions inducing the same levels of protection.

1.4.1 Native vaccines

There are two types of native vaccines inducing high levels of protection against *Haemonchus contortus*: 1) soluble excretory / secretory (E/S) products from the nematode and 2) gut-derived 'hidden' antigens from the blood-feeding adult nematode. Both vaccines have shown to reliably reduce fecal egg counts (FEC) by up to 95 % and worm burden by up to 82 % (reviewed by [7, 85-89]).

Immunizations with the first type of native vaccine, the adult somatic *H. contortus* extracts enriched for the 15 and 24 kDa E/S antigens, showed significant protection against challenge infection (reductions of 77% in FEC and 85% in worm burden) [90-92]. Vaccination induced primarily a Th2 response to the ES antigens with specific IgG1 antibodies and increased mastocytosis in the abomasal tissue of vaccinated sheep [92]. The authors suggested that vaccination may mimic the natural immunity mechanism [92], but the actual mechanism has not been defined.

The second type of native vaccine are the protective gut fractions, which derive from detergent-soluble extracts of the adult parasite and contain 'hidden' antigens. These are defined as antigens that do not induce an antibody response during the course of a natural infection [93], but can be used in vaccination strategies. However, there is evidence that 'hidden' antigens are not completely hidden for the immune system, as unvaccinated animals being exposed to *H. contortus* for a long time, are able to generate an immune response against the gut antigens [94, 95]. In addition, the immune response against H11 primed by immunization can be boosted, presumably by H11 being released from dead or dying parasites in the sheep stomach [93, 96].

Protective gut-antigens are expressed on the microvillar surface of the intestine and show protease activity, believed to be required for the degradation of host blood proteins like albumin and hemoglobin, the major food source of blood-feeding parasites [97]. Proteolytic digestion of hemoglobin is conducted in a semi-ordered cascade of proteases, which is notably similar in distantly related organisms such as hookworms, *Plasmodium* and *H. contortus* [89, 97, 98]. It has been postulated that hemoglobin is initially cleaved by aspartic proteases, then degraded to smaller peptides by cysteine proteases, followed by metalloproteases and terminally degraded into free amino acids by exopeptidases (e.g. aminopeptidases) [97].

To date, the most promising gut-derived native vaccines include the aminopeptidase H11 enriched in a Concanavalin A (ConA) binding fraction [99], the glycoprotein complex H-gal-GP containing metallo and aspartic proteases enriched in a galactose-binding fraction [100], and cysteine proteases enriched in a thiol-binding fraction [101]. The activity of the enzymes H11 and H-gal-GP is inhibited by IgG antibodies from vaccine-protected sheep [102, 103]. A limitation of 'hidden' antigen vaccines is that protective antibody levels wane rapidly, as there is little natural boosting, thus, readministration of vaccines need to be done regularly. A major advantage of the gut antigen approach is, that it is effective in situations where natural immunity is weak, such as in young lambs [104], and young goats [105] and peri-parturient ewes, where some protective immunity was shown to be transferred to lambs by maternal antibodies [106].

A commercial native anti-*H. contortus* vaccine (Barbervax®, WormVax) based on H11 and H-gal-GP was released in Australia in 2014, where the antigen preparation needs to be extracted from local adult worms derived from intentionally infected sheep. Barbervax has a restricted geographic coverage and a cold chain requirement limits its distribution, features that make it difficult to reach a global market [107].

Native H11

The best characterized native vaccine against *H. contortus* is the native H11 vaccine. The protective fraction derives from detergent-soluble extracts of the adult parasite and it is purified by lectin-affinity chromatography using Concanavalin A [99], which binds specifically α -D-mannosyl and α -D-glucosyl residues. This fraction results in a mixture of numerous different proteins with the H11 aminopeptidases being the most abundant [108]. H11 is an integral membrane glycoprotein that runs on SDS-PAGE as a doublet with a mean molecular mass of 110 kDa [99]. It is exclusively expressed on the intestinal microvillar surface of the parasitic stages and exhibits both aminopeptidase A and M activity [109] that is believed to degrade small peptides of host blood proteins into free amino acids [97].

Immunization with native H11 induces high levels of protection leading to a reduction of fecal egg counts by up to 95% and worm burden by 82% (reviewed in [7]). Correct conformation and enzymatic activity of the native aminopeptidase H11 are important to induce full protection, as the efficacy of the vaccine is reduced when H11 is denatured with SDS or SDS and DTT [102]. Protective IgG antibodies inhibit the activity of the H11 enzyme *in vitro*, and levels of inhibition highly correlate with levels of protection *in vivo* [102]. Protection is thought to be mediated by antibodies that, when taken up with the blood meal, inhibit the enzyme activity and disrupt nutrient absorption [109]. Therefore, the enzyme H11 is considered to be critical for parasite survival.

Recently, 13 aminopeptidases were discovered in the complement of the family M1 of *H. contortus* [110], and the native H11 vaccine comprises a family of five different isoforms (termed H11, H11-1, H11-2, H11-4 and H11-5) [85, 109, 111]. The five members of the family share 62-75% amino acid identity and are tandemly arranged in the genome, suggesting that the gene family has arisen through recent duplication and divergence [111]. Different transcriptional pattern of the isoforms were identified by semi-quantitative RT-PCR, where H11 and H11-1 were shown to be expressed in infective L3 larvae and all five isoforms in adult worms, with H11-4 being the most abundant form and H11-5 being female-specific [111]. H11 isoforms form dimers, but little is known about the interaction amongst its isoforms, for example whether they form heterodimers.

Each H11 isoform has the predicted structure of a type II membrane protein, with a short N-terminal cytoplasmic tail, a single transmembrane domain, and an extracellular region organized into the four-domain structure characteristic of the M1 metalloproteinases (Figure 5) [7, 112]. Domain II has a thermolysin-fold and possesses an internal cavity surrounding the active site with the zinc-binding HEXXH₁₈E and exo-peptidase GXMEN catalytic motifs [112]. All H11 proteins contain *N*-glycosylation sites and mass spectrometric analysis of the native H11 vaccine has identified unusual highly fucosylated core structures including core a1-3 fucosylation which is highly immunogenic [51].

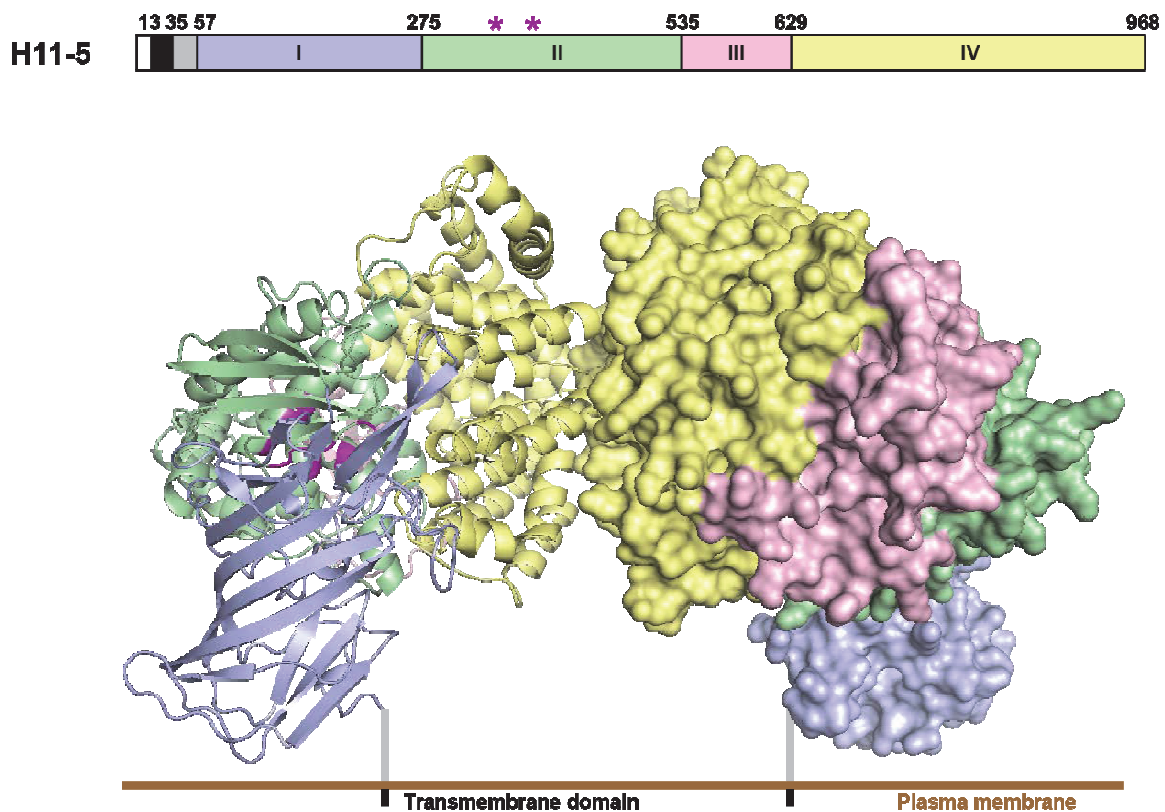


Figure 5: Modeled structure of the H11-5 isoform

Schematic representation of the domain organization of the H11-5 protein. The cytoplasmic tail is white, the transmembrane domain is black, the Ser/Thr rich stalk is gray and the remaining domains of the modeled protein structure are I: lightblue, II: palegreen, III: lightpink, IV: paleyellow. Purple stars show GXMEN and zinc-binding HEXXH₁₈E catalytic motifs. Below the scheme, the modeled H11-5 protein is shown as a dimer. Its protein structure was modeled by PHYRE2 [113] on the crystal structure of the human aminopeptidase N (hAPN [112]). The left monomer is shown in a cartoon and the right monomer in a surface representation. The dimer is depicted in its proposed orientation with respect to the plasma membrane and colored according to domains (as described above).

Native H-gal-GP

The glycoprotein complex H-gal-GP is enriched from detergent-soluble extracts of whole worms in a galactose-binding fraction by peanut agglutinin affinity chromatography, with specificity for Galactose- β 1,3-*N*-acetylgalactosamine disaccharide motifs [100]. This fraction was therefore termed *Haemonchus* galactose-containing glycoprotein complex (H-gal-GP) and contains a mixture of various proteins including a family of zinc metalloproteases (MEP1-4) and pepsinogen-like aspartic proteases (PEP1-2) [114-116]. In addition, multiple galectins, a cystatin and a thrombospondin are part of the 1 MDa complex and all are proposed to aid in the digestion of the blood meal [117-119]. The three-dimensional structure of the protective H-gal-GP complex, revealed by single particle/cryoelectron microscopy, depicts an arch conformation formed by two PEP1 proteins over a base laid by multiple MEP3 proteins [120].

H-gal-GP is a highly effective immunogen when used as a vaccine in immunization trials, leading to consistent reductions of 93 % in fecal egg counts and 72 % in worm burden [100]. IgG antibodies from vaccine-protected sheep inhibit the hemoglobinase activity of the protease complex *in vitro* [103] and IgG concentration correlates with the degree of protection *in vivo*, suggesting that protection is antibody-mediated [121]. Conformational epitopes of the glycoprotein complex are important to induce full protection, as the H-gal-GP loses its protective capacity when it is reduced by SDS and DTT, but not when it is dissociated by SDS alone [122].

Native cysteine proteases

Another group of gut-surface antigens with vaccine potential against *H. contortus* are the cathepsin B-like cysteine proteases (CBLs). Two distinct cysteine protease containing fractions have been evaluated in *H. contortus* immunization trials, a fibrinogen-degrading complex and the thiol sepharose-binding proteins (TSBP), both isolated from adult *H. contortus* worms.

The fibrinogen-degrading complex was protective in immunised lambs against challenge infection with a 93% reduction in fecal egg counts and a 87% reduction in worm burden [123]. TSBP proteins get their name from their enrichment by thiol-sepharose affinity chromatography, which is performed after the removal of H-gal-GP by peanut lectin affinity chromatography [101]. Three trials with lambs immunized with TSBP followed by a *H. contortus* challenge infection showed substantial reductions in faecal egg outputs by 77% and worm burdens by 47% compared to challenge controls [101].

CBLs have been implicated in the parasites blood meal digestion due to their ability to degrade fibrinogen and haemoglobin [124]. They constitute large multigene families [38] and their designation originates from their sequence similarity to mammalian cathepsin B [125]. Initially a small family with five proteins was described (AC-1 to AC-5), which share 64-77% amino acid identity and one of these proteins, AC-3, was found to be linked in tandem with AC-2 [126]. Additionally, genes encoding a polypeptide CBL (GCP) [127] and another cysteine family with six members (hmcp1-6) [128], were identified. Hmcp1, 4, and 6 are the protective targets in the TSBP vaccine [129]. CBL gene families have undergone a large expansion, probably through recent gene duplication [31], and represent the most abundant mRNAs (~16%) detected in the adult female intestine of *H. contortus* [27].

1.4.2 Recombinant vaccines

Despite the success of the native anti-*H. contortus* vaccines, issues with the use of native antigens include high production costs, quality control and the risk of contaminants being exported to other countries with the vaccine [88]. Therefore, the long term aim is to produce a defined, uniform, safe and effective recombinant vaccine using a cost-effective method. For this reason, there has been a continuous effort to produce recombinant versions of the components of the native antigens using various expression systems, like bacteria, yeast, insect cells or *Caenorhabditis elegans* (reviewed in [88]). However, all recombinant proteins so far, elicit only a partially protective immune response against *H. contortus* challenge.

Recombinant H11

Native H11 contains a mixture of various proteins, whereof the most abundant protein, the aminopeptidase H11, comprises a family of five different isoforms (H11, H11-1, H11-2, H11-4, H11-5). Each of these isoforms was recombinantly expressed in one of the multiple studies, evaluating distinct recombinant expression systems and vaccine efficacy, using single isoforms or combinations of multiple isoforms for vaccination.

Initially, an immunization trial with the active site domain of H11-1, expressed in inactive form in *E. coli* as inclusion bodies, reported partial protection with 30–40% reduction in fecal egg counts (Newton et al., unpublished; cited in [85]). Another vaccination trial by the same research group using combinations of baculovirus-insect cell expressed and enzymatically active isoforms for immunization, failed to offer any protection against challenge infection (Newton et al., unpublished; cited in [85]). Interestingly, a later study, showed partial protection in sheep immunized with baculovirus expressed, full-length H11-1 (30% reduction in worm burden) or GST-H11-1 fusion protein (20% reduction in worm burden) [130].

The lack of another protein present in the protective native H11 or inappropriate glycosylation of the baculovirus-expressed antigens were discussed to be responsible for the reduced protective capacity of the recombinant vaccines.

Therefore, the free-living nematode *Caenorhabditis elegans* was examined as an alternative recombinant expression system, as similar glycan modifications are found in *C. elegans* and in *H. contortus* [111]. Combinations of two H11 isoforms, H11-1 and

H11-4 together or H11-1 and H11-5 together, did not yield any protection in a vaccine / challenge trial, in spite of similar di- and tri-fucosylated glycan structure decoration on the *C. elegans*-expressed recombinant H11 compared to those on native H11. Some glycan structural differences were observed (e.g. lack of LDNF) compared to the native H11 vaccine, but the fucosylated *N*-glycans expressed on recombinant H11 were suggested not to contribute to the protective effect of H11 and that additional components present in the native H11 vaccine are required for inducing full protective immunity [111].

Recombinant versions of H-gal-GP

Multiple components of the H-gal-GP complex were identified by immunoscreening a *H. contortus* cDNA library with antisera from lambs successfully vaccinated with H-gal-GP [118, 119] and subsequently cloned and expressed recombinantly to test their vaccine efficacy in sheep immunization trials.

Initially, recombinant cystatin CYS-1 was bacterially expressed in a soluble and functionally active form, as shown by cysteine proteinase inhibition of both mammalian cathepsin B and native *H. contortus* cysteine proteases [119]. Vaccination with rCYS-1 did not confer any protection against infection with *H. contortus*, despite inducing high antigen-specific antibody levels [119].

Then, other predominant antigens of the complex were examined, namely the *H. contortus* zinc metalloproteases MEP1 and MEP3. No protection was observed, when sheep were immunized with bacterially expressed recombinant GST-fusion proteins of MEP1 [131]. Moreover, the principal domains of the MEP3 metalloprotease (41- and 47-kDa domains), did not confer any protection. Sheep immunized with recombinant MEP1 or MEP3 had significantly lower serum titers against the native H-gal-GP extract compared to animals immunized with the native extract [131], suggesting that conformational epitopes were lacking to induce full protectivity, due to incorrect folding or absence of glycosylation on the bacterially expressed molecules.

In another vaccine/challenge trial, the pepsinogen-like aspartic protease (PEP1), was bacterially expressed and refolded to a partially soluble and enzymatically active protein, but did not lead to any protection, although high native PEP1-binding IgG titers were induced [132].

Therefore an alternative, eukaryotic expression system based on insect cell cultures was evaluated for expression of recombinant vaccines. A cocktail vaccine of three insect cell-expressed metalloproteases (MEP1, MEP3 and MEP4) combined with refolded bacterially-expressed PEP1 failed to significantly reduce FEC or worm burden in a sheep vaccination trial, although high levels of H-gal-GP-recognizing serum antibodies were induced [133].

Recombinant cysteine proteases

Recombinant cysteine proteases hmcp1, 4, and 6 were bacterially expressed as inactive GST-fusion proteins and used to vaccinate lambs with a cocktail of the three recombinants prior to *H. contortus* challenge [129]. Worm burdens in vaccinated animals were reduced by 38% compared to controls, but the fecal egg counts were not [129]. The majority of the vaccination induced IgG response was directed to the GST component and the trial was repeated using non-fusion proteins with similar results (29% worm burden reduction) [134]. Unlike their native counterparts, the recombinant cysteine proteases were not enzymatically active, therefore conformational epitopes might play a role in protection. In addition, other components of the native TSBP vaccine and glycans are lacking on the bacterially expressed recombinant proteins.

1.4.3 Concluding remarks

Vaccination against the blood-sucking parasite *H. contortus* has shown to be successful when using native hidden gut-antigens extracted from adult worms. This approach has led to the commercialization of the first effective vaccine in the world against a nematode parasite of sheep [88]. The protective mechanism does not mirror natural immunity and requires high antigen-specific IgG levels that need to be restimulated by repeated immunizations. So far, every attempt to produce an effective recombinant vaccine against *H. contortus* failed to induce full protective immunity. The reduced protective capacity of the recombinant vaccines might be due to the absence of other proteins critical in conferring protection in the native vaccine, due to suboptimal or incorrect folding or lack of various post-translational modifications, such as *N*-glycosylation.

The plethora of studies disclosing only partially protective recombinant vaccines in sheep immunization trials over the last decades, demonstrates the huge challenge and complexity of developing a protective recombinant vaccine. Therefore it is important, that future research unravels the requirements for a correct epitope presentation, the underlying mechanisms of vaccine protection and the complex immune response conferring resistance.

References

1. WHO, *Soil-transmitted helminth infections, Fact sheet*. <http://www.who.int/mediacentre/factsheets/fs366/en/>, January 2017.
2. Maizels, R.M. and M. Yazdanbakhsh, *Immune regulation by helminth parasites: cellular and molecular mechanisms*. *Nat Rev Immunol*, 2003. **3**(9): p. 733-44.
3. Hewitson, J.P. and R.M. Maizels, *Vaccination against helminth parasite infections*. *Expert Rev Vaccines*, 2014. **13**(4): p. 473-87.
4. Blaxter, M.L., et al., *A molecular evolutionary framework for the phylum Nematoda*. *Nature*, 1998. **392**(6671): p. 71-5.
5. Clark, C.H., G.K. Kiesel, and C.H. Goby, *Measurements of blood loss caused by Haemonchus contortus infection in sheep*. *Am J Vet Res*, 1962. **23**: p. 977-80.
6. Gasser, R.B., et al., *Understanding Haemonchus contortus Better Through Genomics and Transcriptomics*. *Adv Parasitol*, 2016. **93**: p. 519-67.
7. Newton, S.E. and E.A. Munn, *The development of vaccines against gastrointestinal nematode parasites, particularly Haemonchus contortus*. *Parasitol Today*, 1999. **15**(3): p. 116-22.
8. Schwarz, E.M., et al., *The genome and developmental transcriptome of the strongylid nematode Haemonchus contortus*. *Genome Biol*, 2013. **14**(8): p. R89.
9. Gilleard, J.S. and E. Redman, *Genetic Diversity and Population Structure of Haemonchus contortus*. *Adv Parasitol*, 2016. **93**: p. 31-68.
10. Veglia, F., *Anatomy and Life-History of the Haemonchus Contortus (Rud.)*. 1915. 347-500 pp.
11. Gilleard, J.S., *Haemonchus contortus as a paradigm and model to study anthelmintic drug resistance*. *Parasitology*, 2013. **140**(12): p. 1506-22.
12. Otsen, M., et al., *Amplified fragment length polymorphism analysis of genetic diversity of Haemonchus contortus during selection for drug resistance*. *Int J Parasitol*, 2001. **31**(10): p. 1138-43.

13. Kotze, A.C. and R.K. Prichard, *Anthelmintic Resistance in Haemonchus contortus: History, Mechanisms and Diagnosis*. Adv Parasitol, 2016. **93**: p. 397-428.
14. <http://www.faecaleggcountkit.com.au/parasites/life-cycle-of-fasciola-hepatica-liverfluke/>.
15. <https://www.flickr.com/>.
16. Grant, W.N., *Genetic variation in parasitic nematodes and its implications*. Int J Parasitol, 1994. **24**(6): p. 821-30.
17. Hoberg, E.R., J.R. Lichtenfels, and L. Gibbons, *Phylogeny for species of Haemonchus (Nematoda: Trichostrongyloidea): considerations of their evolutionary history and global biogeography among Camelidae and Pecora (Artiodactyla)*. J Parasitol, 2004. **90**(5): p. 1085-102.
18. Le Jambre, L.F., *Genetics of vulvar morph types in Haemonchus contortus: Haemonchus contortus cayugensis from the Finger Lakes Region of New York*. Int J Parasitol, 1977. **7**(1): p. 9-14.
19. Prichard, R., *Genetic variability following selection of Haemonchus contortus with anthelmintics*. Trends Parasitol, 2001. **17**(9): p. 445-53.
20. Anderson, T.J., M.S. Blouin, and R.N. Beech, *Population biology of parasitic nematodes: applications of genetic markers*. Adv Parasitol, 1998. **41**: p. 219-83.
21. Blouin, M.S., et al., *Host movement and the genetic structure of populations of parasitic nematodes*. Genetics, 1995. **141**(3): p. 1007-14.
22. Troell, K., et al., *Global patterns reveal strong population structure in Haemonchus contortus, a nematode parasite of domesticated ruminants*. Int J Parasitol, 2006. **36**(12): p. 1305-16.
23. Redman, E., et al., *Microsatellite analysis reveals marked genetic differentiation between Haemonchus contortus laboratory isolates and provides a rapid system of genetic fingerprinting*. Int J Parasitol, 2008. **38**(1): p. 111-22.
24. Hunt, P.W., et al., *Genetic and phenotypic differences between isolates of Haemonchus contortus in Australia*. Int J Parasitol, 2008. **38**(8-9): p. 885-900.

25. Blouin, M.S., *Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer*. Int J Parasitol, 2002. **32**(5): p. 527-31.
26. Bentounsi, B. and J. Cabaret, *Analysis of helminth genetic data: comparative examples with Haemonchus contortus isozymes using exact tests or resampling procedures*. Parasitol Res, 1999. **85**(10): p. 855-7.
27. Jasmer, D.P., M.D. Mitreva, and J.P. McCarter, *mRNA sequences for Haemonchus contortus intestinal cathepsin B-like cysteine proteases display an extreme in abundance and diversity compared with other adult mammalian parasitic nematodes*. Mol Biochem Parasitol, 2004. **137**(2): p. 297-305.
28. Otsen, M., et al., *Microsatellite diversity of isolates of the parasitic nematode Haemonchus contortus*. Mol Biochem Parasitol, 2000. **110**(1): p. 69-77.
29. Hoekstra, R., et al., *Microsatellites of the parasitic nematode Haemonchus contortus: polymorphism and linkage with a direct repeat*. Mol Biochem Parasitol, 1997. **89**(1): p. 97-107.
30. Hoekstra, R., et al., *Transposon associated markers for the parasitic nematode Haemonchus contortus*. Mol Biochem Parasitol, 2000. **105**(1): p. 127-35.
31. Laing, R., et al., *The genome and transcriptome of Haemonchus contortus, a key model parasite for drug and vaccine discovery*. Genome Biol, 2013. **14**(8): p. R88.
32. Gilleard, J.S. and R.N. Beech, *Population genetics of anthelmintic resistance in parasitic nematodes*. Parasitology, 2007. **134**(Pt 8): p. 1133-47.
33. Denver, D.R., et al., *High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome*. Nature, 2004. **430**(7000): p. 679-82.
34. Maizels, R.M. and A. Kurniawan-Atmadja, *Variation and polymorphism in helminth parasites*. Parasitology, 2002. **125 Suppl**: p. S25-37.
35. Ruiz, A., et al., *Genetic variability in cysteine protease genes of Haemonchus contortus*. Parasitology, 2004. **128**(Pt 5): p. 549-59.
36. Martin, S., et al., *Influence of immunoprotection on genetic variability of cysteine proteinases from Haemonchus contortus adult worms*. Int J Parasitol, 2015. **45**(13): p. 831-40.

37. Rehman, A. and D.P. Jasmer, *Defined characteristics of cathepsin B-like proteins from nematodes: inferred functional diversity and phylogenetic relationships*. *Molecular and Biochemical Parasitology*, 1999. **102**(2): p. 297-310.
38. Shompole, S. and D.P. Jasmer, *Cathepsin B-like cysteine proteases confer intestinal cysteine protease activity in Haemonchus contortus*. *J Biol Chem*, 2001. **276**(4): p. 2928-34.
39. Qiang, S., et al., *Variation between ASP-1 molecules from Ancylostoma caninum in China and the United States*. *J Parasitol*, 2000. **86**(1): p. 181-5.
40. Dell, A., et al., *Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes*. *Int J Microbiol*, 2010. **2010**: p. 148178.
41. Helenius, A. and M. Aebi, *Intracellular functions of N-linked glycans*. *Science*, 2001. **291**(5512): p. 2364-9.
42. Helenius, A. and M. Aebi, *Roles of N-linked glycans in the endoplasmic reticulum*. *Annu Rev Biochem*, 2004. **73**: p. 1019-49.
43. Kornfeld, R. and S. Kornfeld, *Assembly of asparagine-linked oligosaccharides*. *Annu Rev Biochem*, 1985. **54**: p. 631-64.
44. Varki, A., H.H. Freeze, and P. Gagneux, *Evolution of Glycan Diversity*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009: Cold Spring Harbor (NY).
45. Aebi, M., *N-linked protein glycosylation in the ER*. *Biochim Biophys Acta*, 2013. **1833**(11): p. 2430-7.
46. Alberts, B., et al., *Molecular biology of the cell. 4th edn.* 2002.
47. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation: diversity, synthesis and function*. *Nat Rev Mol Cell Biol*, 2012. **13**(7): p. 448-62.
48. Varki, A. and J.B. Lowe, *Biological Roles of Glycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009: Cold Spring Harbor (NY).
49. Nyame, A.K., Z.S. Kwar, and R.D. Cummings, *Antigenic glycans in parasitic infections: implications for vaccines and diagnostics*. *Arch Biochem Biophys*, 2004. **426**(2): p. 182-200.

50. Prasanphanich, N.S., et al., *Glycoconjugates in host-helminth interactions*. Front Immunol, 2013. **4**: p. 240.
51. Haslam, S.M., et al., *Haemonchus contortus glycoproteins contain N-linked oligosaccharides with novel highly fucosylated core structures*. J Biol Chem, 1996. **271**(48): p. 30561-70.
52. Haslam, S.M., et al., *The novel core fucosylation of Haemonchus contortus N-glycans is stage specific*. Mol Biochem Parasitol, 1998. **93**(1): p. 143-7.
53. Yan, S., et al., *Array-assisted characterization of a fucosyltransferase required for the biosynthesis of complex core modifications of nematode N-glycans*. J Biol Chem, 2013. **288**(29): p. 21015-28.
54. Tretter, V., et al., *Fucose alpha 1,3-linked to the core region of glycoprotein N-glycans creates an important epitope for IgE from honeybee venom allergic individuals*. Int Arch Allergy Immunol, 1993. **102**(3): p. 259-66.
55. van Die, I., et al., *Core alpha1-->3-fucose is a common modification of N-glycans in parasitic helminths and constitutes an important epitope for IgE from Haemonchus contortus infected sheep*. FEBS Lett, 1999. **463**(1-2): p. 189-93.
56. Paschinger, K. and I.B. Wilson, *Two types of galactosylated fucose motifs are present on N-glycans of Haemonchus contortus*. Glycobiology, 2015. **25**(6): p. 585-90.
57. Yan, S., et al., *Galactosylated fucose epitopes in nematodes: increased expression in a Caenorhabditis mutant associated with altered lectin sensitivity and occurrence in parasitic species*. J Biol Chem, 2012. **287**(34): p. 28276-90.
58. Titz, A., et al., *Molecular basis for galactosylation of core fucose residues in invertebrates: identification of caenorhabditis elegans N-glycan core alpha1,6-fucoside beta1,4-galactosyltransferase GALT-1 as a member of a novel glycosyltransferase family*. J Biol Chem, 2009. **284**(52): p. 36223-33.
59. Butschi, A., et al., *Caenorhabditis elegans N-glycan core beta-galactoside confers sensitivity towards nematotoxic fungal galectin CGL2*. PLoS Pathog, 2010. **6**(1): p. e1000717.

60. Schubert, M., et al., *Plasticity of the beta-trefoil protein fold in the recognition and control of invertebrate predators and parasites by a fungal defence system*. PLoS Pathog, 2012. **8**(5): p. e1002706.
61. Heim, C., et al., *Inhibition of Haemonchus contortus larval development by fungal lectins*. Parasit Vectors, 2015. **8**: p. 425.
62. Wimmerova, M., et al., *Crystal structure of fungal lectin: six-bladed beta-propeller fold and novel fucose recognition mode for Aleuria aurantia lectin*. J Biol Chem, 2003. **278**(29): p. 27059-67.
63. Fujihashi, M., et al., *Crystal structure of fucose-specific lectin from Aleuria aurantia binding ligands at three of its five sugar recognition sites*. Biochemistry, 2003. **42**(38): p. 11093-9.
64. Wohlschlager, T., et al., *Nematotoxicity of Marasmius oreades agglutinin (MOA) depends on glycolipid binding and cysteine protease activity*. J Biol Chem, 2011. **286**(35): p. 30337-43.
65. Cordara, G., et al., *Marasmius oreades agglutinin (MOA) is a chimerolectin with proteolytic activity*. Biochem Biophys Res Commun, 2011. **408**(3): p. 405-10.
66. Walti, M.A., et al., *Structural basis for chitotetraose coordination by CGL3, a novel galectin-related protein from Coprinopsis cinerea*. J Mol Biol, 2008. **379**(1): p. 146-59.
67. Wohlschlager, T., et al., *Methylated glycans as conserved targets of animal and fungal innate defense*. Proc Natl Acad Sci U S A, 2014. **111**(27): p. E2787-96.
68. Schallig, H.D. and M.A. van Leeuwen, *Carbohydrate epitopes on Haemonchus contortus antigens*. Parasitol Res, 1996. **82**(1): p. 38-42.
69. Geldhof, P., et al., *Presence of the LDNF glycan on the host-protective H-gal-GP fraction from Haemonchus contortus*. Parasite Immunol, 2005. **27**(1-2): p. 55-60.
70. Vervelde, L., et al., *Vaccination-induced protection of lambs against the parasitic nematode Haemonchus contortus correlates with high IgG antibody responses to the LDNF glycan antigen*. Glycobiology, 2003. **13**(11): p. 795-804.
71. Kooyman, F.N., et al., *Identification of a thrombospondin-like immunodominant and phosphorylcholine-containing glycoprotein (GP300) in Dictyocaulus*

- viviparus and related nematodes*. Mol Biochem Parasitol, 2009. **163**(2): p. 85-94.
72. Van den Eijnden, D.H., et al., *Novel glycosylation routes for glycoproteins: the lacdiNAc pathway*. Biochem Soc Trans, 1995. **23**(1): p. 175-9.
73. Haslam, S.M., et al., *Structural characterization of the N-glycans of Dictyocaulus viviparus: discovery of the Lewis(x) structure in a nematode*. Glycobiology, 2000. **10**(2): p. 223-9.
74. Morelle, W., et al., *Phosphorylcholine-containing N-glycans of Trichinella spiralis: identification of multiantennary lacdiNAc structures*. Glycobiology, 2000. **10**(9): p. 941-50.
75. Reason, A.J., et al., *Novel tyvelose-containing tri- and tetra-antennary N-glycans in the immunodominant antigens of the intracellular parasite Trichinella spiralis*. Glycobiology, 1994. **4**(5): p. 593-603.
76. Ellis, L.A., et al., *Terminal beta-linked tyvelose creates unique epitopes in Trichinella spiralis glycan antigens*. Glycobiology, 1997. **7**(3): p. 383-90.
77. Wilson, I.B. and K. Paschinger, *Sweet secrets of a therapeutic worm: mass-spectrometric N-glycomic analysis of Trichuris suis*. Anal Bioanal Chem, 2016. **408**(2): p. 461-71.
78. Pottl, G., et al., *N-glycans of the porcine nematode parasite Ascaris suum are modified with phosphorylcholine and core fucose residues*. FEBS J, 2007. **274**(3): p. 714-26.
79. Haslam, S.M., et al., *Structural studies of N-glycans of filarial parasites. Conservation of phosphorylcholine-substituted glycans among species and discovery of novel chito-oligomers*. J Biol Chem, 1999. **274**(30): p. 20953-60.
80. Noll, A.J., et al., *Human DC-SIGN binds specific human milk glycans*. Biochem J, 2016. **473**(10): p. 1343-53.
81. Hokke, C.H. and A. van Diepen, *Helminth glycomics - glycan repertoires and host-parasite interactions*. Mol Biochem Parasitol, 2016.
82. Borloo, J., et al., *In-depth proteomic and glycomic analysis of the adult-stage Cooperia oncophora excretome/secretome*. J Proteome Res, 2013. **12**(9): p. 3900-11.

83. Yan, S., I.B. Wilson, and K. Paschinger, *Comparison of RP-HPLC modes to analyse the N-glycome of the free-living nematode Pristionchus pacificus*. Electrophoresis, 2015. **36**(11-12): p. 1314-29.
84. Schiller, B., et al., *Complicated N-linked glycans in simple organisms*. Biol Chem, 2012. **393**(8): p. 661-73.
85. Newton, S.E. and E.N. Meeusen, *Progress and new technologies for developing vaccines against gastrointestinal nematode parasites of sheep*. Parasite Immunol, 2003. **25**(5): p. 283-96.
86. Knox, D.P., et al., *The nature and prospects for gut membrane proteins as vaccine candidates for Haemonchus contortus and other ruminant trichostrongyloids*. Int J Parasitol, 2003. **33**(11): p. 1129-37.
87. Smith, W.D. and D.S. Zarlenga, *Developments and hurdles in generating vaccines for controlling helminth parasites of grazing ruminants*. Vet Parasitol, 2006. **139**(4): p. 347-59.
88. Nisbet, A.J., et al., *Immunity to Haemonchus contortus and Vaccine Development*. Adv Parasitol, 2016. **93**: p. 353-96.
89. Knox, D., *Proteases in blood-feeding nematodes and their potential as vaccine candidates*. Adv Exp Med Biol, 2011. **712**: p. 155-76.
90. Kooyman, F.N., et al., *Production of a monoclonal antibody specific for ovine immunoglobulin E and its application to monitor serum IgE responses to Haemonchus contortus infection*. Parasitology, 1997. **114 (Pt 4)**: p. 395-406.
91. Schallig, H.D. and M.A. Van Leeuwen, *Protective immunity to the blood-feeding nematode Haemonchus contortus induced by vaccination with parasite low molecular weight antigens*. Parasitology, 1997. **114 (Pt 3)**: p. 293-9.
92. Schallig, H.D., M.A. van Leeuwen, and A.W. Cornelissen, *Protective immunity induced by vaccination with two Haemonchus contortus excretory secretory proteins in sheep*. Parasite Immunol, 1997. **19**(10): p. 447-53.
93. Munn, E.A., *Rational design of nematode vaccines: hidden antigens*. Int J Parasitol, 1997. **27**(4): p. 359-66.
94. Smith, W.D., J.A. van Wyk, and M.F. van Strijp, *Preliminary observations on the potential of gut membrane proteins of Haemonchus contortus as candidate*

- vaccine antigens in sheep on naturally infected pasture*. Veterinary Parasitology, 2001. **98**(4): p. 285-297.
95. Meier, L., P.R. Torgerson, and H. Hertzberg, *Vaccination of goats against Haemonchus contortus with the gut membrane proteins H11/H-gal-GP*. Veterinary Parasitology, 2016. **229**: p. 15-21.
 96. Andrews, S.J., T.P. Rolph, and E.A. Munn, *Duration of protective immunity against ovine haemonchosis following vaccination with the nematode gut membrane antigen H11*. Res Vet Sci, 1997. **62**(3): p. 223-7.
 97. Williamson, A.L., et al., *Digestive proteases of blood-feeding nematodes*. Trends Parasitol, 2003. **19**(9): p. 417-23.
 98. Williamson, A.L., et al., *A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms*. J Biol Chem, 2004. **279**(34): p. 35950-7.
 99. Smith, T.S., et al., *Purification and evaluation of the integral membrane protein H11 as a protective antigen against Haemonchus contortus*. Int J Parasitol, 1993. **23**(2): p. 271-80.
 100. Smith, W.D., S.K. Smith, and J.M. Murray, *Protection studies with integral membrane fractions of Haemonchus contortus*. Parasite Immunol, 1994. **16**(5): p. 231-41.
 101. Knox, D.P., S.K. Smith, and W.D. Smith, *Immunization with an affinity purified protein extract from the adult parasite protects lambs against infection with Haemonchus contortus*. Parasite Immunol, 1999. **21**(4): p. 201-10.
 102. Munn, E.A., et al., *Vaccination against Haemonchus contortus with denatured forms of the protective antigen H11*. Parasite Immunol, 1997. **19**(6): p. 243-8.
 103. Ekoja, S.E. and W.D. Smith, *Antibodies from sheep immunized against Haemonchus contortus with H-gal-GP inhibit the haemoglobinase activity of this protease complex*. Parasite Immunology, 2010. **32**(11-12): p. 731-738.
 104. Smith, W.D., *Protection in lambs immunised with Haemonchus contortus gut membrane proteins*. Res Vet Sci, 1993. **54**(1): p. 94-101.

105. Jasmer, D.P. and T.C. McGuire, *Protective immunity to a blood-feeding nematode (Haemonchus contortus) induced by parasite gut antigens*. Infect Immun, 1991. **59**(12): p. 4412-7.
106. Andrews, S.J., et al., *Vaccination of sheep against haemonchosis with H11, a gut membrane-derived protective antigen from the adult parasite: prevention of the periparturient rise and colostral transfer of protective immunity*. Int J Parasitol, 1995. **25**(7): p. 839-46.
107. Matthews, J.B., et al., *Progress in the development of subunit vaccines for gastrointestinal nematodes of ruminants*. Parasite Immunology, 2016. **38**(12): p. 744-753.
108. Dicker, A.J., et al., *Proteomic analysis of Mecistocirrus digitatus and Haemonchus contortus intestinal protein extracts and subsequent efficacy testing in a vaccine trial*. PLoS Negl Trop Dis, 2014. **8**(6): p. e2909.
109. Smith, T.S., et al., *Cloning and characterization of a microsomal aminopeptidase from the intestine of the nematode Haemonchus contortus*. Biochim Biophys Acta, 1997. **1338**(2): p. 295-306.
110. Mohandas, N., et al., *The complement of family M1 aminopeptidases of Haemonchus contortus--Biotechnological implications*. Biotechnol Adv, 2016. **34**(2): p. 65-76.
111. Roberts, B., et al., *Novel expression of Haemonchus contortus vaccine candidate aminopeptidase H11 using the free-living nematode Caenorhabditis elegans*. Vet Res, 2013. **44**: p. 111.
112. Wong, A.H., D. Zhou, and J.M. Rini, *The X-ray crystal structure of human aminopeptidase N reveals a novel dimer and the basis for peptide processing*. J Biol Chem, 2012. **287**(44): p. 36804-13.
113. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis*. Nat Protoc, 2015. **10**(6): p. 845-58.
114. Longbottom, D., et al., *Molecular cloning and characterisation of a putative aspartate proteinase associated with a gut membrane protein complex from adult Haemonchus contortus*. Mol Biochem Parasitol, 1997. **88**(1-2): p. 63-72.

115. Newlands, G.F., et al., *Molecular characterization of a family of metalloendopeptidases from the intestinal brush border of Haemonchus contortus*. Parasitology, 2006. **133**(Pt 3): p. 357-68.
116. Redmond, D.L., et al., *Molecular cloning and characterisation of a developmentally regulated putative metallopeptidase present in a host protective extract of Haemonchus contortus*. Mol Biochem Parasitol, 1997. **85**(1): p. 77-87.
117. Skuce, P.J., et al., *Cloning and characterisation of thrombospondin, a novel multidomain glycoprotein found in association with a host protective gut extract from Haemonchus contortus*. Mol Biochem Parasitol, 2001. **117**(2): p. 241-4.
118. Newlands, G.F., et al., *Cloning and characterization of a beta-galactoside-binding protein (galectin) from the gut of the gastrointestinal nematode parasite Haemonchus contortus*. Parasitology, 1999. **119 (Pt 5)**: p. 483-90.
119. Newlands, G.F., et al., *Cloning and expression of cystatin, a potent cysteine protease inhibitor from the gut of Haemonchus contortus*. Parasitology, 2001. **122**(Pt 3): p. 371-8.
120. Muench, S., et al., *3D Reconstruction of a Large Protease Complex from Haemonchus contortus*. Biophysical Journal, 2008. **94**(2, Supplement): p. 536-539.
121. Smith, S.K., et al., *Further immunization and biochemical studies with a protective antigen complex from the microvillar membrane of the intestine of Haemonchus contortus*. Parasite Immunol, 1999. **21**(4): p. 187-99.
122. Smith, S.K. and W.D. Smith, *Immunisation of sheep with an integral membrane glycoprotein complex of Haemonchus contortus and with its major polypeptide components*. Res Vet Sci, 1996. **60**(1): p. 1-6.
123. Boisvenue, R.J., et al., *Fibrinogen-degrading proteins from Haemonchus contortus used to vaccinate sheep*. Am J Vet Res, 1992. **53**(7): p. 1263-5.
124. Knox, D.P., D.L. Redmond, and D.G. Jones, *Characterization of proteinases in extracts of adult Haemonchus contortus, the ovine abomasal nematode*. Parasitology, 1993. **106 (Pt 4)**: p. 395-404.

125. Cox, G.N., et al., *Molecular cloning and primary sequence of a cysteine protease expressed by Haemonchus contortus adult worms*. Mol Biochem Parasitol, 1990. **41**(1): p. 25-34.
126. Pratt, D., et al., *Cloning and sequence comparisons of four distinct cysteine proteases expressed by Haemonchus contortus adult worms*. Mol Biochem Parasitol, 1992. **51**(2): p. 209-18.
127. Rehman, A. and D.P. Jasmer, *A tissue specific approach for analysis of membrane and secreted protein antigens from Haemonchus contortus gut and its application to diverse nematode species*. Mol Biochem Parasitol, 1998. **97**(1-2): p. 55-68.
128. Skuce, P.J., et al., *Molecular cloning and characterization of gut-derived cysteine proteinases associated with a host protective extract from Haemonchus contortus*. Parasitology, 1999. **119 (Pt 4)**: p. 405-12.
129. Redmond, D.L. and D.P. Knox, *Protection studies in sheep using affinity-purified and recombinant cysteine proteinases of adult Haemonchus contortus*. Vaccine, 2004. **22**(31-32): p. 4252-61.
130. Reszka, N., et al., *Haemonchus contortus: characterization of the baculovirus expressed form of aminopeptidase H11*. Exp Parasitol, 2007. **117**(2): p. 208-13.
131. Smith, W.D., et al., *Metalloendopeptidases from the intestinal brush border of Haemonchus contortus as protective antigens for sheep*. Parasite Immunol, 2003. **25**(6): p. 313-23.
132. Smith, W.D., et al., *Aspartyl proteases from the intestinal brush border of Haemonchus contortus as protective antigens for sheep*. Parasite Immunol, 2003. **25**(11-12): p. 521-30.
133. Cachat, E., et al., *Attempts to immunize sheep against Haemonchus contortus using a cocktail of recombinant proteases derived from the protective antigen, H-gal-GP*. Parasite Immunol, 2010. **32**(6): p. 414-9.
134. Redmond, D.L. and D.P. Knox, *Further protection studies using recombinant forms of Haemonchus contortus cysteine proteinases*. Parasite Immunol, 2006. **28**(5): p. 213-9.

Chapter 2

A glycoengineering approach for production of *Haemonchus contortus* vaccine candidates in insect cells

Susanna M. Fleurkens¹, Christine Neupert², Chia-wei Lin¹, Hubertus Hertzberg³,
Irene Schiller², Bruno Oesch², Peter Deplazes³, Markus Aebi¹

¹ Institute of Microbiology, Department of Biology, ETH Zurich,
CH-8093 Zurich, Switzerland

² Malcisbo AG, Wagistrasse 27a, CH-8952 Schlieren, Switzerland

³ Institute of Parasitology, Winterthurerstrasse 266a, University of Zurich,
CH-8057 Zurich, Switzerland

Introduction

From a global perspective, *Haemonchus contortus* is the most important blood-feeding parasitic nematode in sheep and goats causing a high economic impact worldwide with major losses in milk, wool and meat production. Anthelmintic drugs currently control the disease, however the rapid spread of drug-resistant strains indicates the need for new strategies [1, 2]. The most practical solution is the development of a safe and effective vaccine. For *Haemonchus contortus*, native vaccines from excretory / secretory products (E/S) as well as gut-derived complexes induce high levels of protection. A reduction of fecal egg counts by up to 95 % and worm burden by 82 % could be detected [1, 3, 4]. The protective gut fractions derive from detergent-soluble extracts of the adult parasite and contain hidden antigens that do not induce an antibody response during the course of a natural infection [5]. These antigens are expressed on the microvillar surface of the intestine and show protease activity, which is believed to degrade host blood proteins such as hemoglobin and albumin [6]. To date, the most promising native vaccines include the aminopeptidase H11 enriched in a Concanavalin A lectin-binding fraction [7], the glycoprotein complex H-gal-GP containing metallo and aspartic proteases enriched in a Peanut lectin binding fraction [8], and cysteine proteases enriched in a thiol-binding fraction [9]. The activity of the enzymes H11 and H-gal-GP is inhibited by IgG antibodies from vaccine-protected sheep [10, 11].

Despite the success of native vaccines, the long term aim is to produce a uniform and safe recombinant vaccine in a cost-effective method regime. For this reason, there has been a continuous effort to produce recombinant versions of the native antigens in bacteria, yeast, insect cells or *Caenorhabditis elegans* as expression systems [12]. The digestive proteases were successfully expressed in an active form, but induced only low levels of protection when tested in sheep vaccination trials [13-15]. One might speculate that this reduced protective capacity of the recombinant vaccines is due to suboptimal folding, absence of other proteins critical in conferring protection in the complex or lack of various post-translational modifications. One of the main differences between recombinant and native proteins is the *N*-linked glycosylation present on native proteins. In *Haemonchus contortus*, unusual carbohydrate modifications with up to three core fucose residues were identified on the native aminopeptidase H11 [16]. One of these modifications, the core α 1,3 Fuc epitope is recognized by parasite

specific IgE antibody from *H. contortus* infected sheep and may contribute to an induction of a Th2 response [17]. The core α 1,3 Fuc epitope is a common antigen synthesized by many parasitic helminths and is highly immunogenic when present on plant or insect glycoproteins [17, 18]. The *Trichoplusia ni* cell line BTI-TN-5B1-4 (High five) produces *N*-glycan structures with core α 1,3 and α 1,6 fucosylation [19] equal to the ones found in nematodes [20] and is therefore an ideal expression system for recombinant glycoproteins. Moreover, the baculovirus insect cell system offers a fast and easy production system for secreting high amounts of recombinant glycoproteins carrying post-translational protein modifications, in particular *N*-linked glycosylation [21, 22]. In nematodes, α 1,3 and α 1,6 core fucoses are often additionally decorated with galactose residues, creating another nematode-specific antigen, the 'GalFuc' epitope [23, 24]. *Coprinopsis* galectin 2 (CGL2, [25]) recognizes the 'GalFuc' antigen and *Coprinopsis cinerea* lectin 2 (CCL2 [26]) binds the α 1,3 Fuc epitope. These two fungal lectins were shown to be toxic towards *Haemonchus contortus* in an *in vitro* larval development assay and bind the digestive tract of the larvae, as well as the brush border of the adult parasite gut [27]. The enzyme GALT-1, a core α 1,6-fucoside β 1,4-galactosyltransferase-1 is generating the GalFuc epitope in *C. elegans* [28]. By overexpressing the galactosyltransferase-1 in High Five insect cells and thereby manipulating the insect *N*-glycosylation pathway, a recombinant glycoprotein carrying this epitope was successfully produced [29].

Glycoengineered insect cells might therefore be a suitable expression system to produce recombinant *H. contortus* vaccine candidates with nematode-like glycans to mimic the *H. contortus* *N*-glycans. In this study, we co-express the *C. elegans* galactosyltransferase-1 in insect cells together with three well-studied *H. contortus* digestive proteases, H11, AC1, and PEP1, and characterize their overall and site-specific glycosylation patterns. We test their efficacy in a sheep immunization trial and analyze the sheep immune response. Our results are relevant for the production of other nematode vaccines in glycoengineered insect cells and requirements for vaccine-induced protective immunity.

Results

Recombinant expression of *H. contortus* proteases H11, AC1 and PEP1 using glycoengineered High five insect cells

The baculovirus insect cell system was chosen for protein expression in High five (*Trichoplusia ni*) cells due to its ability to produce high amounts of recombinant glycoproteins carrying post-translational protein modifications. In addition, High five cells decorate secreted proteins with *N*-glycan structures similar to the ones found in *Haemonchus contortus* [20]. In order to produce recombinant glycoproteins carrying nematode-like *N*-glycans we manipulated the insect *N*-glycosylation pathway by overexpressing an additional glycosyltransferase, the galactosyltransferase GALT-1 from *Caenorhabditis elegans*. GALT-1 adds a β 1,4-galactose to the α 1,6-linked core fucose residue [28] thereby creating the nematode-like glycoepitope GalFuc. Another nematode-like glycoepitope, the α 1,3-Fuc antigen is occurring on proteins secreted by High five cells [30].

Three different *H. contortus* proteins, the proteases H11, AC1 and PEP1, were chosen for recombinant expression, because they were shown to be components of protective native vaccines. The *H. contortus* protease expression constructs were designed for secretion of H11, AC1 and PEP1 into the medium with a gp67 signal peptide sequence targeting the proteins to the secretory pathway (Figure 1A). The N-terminal transmembrane domain of H11 was substituted by the gp67 signal peptide. An N- (H11, AC1) or C-terminal (AC1, PEP1) His₁₀-tag was added for affinity purification of the recombinant proteins. In addition, a MBP-tag was fused to the N-terminus of PEP1 to increase the solubility of the protein and to allow its expression [31]. Over 72 hours, the expression levels of the *H. contortus* proteins with GALT-1 were monitored daily by immunoblotting with antibodies against the tags. The expression of the galactosyltransferase-1 was stable during these three days and the amount of secreted proteases was increasing with time of incubation. Therefore, the digestive proteases were co-expressed with GALT-1 for 72 hours (Figure 1B) and affinity purified from the culture supernatant (Figure 1C) to further analyze the *N*-glycosylation patterns on the recombinant proteins in the next step.

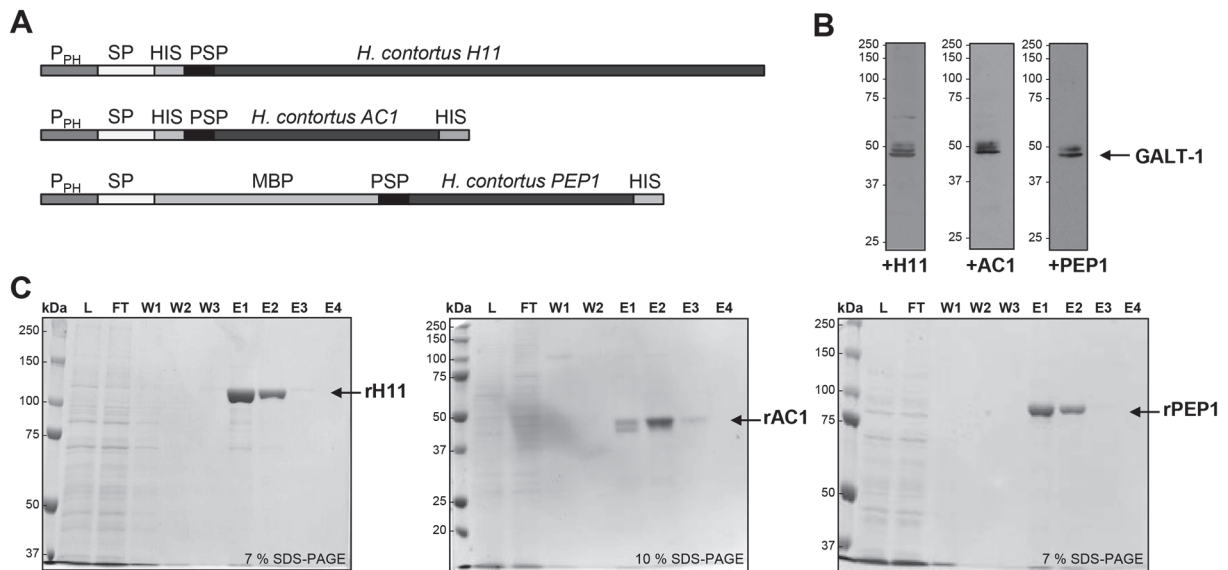


Figure 1: Purification of recombinant *H. contortus* H11, AC1, and PEP1 from glycoengineered High five insect cells

(A) Schematic representation of the H11, AC1, and PEP1 expression constructs cloned into pFastBac1 vector downstream of the Polyhedrin promoter (P_{PH}). All constructs encode a gp67 secretion signal peptide (SP), one or two tags (HIS: His₁₀-tag, MBP: Maltose binding protein) and a PreScission Protease cleavage site (PSP) between the N-terminal tag and the respective gene. **(B)** GALT-1 expression in High five cells co-expressing GALT-1 and H11, AC1 or PEP1 (+H11, +AC1, +PEP1) was confirmed by immunoblotting with anti-FLAG antibody. Equal amounts of cells were harvested and lysed 72 hours post co-infection with *H. contortus* proteins. **(C)** Ni-NTA affinity purification of secreted recombinant *H. contortus* proteins H11, AC1 and PEP1 from culture supernatant 72 hours post co-expression with GALT-1. Samples of load (L), flow-through (FT), wash (W1-W3) and elution (E1-E4) fractions were analyzed by SDS-PAGE and stained with Coomassie. Purified proteins were identified by MS peptide sequencing.

Glycoengineered High five-expressed proteases are decorated with nematode-like *N*-glycans

To identify the specific *N*-glycans present on glycoengineered High five-expressed proteases, an overall *N*-glycan profiling by HPLC analysis was performed as well as a detailed glycosite-specific *N*-glycan analysis using a nanoHPLC-HCD-MS/MS mass spectrometry method [29]. To get the overall *N*-glycan profile, *N*-glycans were fluorescently-labelled with 2-aminobenzamide (2-AB) after enzymatic release with peptide-*N*-glycosidase A (PNGase A) and fractionated by HPLC using a normal-phase-amino column. Mostly, the HPLC chromatograms of the glycoengineered recombinant

proteases appear to be similar. However, three additional elution peaks (P7, P9 and P11) emerge in the chromatogram of H11 when co-expressed with the galactosyltransferase-1 compared to an empty vector control (Figure 2). The same could be shown for AC1 and PEP1 (Figure S1). To identify the composition of the *N*-glycans in the chromatogram, all elution peaks were collected and analyzed by MALDI-TOF MS and MS/MS (Figure S2, Table S1).

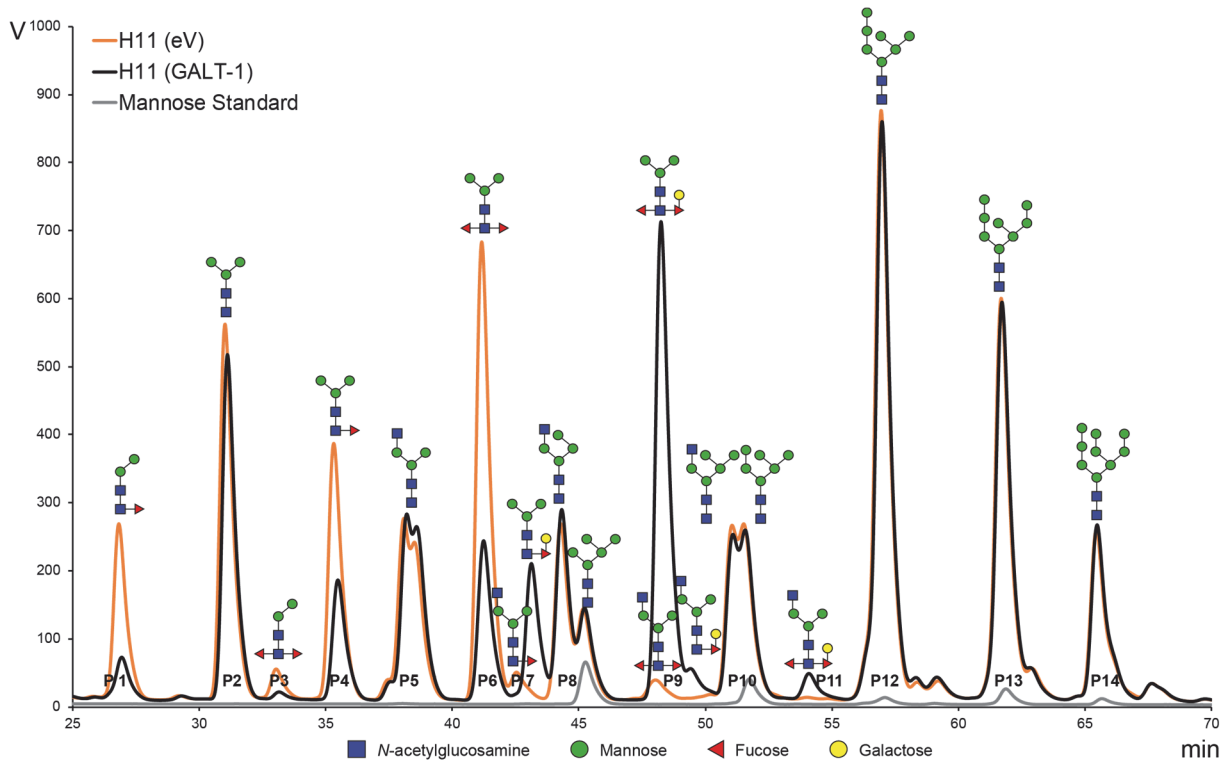


Figure 2: Glycoengineered High five-expressed H11 is decorated with nematode-like *N*-glycans

Overall *N*-glycan HPLC profile of recombinant H11 purified from insect cells co-expressing either an empty vector control (eV, blue) or the *C. elegans* galactosyltransferase-1 (GALT-1, orange). Purified proteins were reduced, alkylated, and proteolyzed with trypsin. The oligosaccharides were released with PNGase A, 2-AB labelled, and separated by HPLC using a normal-phase-amino column and fluorescence detection. Peaks were collected and analyzed by MALDI-TOF MS and MS/MS. The chromatography shows the relative amount of each glycoform. A HPLC profile of a high mannose standard is shown in grey (GlcNAc₂Man₍₅₋₉₎ from RNase B).

MALDI-TOF MS analysis verified *N*-glycan compositions for high mannose structures ranging from GlcNAc₂Man₅ to GlcNAc₂Man₉, preassigned based on the HPLC chromatogram of a mannose standard (Figure 2, grey, peaks 8, 10 and 12-14). The Y ions were observed as single-charged potassium adducts [M+K]⁺. The GlcNAc₂Man₃Fuc₁ composition, pre-assigned to peak 4 by a standard, was also confirmed with a major peak of m/z 1215.4 by MS and upon MS/MS, a dominant fragment of m/z 526.2 (2-AB-HexNAc₁Fuc₁) verified the core position of the fucose residue. Whether the fucose residue is α1,3- or α1,6-linked to the core *N*-acetylglucosamine cannot be distinguished. In addition, two other core monofucosylated *N*-glycans were detected in fraction 1 (m/z 1053.4 (2-AB-HexNAc₂Hex₂)Fuc₁) and fraction 7 (m/z 1418.6 (2-AB-HexNAc₂Hex₃)HexNAc₁Fuc₁, Figure S2A). Furthermore, difucosylated *N*-glycans were detected, as previously identified on proteins secreted by High five cells [30]. MS/MS fragmentation of *N*-glycans in fraction 3 (m/z 1199.4 (2-AB-HexNAc₂Hex₂)Fuc₂), fraction 6 (m/z 1361.5 (2-AB-HexNAc₂Hex₃)Fuc₂) and fraction 9 (m/z 1564.6 (2-AB-HexNAc₂Hex₃)HexNAc₁Fuc₂) yielded a major fragment of m/z 672.2 (2-AB-HexNAc₁Fuc₂), indicative of difucosylation of the core *N*-acetylglucosamine (data not shown).

Finally, the additional peaks emerging uniquely in the chromatogram of the proteases co-expressed with the galactosyltransferase-1 were analyzed (Figure 2 and S1, peaks 7, 9 and 11). The appearance of these peaks suggests that these *N*-glycan structures are further decorated with a galactose residue due to the GALT-1 activity. MS analysis of fraction 7 (m/z 1377.6, Figure S2B), fraction 9 (m/z 1523.5 and m/z 1580.5) and fraction 11 (m/z 1726.6) indicated the addition of a hexose (2-AB-HexNAc₂Hex₃)HexNAc₀₋₁Fuc₁₋₂Hex₁). Additional evidence came from the MS/MS analysis, where dominant fragments of m/z 688.2 and m/z 834.3 (2-AB-HexNAc₁Fuc₁₋₂Hex₁) verified the core position of the additional hexose (Figure S2C and D). This hexose is most likely a galactose modification due to the co-expression with the galactosyltransferase-1.

Taken together, the HPLC *N*-glycan analysis on glycoengineered recombinant proteases H11, AC1, and PEP1 revealed that 42-67 % of all difucosylated structures were decorated with a galactose and overall an amount of 18-26 % galactosylated *N*-glycan structures were found (Table 1).

Table 1: Relative N-glycan abundance of recombinant H11, AC1, and PEP1 purified from insect cells co-expressing the galactosyltransferase-1 (GALT-1) measured by HPLC.

N-glycans resulting from three independent purifications (biological replicate 1-3, BR1-BR3) of H11, AC1, and PEP1 co-expressing GALT-1 were processed for HPLC analysis (see also Figure 2). Quantification of the HPLC peak areas was performed using the Chromleon software and peak area ratios were calculated as the relative amount of each glycoform. Mean values of the peak area ratios are listed with the standard deviation (SD) of the mean. Peak composition was analyzed by MALDI-TOF MS and MS/MS and can be found in the 'Glycan structures' column. Oligomannosidic and hybrid (harboring an additional HexNAc) N-glycans were detected in peaks 2, 5, 8, 10 and 12-14. Core fucosylated structures were found in fractions 1 and 4 and difucosylated in fractions 3 and 6. The addition of a galactose to the core fucose was detected in peaks 7, 9 and 11 and is summarized under 'core galactosylation'.

Peak	H11 (GALT-1)						AC1 (GALT-1)						PEP1 (GALT-1)						Glycan structures
	Peak area ratio		Mean	SD	Peak area ratio		Mean	SD	Peak area ratio		Mean	SD	Peak area ratio		Mean	SD			
	BR1	BR2			BR3	BR1			BR2	BR3			BR1	BR2			BR3	BR1	
P1	1.3	1.5	1.6	1.5	0.2	1.9	1.8	1.6	1.8	0.1	2.6	3.8	2.1	2.8	0.9	HexNAc ₂ Hex ₂ Fuc			
P2	9.6	7.9	12.8	10.1	2.5	19.0	16.9	19.7	18.5	1.5	5.0	4.8	6.2	5.3	0.7	HexNAc ₂ Hex ₃			
P3	0.3	0.4	1.7	0.8	0.8	1.2	1.3	2.1	1.5	0.5	6.2	10.1	1.2	5.8	4.5	HexNAc ₂ Hex ₂ Fuc ₂			
P4	3.4	5.8	6.9	5.4	1.8	15.7	14.6	12.8	14.4	1.5	11.2	13.1	17.6	14.0	3.3	(HexNAc ₂ Hex ₃)Fuc			
P5	8.8	9.8	15.2	11.3	3.4	4.6	4.7	5.9	5.1	0.7	5.6	5.3	5.1	5.3	0.3	(HexNAc ₂ Hex ₃)HexNAc			
P6	4.6	6.6	12.7	8.0	4.2	12.7	13.2	20.9	15.6	4.6	29.6	38.5	21.7	29.9	8.5	(HexNAc ₂ Hex ₃)Fuc ₂			
P7	3.9	3.7	4.3	3.9	0.3	10.7	10.5	5.9	9.1	2.7	6.3	4.3	6.0	5.5	1.1	(HexNAc ₂ Hex ₃)HexNAcFuc, (HexNAc ₂ Hex ₃)FucHex			
P8	8.1	11.0	9.2	9.4	1.5	4.3	4.3	7.0	5.2	1.6	4.4	3.7	4.7	4.3	0.5	HexNAc ₂ Hex ₄ HexNAc, HexNAc ₂ Hex ₅			
P9	14.7	11.9	12.7	13.1	1.5	10.8	10.8	9.1	10.2	1.0	22.2	12.4	20.5	18.4	5.2	(HexNAc ₂ Hex ₃)Fuc ₂ Hex, (HexNAc ₂ Hex ₃)HexNAcFucHex, (HexNAc ₂ Hex ₃)HexNAcFuc ₂			
P10	8.8	9.9	6.3	8.3	1.8	3.1	3.0	3.7	3.3	0.4	2.0	1.4	3.9	2.5	1.3	HexNAc ₂ Hex ₅ HexNAc, HexNAc ₂ Hex ₆			
P11	0.8	0.7	1.6	1.0	0.5	1.2	1.2	0.8	1.1	0.2	2.8	1.4	2.6	2.3	0.7	(HexNAc ₂ Hex ₃)HexNAcFuc ₂ Hex			
P12	18.0	15.4	8.6	14.0	4.9	4.8	5.0	3.6	4.5	0.7	1.0	0.6	3.5	1.7	1.6	HexNAc ₂ Hex ₇			
P13	12.5	10.7	4.8	9.3	4.0	2.8	3.0	1.7	2.5	0.7	0.7	0.4	2.0	1.0	0.9	HexNAc ₂ Hex ₈			
P14	5.3	4.6	1.5	3.8	2.0	7.1	9.7	5.0	7.3	2.4	0.2	0.1	3.2	1.2	1.7	HexNAc ₂ Hex ₉			
	71.1	69.4	58.4	66.3	8.2	45.8	46.6	46.7	46.3	3.5	19.0	16.3	28.5	21.3	3.0	Oligomannose + Hybrid			
	4.7	7.2	8.6	6.8	1.8	17.6	16.4	14.4	16.2	1.5	13.8	16.9	19.7	16.8	3.4	Core fucosylation			
	4.9	7.0	14.5	8.8	4.3	13.9	14.5	23.0	17.2	4.7	35.9	48.6	22.8	35.8	9.6	Core difucosylation			
	19.4	16.4	18.5	18.1	1.6	22.7	22.5	15.9	20.4	2.9	31.3	18.1	29.0	26.2	5.4	Core galactosylation			

The proteases H11 and AC1 have four *N*-glycosylation sites, whereas PEP1 has five. Therefore, together with the overall *N*-glycan analysis by HPLC, a detailed glycosite-specific *N*-glycan analysis was performed. After proteolytic cleavage of purified proteins with trypsin, the trypsin digested glycopeptides were analyzed by nanoHPLC-HCD-MS/MS mass spectrometry to assess the *N*-glycan microheterogeneity at each *N*-glycosylation site. The respective percentage of the carbohydrates attached to these *N*-glycosylation sites was calculated (Table S2, Figure 3). We observed, that the first two *N*-glycosites of H11 and AC1 were modified mainly by oligomannose structures; and the last two *N*-glycosites were decorated with further processed *N*-glycans harboring core fucoses and an additional galactose when co-expressed with GALT-1. This explained the low levels of galactosylated *N*-glycan structures found on H11 and AC1 (18 % and 20 % respectively). In contrast, four out of five *N*-glycosites of PEP1 were detected to be decorated with core difucosylated *N*-glycan structures and approximately half of them were galactosylated when co-expressed with GALT-1.

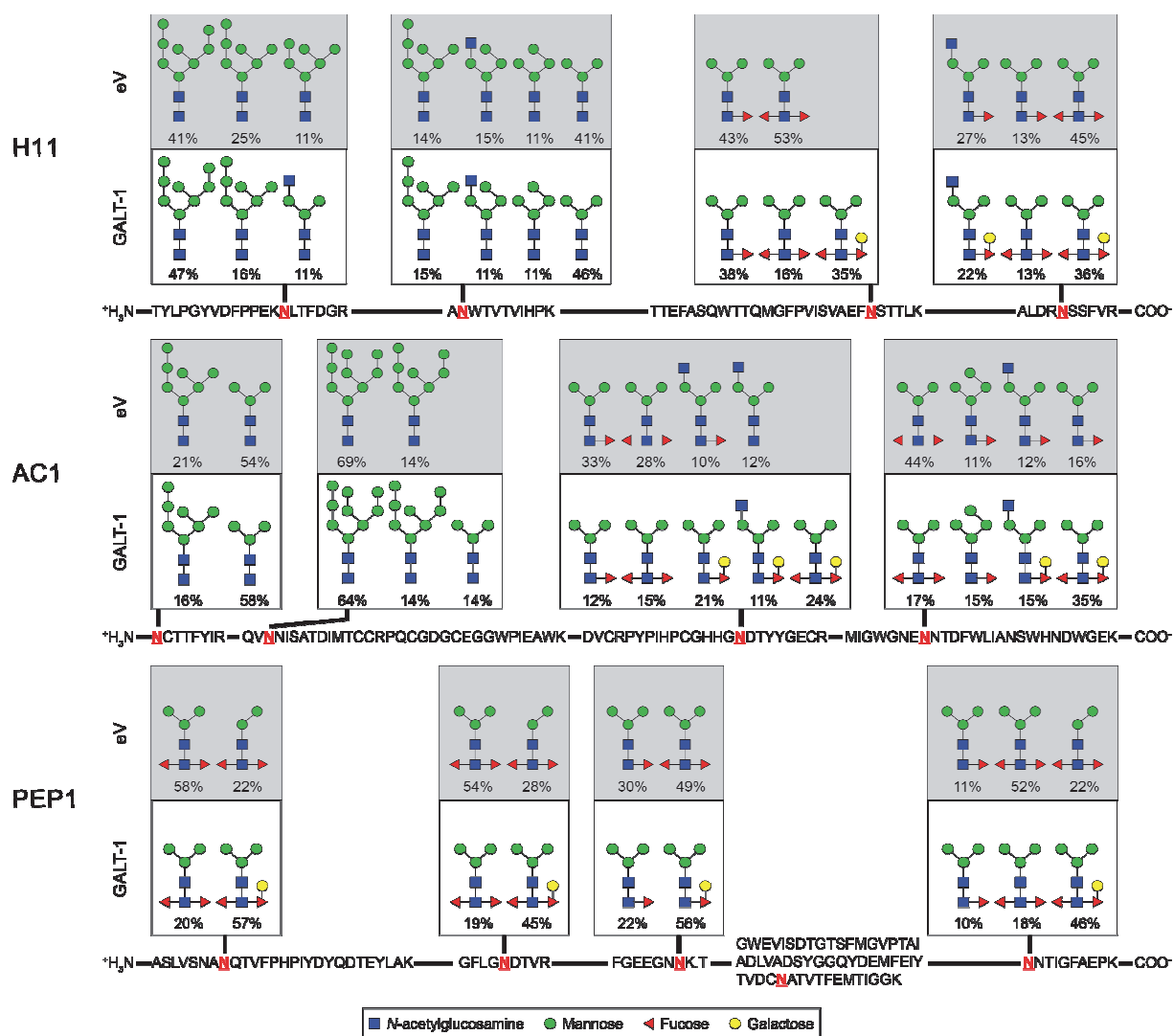


Figure 3: Recombinant H11, AC1, and PEP1 purified from glycoengineered insect cells show site-specific modification with oligomannosidic and nematode-like N-glycans.

Glycosite-specific N-glycosylation profile of H11 (top), AC1 (middle) or PEP1 (bottom) purified from insect cells co-expressing either an empty vector control (eV, upper panel, grey) or the galactosyltransferase GALT-1 (lower panel, white). Trypsin digested glycopeptides were analyzed by nanoHPLC-HCD MS/MS mass spectrometry. H11 and AC1 have four N-glycosylation sites and PEP1 has five, their location is marked by red N's representing the asparagine residue where the glycan is attached. Trypsin digested peptide sequences of these four to five glycosylation sites are shown with the most abundant (>10 %) carbohydrates attached and their respective percentage. Detailed data analysis can be found in Supplementary Table 2.

Vaccine trial using recombinant H11, AC1, and PEP1

To examine the immunoprotective properties of the recombinantly produced glycoengineered proteins, we designed a sheep vaccination trial. Six lambs per group were vaccinated three times at three weekly intervals with recombinant proteins or with native H11 extract [7] (Table 2). Vaccinations were done using PEP1 alone (-) or with a clostridial vaccine (+Covexin 8), as well as combinations of PEP1 with the proteases AC1 and H11 (AP+, HAP+, HAP-). Two different adjuvant combinations were used, the saponin QuilA alone or in combination with the aluminium based adjuvant Rehydragel. As a control group, six lambs were not treated with any vaccine (Control).

Table 2: Group design of the vaccine trial

Sheep were immunized with 30 - 100 µg of protein listed in the 'applied antigen' column with the corresponding adjuvant. Clostridial Covexin 8 vaccination is indicated in the last column.

Trial group	Applied antigen	Amount	Adjuvant	Covexin 8
Native+	Native H11 extract	30 µg	QuilA	+
P+	PEP1	100 µg	QuilA	+
P-	PEP1	100 µg	QuilA	-
AP+	AC1/PEP1	100 µg each	QuilA	+
HAP+	H11/AC1/PEP1	100 µg each	QuilA / Rehydragel	+
HAP-	H11/AC1/PEP1	100 µg each	QuilA / Rehydragel	-
Control	-	-	-	-

All animals were orally infected with 5000 infective *H. contortus* larvae one week after the third vaccination. Fecal egg counts (eggs per gram feces, EPG) were monitored for four weeks until necropsy (Figure 4A). PEP1 vaccinated sheep showed only a minor reduction in egg counts (up to 28.4 % reduction) and no reduction in worm burden as compared to the unvaccinated control group. This reduction was not increased by the additional vaccination with other recombinant proteins (up to 23.9 % EPG reduction, up to 13.5 % worm reduction). In contrast, lambs vaccinated with native H11 extract showed significant reductions in both egg counts (98.7 %) and worm burden (83.4 %) as compared to the control group (Figure 4B and C, Table S3).

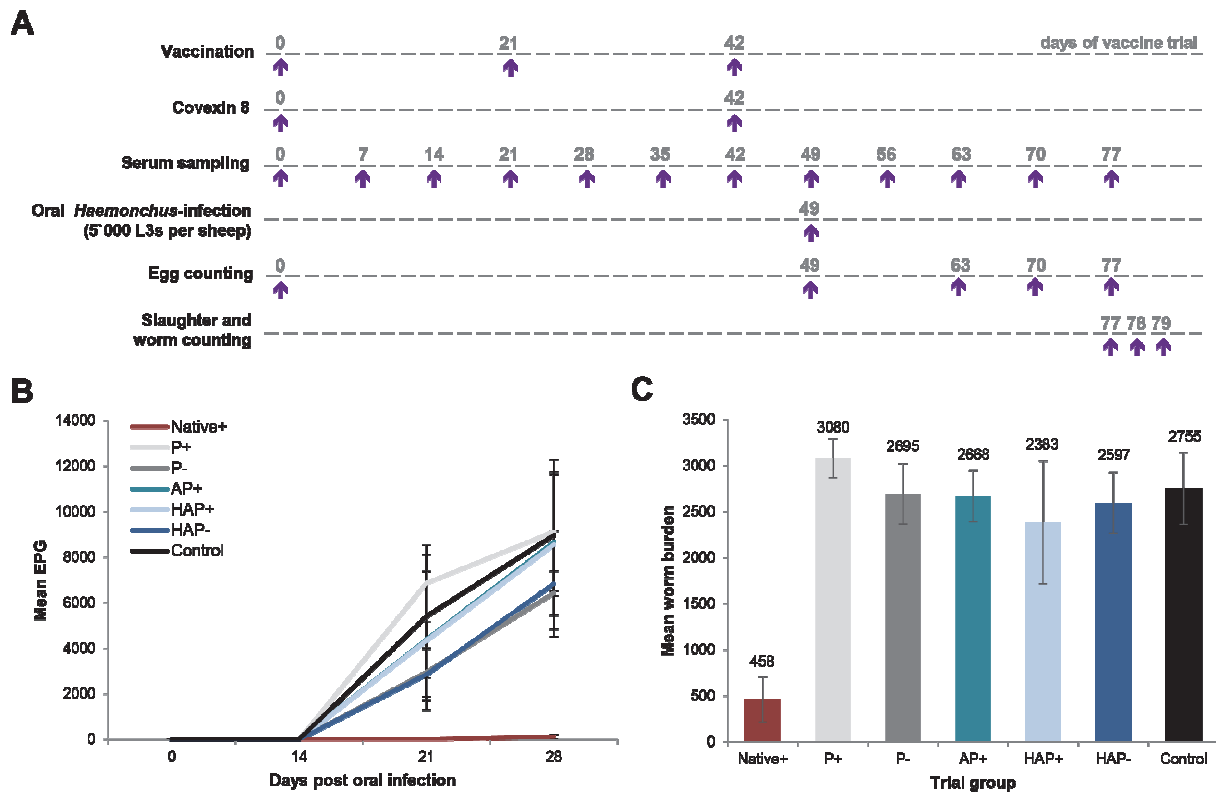


Figure 4: Sheep vaccine trial using glycoengineered recombinant H11, AC1, and PEP1

(A) Timeline of the sheep vaccination and challenge trial with the different treatments indicated to the left (vaccination, serum sampling etc.) and their respective time points to the right (purple arrows). (B) Eggs per gram feces (EPG) were monitored at day 0, 14, 21 and 28 post oral infection with 5000 *H. contortus* larvae and the worm burden (C) was counted at necropsy day. Mean group counts (group size = 6) are shown with error bars representing the standard deviation of the mean.

Antibody response to recombinant and native H11 extract proteins

To test whether the vaccinations with recombinant and native H11 extract proteins induced specific immune responses, sera of vaccinated animals were analyzed by ELISA (Figure 5). Sheep immunized with recombinant proteases showed comparable IgG antibody titres against the tested antigens (H11, AC1, PEP1) that reached the maximum one week after the third vaccination (day 49) and maintained high levels for four weeks (Figure 5B, C, and D). Maximum response was observed against the respective protein used in the vaccination (e.g. highest recognition of PEP1 by PEP1 vaccinated sheep serum). Some cross-recognition to other recombinant proteins was observed, most likely due to anti-glycan antibodies, as all recombinant proteins are decorated by identical nematode-like glycans (Figure 5B). Another possibility are antibodies against the His-tag which they have in common. Recombinant H11, AC1, and PEP1 were equally recognized when lambs were vaccinated with all three proteins

(Figure 5D). Antibody responses to the recombinant proteins rose earlier and to a higher maximum level when lambs were immunized with recombinant proteins as compared to lambs immunized with native H11 extract proteins (Figure 5A). This was expected, as the native H11 extract contains a number of other proteins in addition to H11 and the antibody response towards these extract proteins is probably higher. Nevertheless, lambs immunized with native extracts recognized recombinant H11 decorated with nematode-like glycans, as well as recombinant AC1 and PEP1 even if to a lower extent.

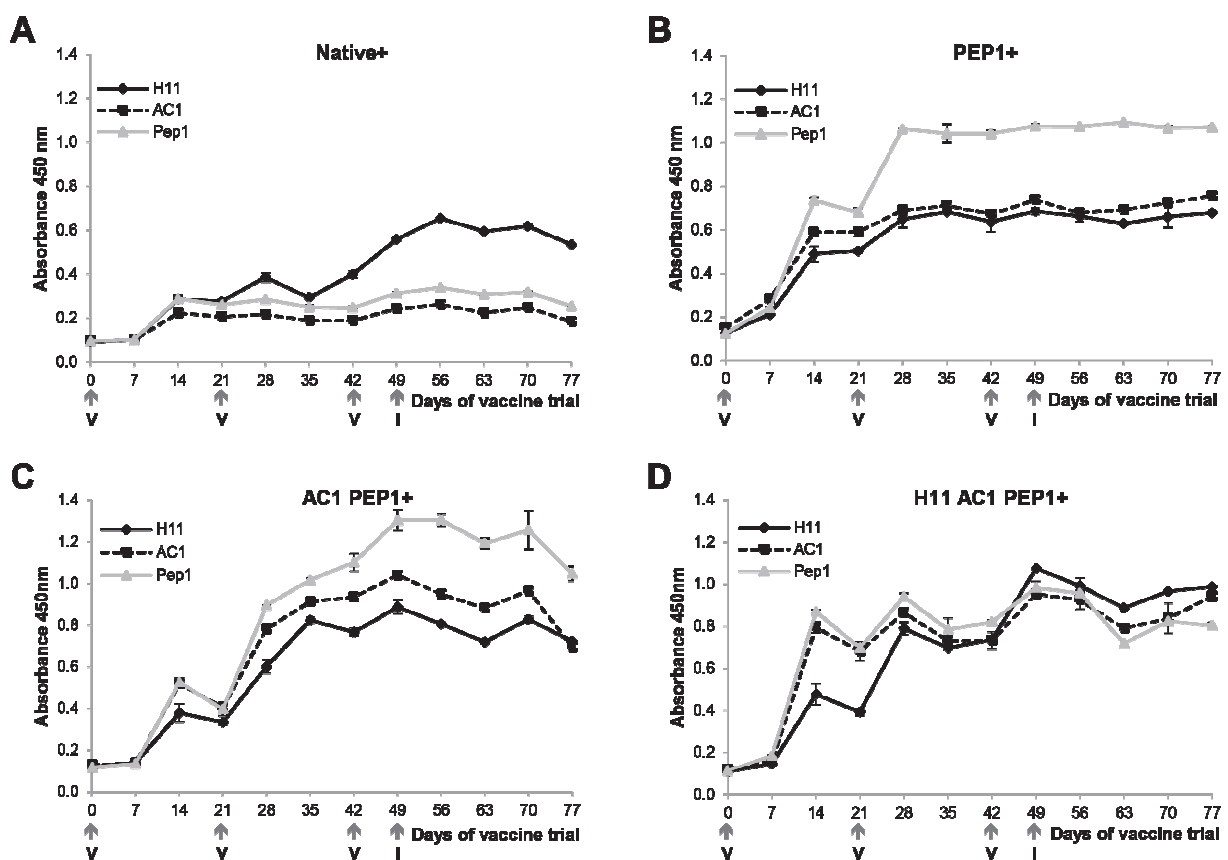


Figure 5: Protein specific immune responses in sheep

ELISA OD₄₅₀ values of sheep IgG antibody reactivity to recombinant proteins H11 (black), AC1 (black dashed) or PEP1 (grey) following immunization with native H11 extract proteins (A) or PEP1+ (B), AC1 PEP1+ (C) or H11 AC1 PEP1+ (D). Blood was sampled at weekly intervals. Vaccinations were done at three weekly intervals (V) before oral infection with *H. contortus* (I). Representative immune responses from single sheep antiserum are shown. Error bars represent the standard deviation of duplicate measurements.

Antibody responses of sheep vaccinated with recombinant or native H11 extract proteins were further examined by immunoblots to identify qualitative differences (Figure 6). Non-protective sheep serum derived from animals vaccinated with the three recombinant proteins H11, AC1 and PEP1 (HAP+) recognized various proteins in the native H11-enriched extract, potentially also native H11 (Figure 6A). However, protective sheep serum derived from animals vaccinated with this native H11 extract detected even more proteins and with at least 10 times greater intensity. The detection of more than three proteins by the non-protective sheep serum could be explained with the potential presence of anti-glycan antibodies identifying the glycan modification of the extract proteins. Indeed, recognition of some native H11 extract proteins by the non-protective serum was lost when the native protein glycans were cleaved by *N*-Glycosidase F (PNGase F) prior to separation on a gradient gel and immunoblotting (Figure 6B). In contrast, the protective serum recognized all the proteins in the native H11 extract even after PNGase F digestion, either by antibodies binding to the proteins themselves or by anti-glycan antibodies recognizing epitopes that are not cleaved by PNGase F (e.g. *N*-glycans with a core α 1,3 fucose). For an effective response, antibodies against these protein antigens and/or antigenic glycans recognized by the protective sheep serum might be required.

To study whether the immunization with recombinant proteins evoked anti-glycan antibodies targeting specifically the nematode-like glycan epitopes on the recombinant proteins we expressed the yeast protein PDI1 (Protein disulfide isomerase 1) with nematode-like glycans in glycoengineered insect cells co-expressing the GALT-1 (Figure 6C). Secreted PDI1 proteins are decorated with both nematode glycan epitopes, the α 1,3 Fuc and the GalFuc epitope [29]. His-tagged PDI1 was purified from insect cell supernatant as described previously and further treated with PreScission protease to enable the cleavage of the His-tag. Similarly, His- and un-tagged PDI1 carrying nematode glycans are recognized by sheep serum derived from animals vaccinated with native H11 extract and even to a higher extent by sheep serum derived from animals vaccinated with recombinant proteins (Figure 6C, lanes 1 and 2). In contrast, the protein PDI1 bearing high-mannose glycans (lane 3) or without *N*-glycans was not detected by any vaccinated nor non-vaccinated sheep serum (Figure 6C, lanes 3 and 4). Thus, vaccination with recombinant proteins or native H11 extract induces a substantial antibody response to nematode-like glycan epitopes.

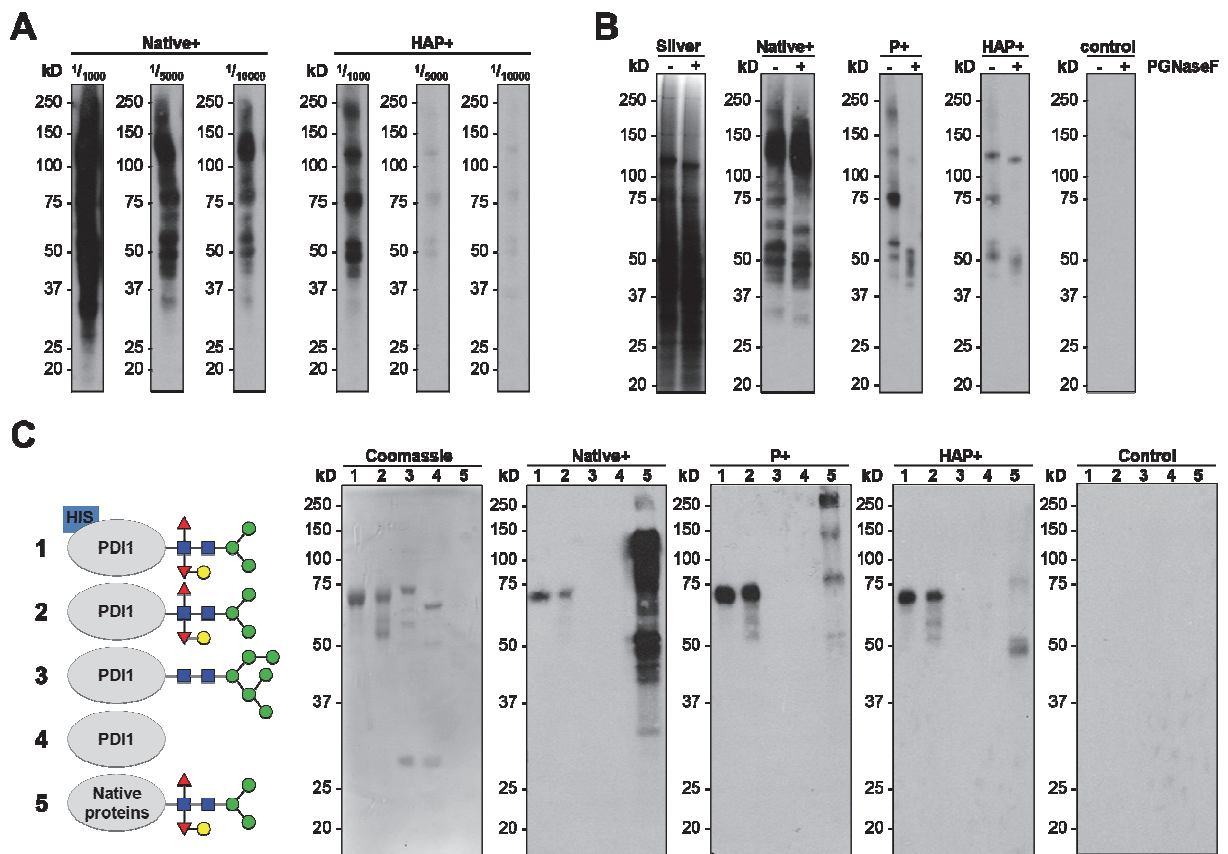


Figure 6: Nematode glycan-specific antibodies in sheep

(A) Native H11 extract was separated by SDS-PAGE, transferred onto a PVDF membrane and probed with different dilutions (1:1000, 1:5000 and 1:10'000) of pooled sheep serum (day 56) vaccinated with either native H11 extract (Native+) or the three recombinant proteins H11, AC1, and PEP1 (HAP+). (B) Native H11 extract was treated either with (+) or without (-) PNGase F prior to separation by SDS-PAGE. The proteins were either silver stained or immunoblotted onto PVDF membrane and probed with different sheep serum (Native+ 1:5000 dilution; P+, HAP+, control 1:1000 dilution) (C) Left: Schematic representation of the different versions of yeast PDI1 proteins (1-4) and native H11 extract proteins (5) were tested for recognition by sheep sera. HIS-PDI1 with nematode-like glycans (1) was co-expressed with GALT-1 and affinity purified from the insect cell culture supernatant, before cleavage of the HIS-tag with PreScission protease (2). ER-retained PDI1 with high-mannose glycans was affinity purified from High five cells. The tag was removed by PreScission protease cleavage (3) and high-mannose glycans were cleaved by PNGaseF digest (4). PDI1 proteins and native H11 extract were separated by SDS-PAGE and either stained with Coomassie or transferred onto a nitrocellulose membrane and probed with pooled sheep serum (day 56, 1:1000 dilution). Blue squares symbolize *N*-acetylglucosamine, red triangles fucose, green circles mannose and yellow circles galactose.

Discussion

The development of an effective recombinant vaccine against the sheep parasite *H. contortus* would be the best solution to effectively control the disease and to combat the anthelmintic resistance problem. Over the last decades, numerous protective native antigens have been identified, but so far all attempts of producing recombinant versions have failed or induced only partial protection against *H. contortus* challenge. High five cells offer a fast and flexible expression system for secreting high amounts of proteins carrying *N*-linked glycans, and allow additional manipulation of the *N*-glycosylation pathway to engineer desired *N*-glycans, such as complex mammalian-type [32]. In this study, we tested the suitability of glycoengineered High five cells as an expression system for the production of recombinant proteins decorated with nematode-like glycans as a vaccine against *H. contortus*. Our experimental system allowed us to characterize the immune response of sheep against protein- and glycoepitopes in more detail.

To date, the most effective native vaccine antigens that protect against *H. contortus* challenge are purified via lectin columns specifically enriching for glycoproteins. Glycans are good immune stimulators and many parasitic modifications are unique to parasites [33]. Also, *H. contortus* possesses a range of unusual *N*-glycan modifications absent from mammals: fucosylated LacdiNAc (LDNF) and trifucosylated cores including the antigenic core α 1,3 fucose, galactosylated fucose (proximal α 1,6 fucose and distal α 1,3 fucose), galactosylated GalNAc and phosphorylcholine [23]. Here we showed by HPLC and mass spectrometry, that glycoengineered High five cells are able to express *H. contortus* proteins decorated with two nematode-like epitopes, the galactosylated core α 1,6 fucose and the immunogenic core α 1,3 fucose. However, other *H. contortus* glycan modifications could not be found. There is potential to further manipulate the *N*-glycosylation pathway in insect cells to produce additional *H. contortus* modifications by overexpressing further glycosyltransferases, e.g. the *C. elegans* fucosyltransferase FUT-6 [34], to produce trifucosylated core *N*-glycans after the removal of the α 1,6 mannose. The fact that proteins are easily produced in high quantities in glycoengineered High five cells and decorated by nematode-like *N*-glycans shows that they are a suitable heterologous system for expressing *H. contortus* vaccine candidates in the future.

The immunogenicity of the glycans and the sensitivity of nematodes towards specific carbohydrate-binding proteins (lectins) lead to the hypothesis that lectins or anti-glycan antibodies might be effective anthelmintics [25, 27]. Moreover, there is evidence that a significant part of the host immune response targets carbohydrate epitopes of the parasite [35]. For example, IgG antibodies from lambs vaccinated with native H-gal-GP complex recognize the LDNF epitope [36, 37] and IgE antibodies bind the core α 1,3 fucose [17]. In this study, we showed by ELISA and immunoblotting that a significant amount of the sheep immune response vaccinated with recombinant proteins is directed against the nematode-like glycan structures. They specifically recognize the galactosylated α 1,6 fucose and/or the core α 1,3 fucose, but not high-mannose structures. This finding is consistent with the response found in sheep vaccinated with native H11 extract, although the binding to these epitopes by the anti-glycan antibodies is lower. Despite the significant anti-glycan response stimulated in sheep vaccinated with recombinant glycoengineered proteins, no protection was observed, suggesting that these glycans might contribute to the antigenicity but do not induce protective immunity. Alternatively, other protective glycan structures might still be missing on the recombinant proteins.

In the vaccination trial of this study, we evaluated the efficacy of three different *H. contortus* proteins, H11, AC1 and PEP1, all shown to be a component of protective native vaccines. These proteases are derived from different detergent-soluble extracts of the adult parasite, namely H11 is enriched in a ConA binding fraction, AC1 in a thiol-binding fraction, and PEP1 is part of the glycoprotein complex H-gal-GP in a galactose-binding fraction. Here, we tested whether combining these diverse digestive proteases in a multivalent vaccine could enhance the protective features of each protease.

Earlier studies failed to express PEP1 as a soluble recombinant protein in insect cells and the *E. coli* expressed and refolded PEP1 was inactive [14]. Here, we report that MBP-fused PEP1 was successfully purified from the insect cell culture medium. PEP1 was enzymatically active between pH 3.4 and 4 (Sabotič, unpublished data), suggesting that the protein is properly folded. Despite the activity of PEP1, it did not provide protection in the sheep vaccination trial. Our explanation is that other proteins are missing in the recombinant PEP1 vaccine for presenting all native conformational epitopes necessary for the induction of protective immunity. For example, including the metalloprotease MEP3 in the vaccine might be necessary, as the three-dimensional structure of the protective H-gal-GP complex depicts an arch conformation formed by

two PEP1 proteins over a base laid by multiple MEP3 proteins [38]. In addition, multiple galectins, a cystatin and a thrombospondin are part of the 1 MDa complex and all are proposed to aid in the digestion of the blood meal [39, 40]. Conformational epitopes of the glycoprotein complex are indeed important to induce full protection, as the H-gal-GP loses its protective capacity when it is reduced by SDS and DTT, but not when it is dissociated by SDS alone [41]. Similar to H-gal-GP, correct conformation of the native aminopeptidase H11 is also essential for protection, as efficacy is reduced when H11 is denatured [10]. To date, only one study could show partial protection using baculovirus-produced H11 in a sheep immunization trial (30 % reduction in worm burden) [13]. It was suggested that the use of a commercial adjuvant instead of the baculovirus extract, might further improve protection levels. Here we used QuilA and Rehydragel as adjuvant for the immunization with H11, AC1 and PEP1, but no protection and even a lower reduction in worm burden (13.5 %) was observed. This was not due to the lack of immune response stimulated by the recombinant vaccines, as all immunized animals showed strong antibody responses against the respective proteins they were vaccinated with, and also displayed cross-reaction with the counterpart proteins (e.g. serum of sheep vaccinated with native H11 extract recognized recombinant proteins and vice versa). In contrast to previous studies [13-15], the sheep immune response towards the recombinant proteins was not short lived and stayed stable for at least 5 weeks after the last immunization. Thus this does not explain the absence of protection. But the protective response might include other Ig isotypes, which were not tested here. Remarkable differences in the recognition of the native *H. contortus* proteins by the protective or non-protective sheep serum were encountered by immunoblot analysis. Not only, did the protective sheep serum recognize all proteins of the native H11 extract before and after PNGase F digestion, but also with at least 10 times greater intensity. This reflects that the native H11 extract is a complex mixture containing different isoforms of aminopeptidases and many other proteins and induces higher levels of antibody responses providing sufficient protection.

We suggest here, that nematode-like *N*-glycans might be important for the proper folding of the recombinantly produced proteins, but it is mainly due to the lack of other proteins that recombinant versions are not protective so far. Therefore, further antigens need to be included in a multivalent vaccine to mimic the overall complexity of the native vaccine and to induce full protective immunity against *H. contortus* infection.

Methods

Construction of plasmids for baculovirus protein expression

Plasmids were generated using standard cloning protocols and are listed in Supplementary Table 4. Open reading frames of all plasmids constructed in this study were confirmed by nucleotide sequencing.

pUC57-H11, pUC57-AC1 and pUC57-PEP1: *Haemonchus contortus* genes *H11* (GenBank X94187.1), *AC1* (GenBank M31112.1) and *PEP1* (GenBankAF079402.1) were codon optimized for expression in insect cells and synthesized at Genscript in a pUC57 plasmid with inserted *Xba*I / *Xho*I restriction sites (*H11*) or *Sac*I / *Xho*I restriction sites (*AC1* and *PEP1*).

pSF10, pSF11 and pSF14: These plasmids were constructed to allow for expression and secretion of N-terminally His-tagged *Haemonchus contortus* digestive proteases *H11* (pSF10), *AC1* (pSF11), and *PEP1* (pSF14). The coding sequences of pUC57-*H11*, pUC57-*AC1* and pUC57-*PEP1* were either excised by *Sac*I and *Xho*I (*H11*) after PCR-amplification using primers H11-TM-Fw*Stu*I*Sac*I (5'-AAA AAA GGC CTG AGC TCT ACT ACT TCA CAA GGA AGG CTT TCG-3') and H11-TM-Rev*Xho*Inew (5'-AAA AAA CTC GAG TTA CAG GGT AGC CTT CTT GAA GAA AG-3') or cut directly by *Sac*I and *Xho*I (*AC1* and *PEP1*). Fragments were ligated into the respective restriction sites of plasmid pRG105 downstream of the PreScission protease cleavage site and thus replacing the *PDI1* gene.

pSF12 and pSF15: Cloning of an additional C-terminal His₁₀-tag to improve expression and purification of secretory *AC1* (pSF12) and *PEP1* constructs (pSF15). The 3'-region of the genes was PCR-amplified using plasmid pSF11 (*AC1*) or pSF14 (*PEP1*) as template and the following primer pairs: *AC1*-FwHis_*Xba*I (5'-AAA AAA ATC TAG AAA TGA AGT ACC TGG TTC TCG CC-3') and *AC1*-RvHis_*Xho*I (5'-AAA AAA ACT CGA GTC AGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGA GGG ATT CTG TGT CCA CGA TA-3') for pSF11; *PEP1*-FwHis_*Xba*I (5'-AAA AAA ATC TAG AAA TGC TGT ACT TGC TCC TCC TC-3') and *PEP1*-Rvhis_*Xho*I (5'-AAA AAA ACT CGA GTC AGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GCT TAG GTT CAG CGA AGC CGAT-3') for pSF14. The respective PCR products were either cloned into the *A*l*l*I and *Xho*I restriction sites of pSF11 or into the *Stu*I and *Xho*I restriction sites of pSF14, replacing the 3'-region of the genes and adding a C-terminal His₁₀-tag.

pSF22: Substitution of the N-terminal His-tag by a maltose binding protein (MBP)-tag to improve expression and purification of secretory PEP1 (C-terminally His-tagged). The coding sequence was subcloned as *SacI*-*XhoI* fragment from pSF15 into plasmid pRG132 downstream of the PreScission protease cleavage site.

Expression in insect cells

Trichoplusia ni "High Five" and *Spodoptera frugiperda* Sf21 cells were obtained from Prof. K. Locher (ETH Zürich, Switzerland) and cultivated in Sf-900 II SFM medium (Invitrogen # 10902104) without fetal calf serum at 27 °C in shaker flasks. Recombinant bacmids were generated using DH10Bac *E. coli* cells and viruses were produced in Sf21 cells according to standard procedures (Invitrogen).

Proteins were expressed in High Five cells by diluting the culture to 1 million cells per ml and either single or co-infected 1:100 (v:v) with each recombinant viruses. Expression levels were detected 72 hours post infection as described in [28]. Briefly, 1 million cells were pelleted by centrifugation (600 x g, 5 min), washed with PBS and lysed in 125 µL lysis buffer (PBS, 1 % Triton X-100, 1x Protease Inhibitor cocktail (Roche #11873580001)) by rotating at 4 °C for 10 min. The whole cell lysates were centrifuged (16'000 x g, 5 min), and the supernatant was analyzed by immunoblotting. For secreted proteins, the culture supernatants of cells were harvested 72 hours post infection by centrifugation at 3800 x g for 10 min and filtered through a PES-membrane 0.22 µM filter (Techno Plastic Products).

Protein purification

Affinity purification was performed on gravity flow columns filled with 1 ml Ni-NTA agarose beads (Protino #745400.100) equilibrated with 10 column volumes (CV) of Sf900 II SFM medium. The culture supernatant was loaded and the resin was washed afterwards with 10 to 15 CV washing buffer (10 mM imidazole in PBS, pH 7.4). Proteins were eluted four times with 1 CV of elution buffer (250 mM imidazole in PBS, pH 7.4). Elution fractions were concentrated and buffer exchanged to PBS on Amicon Ultra-4 Centrifugal Filter Devices (Millipore). Purification of GST-tagged ER retained-Pdi1p and secreted His-tagged Pdi1p and subsequent tag cleavage using PreScission protease (VWR #27-0843-01) were described in (Chapter 4, [29]). Native H11 extract

was prepared from *H. contortus* adult worms by solubilization of a PBS homogenate pellet in 2 % Triton X-100 (Sigma-Aldrich #T9284) and purified using Concanavalin A Sepharose (Sigma-Aldrich #C9017) according to patent WO1990011086 A1 by Munn and Smith [42].

SDS-PAGE and Immunoblotting

Protein samples were boiled at 95 °C for 5 min in reducing sample buffer (62.5 mM Tris-HCl pH 6.8, 2 % SDS (v/w), 5 % β -mercaptoethanol (v/v), 10 % glycerol (v/v), 0.01 % bromophenol blue (w/v)) and separated by SDS-PAGE (10 % acrylamide, 120 Volt, 150 min or 4–12 % Bis-Tris gradient gels BioRad #3450125, XT MOPS running buffer, 200 Volt, 55 min). Proteins were visualized with Coomassie blue or silver staining or immunoblotting on PVDF (Merck Millipore #SEQ00010) or nitrocellulose (GE Healthcare #10600002). Blots were blocked with 5 % milk in PBST (PBS, 0.1 % Tween-20) and probed with sheep antiserum to different proteins (1:1000 - 1:10'000 dilutions in PBST with 5 % milk), followed by bovine anti-sheep IgG-HRP (Santa Cruz #sc-2701; 1:5000 dilution in PBST with 5 % milk). After extensive washing with PBST, the membranes were developed using ECL detection solution (GE Healthcare #RPN2105) and exposure to photographic films (Super RX-N Fujifilm #57164152).

Protein identification by mass spectrometry

Major bands were excised from Coomassie-stained SDS-PAGE gels of purified recombinant protein sample. Standard mass spectrometry procedure was performed with in-gel trypsin digest, extraction of trypsin digested peptides and fractionation by LC-MS/MS. Samples were analyzed on a calibrated LTQ-Orbitrap Velos mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to an Eksigent-Nano-HPLC system (Eksigent Technologies, Dublin (CA), USA). Peptides were resuspended in 2.5% acetonitrile and 0.1% formic acid and loaded on a self-made tip column (75 μ m \times 80 mm) packed with reverse phase C18 material (AQ, 3 μ m 200 Å, Bischoff GmbH, Leonberg, Germany) and eluted with a flow rate of 200 nl per min by a gradient from 3 to 30% of B in 22 min, 50% B in 25min, 97% B in 27 min. One scan cycle comprised of a full scan MS survey spectrum, followed by up to 20 sequential CID MS/MS on the most intense signals above a threshold of 1500. Full-scan MS

spectra (300–2000 m/z) were acquired in the FT-Orbitrap at a resolution of 60,000 at 400 m/z, while CID MS/MS spectra were recorded in the linear ion trap. CID was performed with a target value of 1e4 in the linear trap, collision energy at 35 V, Q value at 0.25 and activation time at 30 ms. AGC target values were 5e5 for full FTMS scans and 1e4 for ion trap MSn scans. For all experiments, dynamic exclusion was used with 1 repeat count, 15 s repeat duration, and 60 s exclusion duration. Data was searched against the NCBI and SwissProt (Version 201408) database using the Mascot search engine (Version 2.4).

Sample preparation for glycopeptides

Purified proteins (usually 50-200 µg) were reduced by 50 mM dithiothreitol in 50 mM ammonium bicarbonate buffer (AmBic buffer, pH 8.5) at 37°C for 1 hour on a filter device (Microcon YM-30, Millipore) and subsequently alkylated by 65 mM iodoacetamide in 50 mM AmBic buffer at 37°C in the dark for 1 hour. After extensive washing of the filter with AmBic buffer, the proteins were digested by trypsin (molar ratio 60:1, Promega #V5111) at 37°C for 16 hours. Trypsin digested peptides and glycopeptides were collected by centrifugation and vacuum-dried in a speedvac. Samples were either desalted by Zip-Tip C18 (Millipore #ZTC18S960) prior to glycosite-specific nanoHPLC-HCD-MS/MS analysis or further processed for HPLC analysis.

Glycosite-specific *N*-glycosylation analysis by NanoHPLC-HCD-Mass Spectrometry

For a glycosite-specific *N*-glycan analysis of intact glycopeptides the nanoHPLC-HCD-MS/MS method was used as previously described (Chapter 4, [29]). A calibrated LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific) coupled to an Eksigent-NanoHPLC system (Eksigent Technologies) was used for sample analysis. Peptides were resuspended in 2.5% acetonitrile (ACN) with 0.1% formic acid (FA) and loaded on a self-made tip column (75 µm × 80 mm) packed with reverse phase C18 material (AQ, 3 µm 200 Å, Bischoff GmbH) and eluted by a gradient (from 3 to 30% of solution B (99.9% ACN, 0.1% FA) for 22 min, 50% of B for 25 min, 97% of B for 27 min, with a flow rate of 200 nl/min). One scan cycle included a full scan MS survey spectrum,

followed by up to 10 sequential HCD MS/MS on the most intense signals (>2,000). Full-scan MS spectra (700–2,000 m/z) and HCD MS/MS spectra were recorded in the FT-Orbitrap (resolution of 60,000 at 400 m/z for MS and 15,000 at 400 m/z for MS/MS). HCD was performed with a target value of 1e5 and stepped collision energy rolling from 35, 40 and 45 V was applied. For full Fourier Transform MS, AGC target values were 5e5. Dynamic exclusion with a single repeat count, 15 s repeat duration, and 60 s exclusion duration was used for all experiments.

MS and MS/MS data were processed into the Mascot generic format (mgf) files and searched against the SwissProt database (Version 201408) database using the Mascot search engine (Version 2.4) with the consideration of carbamidomethylation at cysteine and oxidation at methionine. The monoisotopic masses of charged peptides (2+ or more) were searched with a peptide tolerance of 10 ppm and a MS/MS tolerance of 0.25 Da for fragment ions. XCalibur (Version 2.2 sp1.48) was used for glycosite-specific *N*-glycosylation analysis and all data were inspected manually. First extractmgf was written in Perl and then glycopeptides were extracted from mgf file by identifying glycan oxonium ions ([HexNAc]⁺ 204.09 and [HexNAc+Hex]⁺ 366.12) and glycopeptides specific for each glycosite by sorting out the Y1 ions individually from extracted mgf files. For quantification, extracted ion chromatography (XIC) of each glycoform was plotted by its individual m/z with the mass tolerance of 10 ppm. Peak area was defined manually and integrated by XCalibur. The relative amount of each glycoform sharing the same peptide backbone was calculated with the following equation: Relative amount of each glycoform (%) = $\frac{\text{Peak area of each glycoform}}{\text{Sum of peak area of all glycoforms}} \times 100\%$

***N*-glycan analysis by high pressure liquid chromatography (HPLC)**

N-glycans were enzymatically released from trypsin digested glycopeptides with 1 μL of PNGase A (Roche #11642995001) in 100 μL of 50 mM sodium acetate buffer (pH 5) at 37°C for 16 hours. Released glycans were supplemented with ACN to a 2 % final concentration and purified using prepacked C18 Sep-Pak columns (Waters #WAT051910) and self-made fritted carbon columns (Supelclean ENVI-carbon 120/400, Sigma-Aldrich #57210-U). The glycans were eluted from the latter column in 25 % ACN and vacuum-dried. *N*-glycan samples were fluorescently labeled with 2-aminobenzamide (2-AB, Sigma #A8980-4) by incubation in 25 μL of labeling solution (1 M 2-AB, 70 % DMSO, 30 % acetic acid, 0.35 M cyanoborohydride (Sigma-Aldrich

#71435) at 65 °C for 2 hours. Labelled glycans were purified over 4 layers Whatman filter paper in 95 % ACN using 100 µL deionized water for elution and vacuum-dried in a speedvac. Then, they were dissolved in acetonitrile:H₂O (75:25; v:v) and filtered (0.45 µm, Millipore #UFC30HVNB) before loading on the HPLC.

2-AB labelled glycans were separated using a linear gradient (from acetonitrile:H₂O (75:25; v:v) to acetonitrile:H₂O (50:50; v:v) for 90 min with a flow rate of 0.8 ml/min) on a normal-phase column (Supelcosil LC-NH₂, Sigma #58338 with a Supelguard Cartridge Sigma #59568) and fluorescence detection (λ_{ex} =330 nm, λ_{em} =420) at room temperature. The Dionex HPLC system consisted of a LC20 chromatography enclosure, one GP40 gradient pump, an autosampling device and a Jasco FP920 fluorescence detector controlled by a personal computer using Chromeleon 6.0 software. Two standards were used for calibration and pre-assignment of *N*-glycan retention times, a purchased standard of 2-AB labelled GlcNAc₂Man₃Fuc (Prozyme #GKSB-102) and a mannose standard ranging from GlcNAc₂Man₅ to GlcNAc₂Man₉ isolated from RNase B (Sigma). All HPLC elution fractions were collected and analyzed by MALDI-TOF MS and MS/MS to identify the composition of the *N*-glycans in the chromatogram. Peak areas were quantified using the Chromeleon program (Dionex).

MALDI-TOF Mass Spectrometry for 2-AB labelled glycans

2-AB labelled glycans, collected from the HPLC and vacuum-dried, were mixed 1:1 with dihydroxybenzoic acid matrix (DHB, 10 mg/mL in 75 % ACN with 0.1 % formic acid) and analyzed by MALDI-TOF mass spectrometry in positive ion mode on a MALDI-TOF/TOF machine (SCIEX 5800). A 20 pmole permethylated Lewis B standard was used for external calibration. All *N*-glycan structures with given *m/z* values were further analyzed by MALDI-TOF MS/MS. The spectra were assigned manually and annotated with the GlycoWorkbench software (version 1).

Sheep vaccine trial

Lambs (White Mountain breed, approx. 6 months old), which were raised indoors under conditions that minimized helminth infections, were distributed into groups of six, balanced for sex and weight. Intramuscular vaccinations were done three times at three weekly intervals with either 100 µg recombinant protein (100 µg of each protein

and administered separately) or with 30 µg native H11 extract. Additionally, a clostridial vaccine (Covexin 8, Merck Animal Health) was applied at the first and third vaccination in the opposing leg to the vaccinated one. One week after the last vaccination, the sheep were orally challenged with 5000 infective larvae (L3) of *H. contortus*. One control group was orally infected with *H. contortus* but not treated with any vaccine. QuilA (SEPPIC, 2.5 mg / 0.5 mL dose) or QuilA in combination with Rehydragel (Fisher Scientific, 1 mg / 1 mL dose) was used as adjuvant. Blood was sampled at weekly intervals and EPG (eggs per gram of feces) monitored at day 0, 14, 21 and 28 after oral infection and until necropsy at days 28 to 30, when worm burden was counted. All experimental procedures were approved by national veterinary authorities.

Enzyme-linked Immunosorbent Assay (ELISA)

Coating of 96 well plates (TPP #92096) was done by 24 hours incubation at 4 °C with either 0.5 µg (AC1 or PEP1) or 1µg (H11) recombinant protein per well (PBS pH 8 coating buffer). Plates were washed with an excess of washing solution (PBS, pH 8, 0.05 % Tween, 0.05 % Casein) and blocked with blocking solution (PBS, pH 8, 0.05 % Tween, 1 % Casein). After one hour incubation on a shaking platform at room temperature, a 1:1000 dilution of primary sheep serum or a 1:5000 dilution of anti-His antibody (Qiagen #34670) in washing solution was added per well. Plates were incubated one hour on a shaking platform at room temperature and washed 5 times with an excess of washing solution. A 1:1000 dilution of secondary antibody in washing solution was added (either anti-sheep IgG-HRP antibody (Santa Cruz #sc-2701) for sheep serum or goat anti-mouse IgG-HRP (Santa Cruz #sc-2005) for anti-His antibody). Plates were incubated one hour on a shaking platform at room temperature and washed 5 times with an excess of washing solution. Development of the ELISA plates was performed by adding TMB development solution (0.05 M phosphate-citrate buffer, pH 5 containing 0.1 mg/mL 3,3',5,5'-Tetramethylbenzidine dihydrochloride, 10 % DMSO and 0.004 % H₂O₂). The reaction was stopped with an equal amount of stop reagent BioFX (SurModics #STPR-1000-01) and plates were analyzed on a Tecan reader at 450nm.

References

1. Newton, S.E. and E.A. Munn, *The development of vaccines against gastrointestinal nematode parasites, particularly Haemonchus contortus*. Parasitol Today, 1999. **15**(3): p. 116-22.
2. Kotze, A.C. and R.K. Prichard, *Anthelmintic Resistance in Haemonchus contortus: History, Mechanisms and Diagnosis*. Adv Parasitol, 2016. **93**: p. 397-428.
3. Newton, S.E. and E.N. Meeusen, *Progress and new technologies for developing vaccines against gastrointestinal nematode parasites of sheep*. Parasite Immunol, 2003. **25**(5): p. 283-96.
4. Knox, D.P., et al., *The nature and prospects for gut membrane proteins as vaccine candidates for Haemonchus contortus and other ruminant trichostrongyloids*. Int J Parasitol, 2003. **33**(11): p. 1129-37.
5. Munn, E.A., *Rational design of nematode vaccines: hidden antigens*. Int J Parasitol, 1997. **27**(4): p. 359-66.
6. Williamson, A.L., et al., *Digestive proteases of blood-feeding nematodes*. Trends Parasitol, 2003. **19**(9): p. 417-23.
7. Smith, T.S., et al., *Purification and evaluation of the integral membrane protein H11 as a protective antigen against Haemonchus contortus*. Int J Parasitol, 1993. **23**(2): p. 271-80.
8. Smith, W.D., S.K. Smith, and J.M. Murray, *Protection studies with integral membrane fractions of Haemonchus contortus*. Parasite Immunol, 1994. **16**(5): p. 231-41.
9. Knox, D.P., S.K. Smith, and W.D. Smith, *Immunization with an affinity purified protein extract from the adult parasite protects lambs against infection with Haemonchus contortus*. Parasite Immunol, 1999. **21**(4): p. 201-10.
10. Munn, E.A., et al., *Vaccination against Haemonchus contortus with denatured forms of the protective antigen H11*. Parasite Immunol, 1997. **19**(6): p. 243-8.
11. Ekoja, S.E. and W.D. Smith, *Antibodies from sheep immunized against Haemonchus contortus with H-gal-GP inhibit the haemoglobinase activity of this protease complex*. Parasite Immunology, 2010. **32**(11-12): p. 731-738.

12. Nisbet, A.J., et al., *Immunity to Haemonchus contortus and Vaccine Development*. Adv Parasitol, 2016. **93**: p. 353-96.
13. Reszka, N., et al., *Haemonchus contortus: characterization of the baculovirus expressed form of aminopeptidase H11*. Exp Parasitol, 2007. **117**(2): p. 208-13.
14. Cachat, E., et al., *Attempts to immunize sheep against Haemonchus contortus using a cocktail of recombinant proteases derived from the protective antigen, H-gal-GP*. Parasite Immunol, 2010. **32**(6): p. 414-9.
15. Roberts, B., et al., *Novel expression of Haemonchus contortus vaccine candidate aminopeptidase H11 using the free-living nematode Caenorhabditis elegans*. Vet Res, 2013. **44**: p. 111.
16. Haslam, S.M., et al., *Haemonchus contortus glycoproteins contain N-linked oligosaccharides with novel highly fucosylated core structures*. J Biol Chem, 1996. **271**(48): p. 30561-70.
17. van Die, I., et al., *Core alpha1-->3-fucose is a common modification of N-glycans in parasitic helminths and constitutes an important epitope for IgE from Haemonchus contortus infected sheep*. FEBS Lett, 1999. **463**(1-2): p. 189-93.
18. Tretter, V., et al., *Fucose alpha 1,3-linked to the core region of glycoprotein N-glycans creates an important epitope for IgE from honeybee venom allergic individuals*. Int Arch Allergy Immunol, 1993. **102**(3): p. 259-66.
19. Rendic, D., et al., *Adaptation of the "in-gel release method" to N-glycome analysis of low-milligram amounts of material*. Electrophoresis, 2007. **28**(23): p. 4484-92.
20. Schiller, B., et al., *Complicated N-linked glycans in simple organisms*. Biol Chem, 2012. **393**(8): p. 661-73.
21. Altmann, F., et al., *Insect cells as hosts for the expression of recombinant glycoproteins*. Glycoconj J, 1999. **16**(2): p. 109-23.
22. Jarvis, D.L., *Baculovirus-insect cell expression systems*. Methods Enzymol, 2009. **463**: p. 191-222.
23. Paschinger, K. and I.B. Wilson, *Two types of galactosylated fucose motifs are present on N-glycans of Haemonchus contortus*. Glycobiology, 2015. **25**(6): p. 585-90.

24. Yan, S., et al., *Galactosylated fucose epitopes in nematodes: increased expression in a Caenorhabditis mutant associated with altered lectin sensitivity and occurrence in parasitic species*. J Biol Chem, 2012. **287**(34): p. 28276-90.
25. Butschi, A., et al., *Caenorhabditis elegans N-glycan core beta-galactoside confers sensitivity towards nematotoxic fungal galectin CGL2*. PLoS Pathog, 2010. **6**(1): p. e1000717.
26. Schubert, M., et al., *Plasticity of the beta-trefoil protein fold in the recognition and control of invertebrate predators and parasites by a fungal defence system*. PLoS Pathog, 2012. **8**(5): p. e1002706.
27. Heim, C., et al., *Inhibition of Haemonchus contortus larval development by fungal lectins*. Parasit Vectors, 2015. **8**: p. 425.
28. Titz, A., et al., *Molecular basis for galactosylation of core fucose residues in invertebrates: identification of caenorhabditis elegans N-glycan core alpha1,6-fucoside beta1,4-galactosyltransferase GALT-1 as a member of a novel glycosyltransferase family*. J Biol Chem, 2009. **284**(52): p. 36223-33.
29. Hang, I., et al., *Analysis of site-specific N-glycan remodeling in the endoplasmic reticulum and the Golgi*. Glycobiology, 2015. **25**(12): p. 1335-49.
30. Hsu, T.A., et al., *Differential N-glycan patterns of secreted and intracellular IgG produced in Trichoplusia ni cells*. J Biol Chem, 1997. **272**(14): p. 9062-70.
31. Nallamsetty, S. and D.S. Waugh, *Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners*. Protein Expr Purif, 2006. **45**(1): p. 175-82.
32. Palmberger, D., et al., *SweetBac: a new approach for the production of mammalianised glycoproteins in insect cells*. PLoS One, 2012. **7**(4): p. e34226.
33. Nyame, A.K., Z.S. Kwarar, and R.D. Cummings, *Antigenic glycans in parasitic infections: implications for vaccines and diagnostics*. Arch Biochem Biophys, 2004. **426**(2): p. 182-200.
34. Yan, S., et al., *Array-assisted characterization of a fucosyltransferase required for the biosynthesis of complex core modifications of nematode N-glycans*. J Biol Chem, 2013. **288**(29): p. 21015-28.

35. Schallig, H.D. and M.A. van Leeuwen, *Carbohydrate epitopes on Haemonchus contortus antigens*. Parasitol Res, 1996. **82**(1): p. 38-42.
36. Geldhof, P., et al., *Presence of the LDNF glycan on the host-protective H-gal-GP fraction from Haemonchus contortus*. Parasite Immunol, 2005. **27**(1-2): p. 55-60.
37. Vervelde, L., et al., *Vaccination-induced protection of lambs against the parasitic nematode Haemonchus contortus correlates with high IgG antibody responses to the LDNF glycan antigen*. Glycobiology, 2003. **13**(11): p. 795-804.
38. Muench, S., et al., *3D Reconstruction of a Large Protease Complex from Haemonchus contortus*. Biophysical Journal, 2008. **94**(2, Supplement): p. 536-539.
39. Skuce, P.J., et al., *Cloning and characterisation of thrombospondin, a novel multidomain glycoprotein found in association with a host protective gut extract from Haemonchus contortus*. Mol Biochem Parasitol, 2001. **117**(2): p. 241-4.
40. Newlands, G.F., et al., *Cloning and characterization of a beta-galactoside-binding protein (galectin) from the gut of the gastrointestinal nematode parasite Haemonchus contortus*. Parasitology, 1999. **119 (Pt 5)**: p. 483-90.
41. Smith, S.K. and W.D. Smith, *Immunisation of sheep with an integral membrane glycoprotein complex of Haemonchus contortus and with its major polypeptide components*. Res Vet Sci, 1996. **60**(1): p. 1-6.
42. Munn, E.A. and T.S. Smith, *Production and use of anthelmintic agents and protective immunogens*. 1990, Google Patents.

Supplementary Data

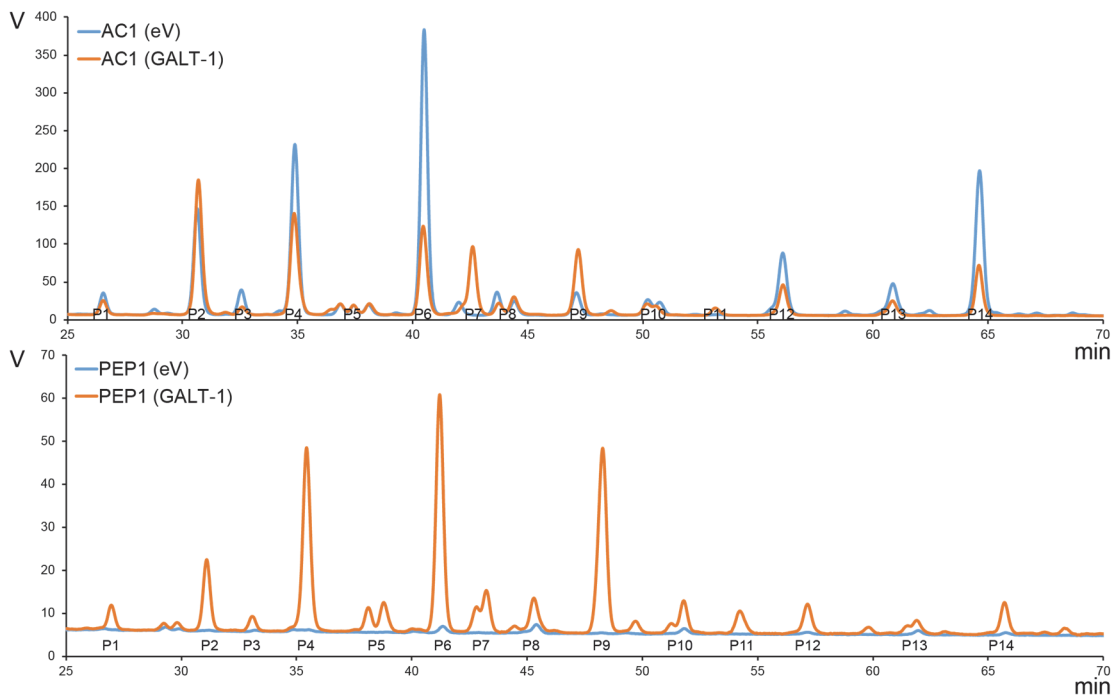


Figure S1: Glycoengineered High five-expressed AC1 and PEP1 are decorated with nematode-like N-glycans

Overall *N*-glycan HPLC profiles of recombinant AC1 (top) and PEP1 (bottom) purified from insect cells co-expressing either an empty vector control (eV, blue) or the *C. elegans* galactosyltransferase-1 (GALT-1, orange). Purified proteins were reduced, alkylated, and proteolyzed with trypsin. The oligosaccharides were 2-AB labelled after their release with PNGase A, and separated by HPLC using a NP-amino column and fluorescence detection. Peaks were collected and analyzed by MALDI-TOF MS and MS/MS. The chromatography shows the relative amount of each glycoform.

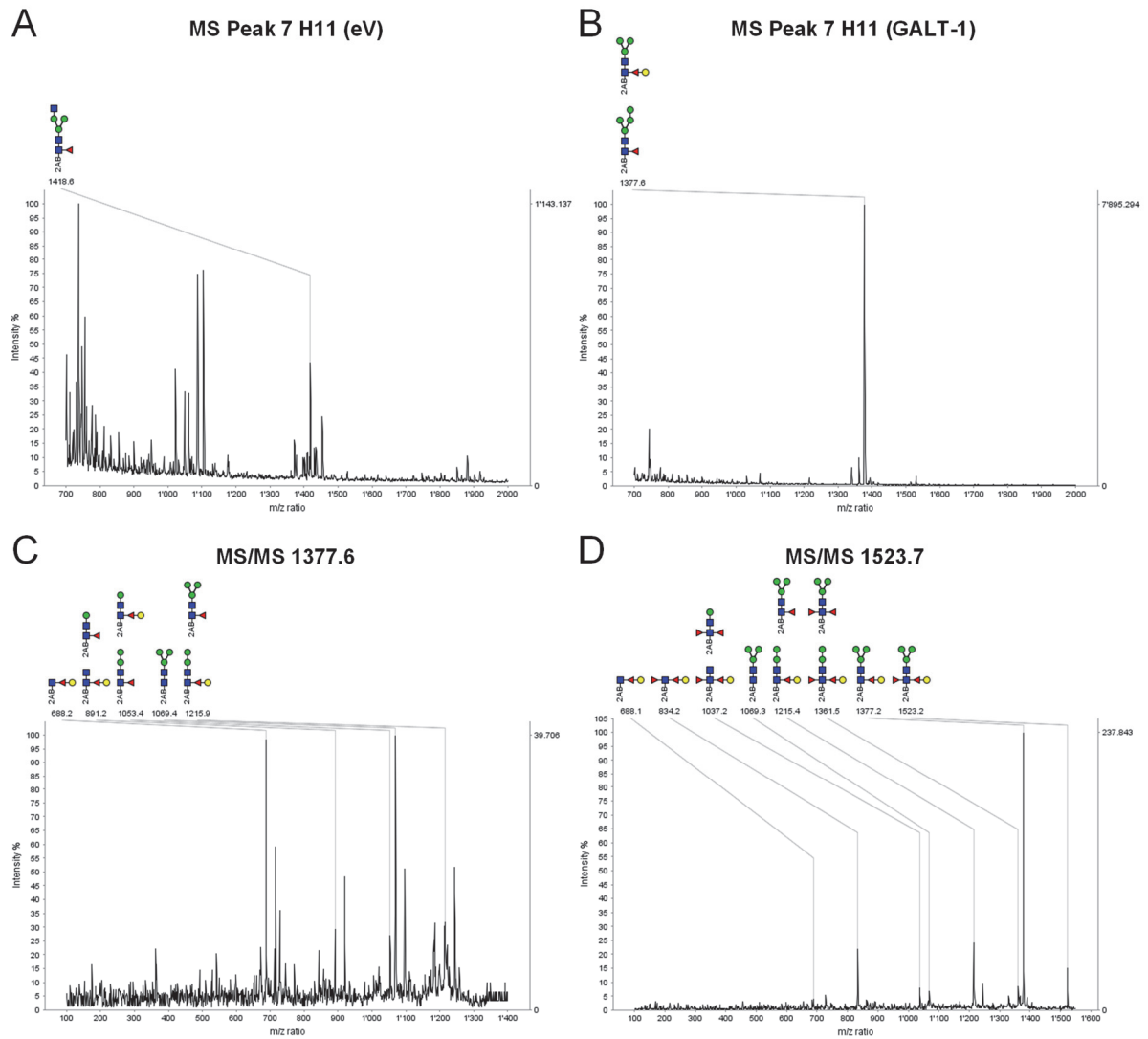


Figure S2: MALDI-TOF MS and MS/MS of 2-AB labelled *N*-glycans collected from HPLC peaks
(A) MALDI-TOF MS spectrum of Peak 7 collected from HPLC chromatogram of H11 (eV) shown in Figure 2. **(B)** MALDI-TOF MS spectrum of Peak 7 collected from HPLC chromatogram of H11 (GALT-1) shown in Figure 2. **(C)** MALDI-TOF MS/MS of molecular ion m/z 1377.6 detected in MS spectrum of H11 (GALT-1) HPLC Peak 7 in spectrum B. **(D)** MALDI-TOF MS/MS of molecular ion m/z 1523.7 detected in MS spectrum of H11 (GALT-1) HPLC Peak 9 in Figure 2. All molecular ions are $[M + K]^+$. The spectra were assigned manually and annotated with the GlycoWorkbench software.

Table S1: Quantification of the N-glycan HPLC profiles of recombinant H11, AC1, and PEP1 purified from insect cells co-expressing either an empty vector control (EV) or the galactosyltransferase-1 (GALT-1).

Quantification of the peak areas was performed using the Chromeleon software. Peak area ratios were calculated as the relative amount of each glycoform. Peak composition, listed in the glycan structures column, was analyzed by MALDI-TOF MS and MS/MS.

Peak	Peak area						Peak area ratio						Glycan structures
	H11		PEP1		AC1		H11		PEP1		AC1		
	EV	GALT-1	EV	GALT-1	EV	GALT-1	EV	GALT-1	EV	GALT-1	EV	GALT-1	
P1	155.0	40.8	10.1	6.9	0.1	2.1	4.8	1.3	1.9	1.9	3.0	2.1	HexNAc ₂ Hex ₂ Fuc
P2	337.8	310.4	53.7	69.2	0.0	6.1	10.4	9.6	10.1	19.0	1.2	6.2	HexNAc ₂ Hex ₃
P3	29.4	8.5	11.8	4.5	0.1	1.2	0.9	0.3	2.2	1.2	2.3	1.2	HexNAc ₂ Hex ₂ Fuc ₂
P4	235.2	111.5	90.5	57.2	0.4	17.4	7.3	3.4	17.0	15.7	10.9	17.6	(HexNAc ₂ Hex ₃)Fuc
P5	276.6	285.7	15.6	16.8	0.1	5.0	8.5	8.8	2.9	4.6	1.4	5.1	(HexNAc ₂ Hex ₃)HexNAc
P6	421.1	148.5	139.2	46.1	0.5	21.5	13.0	4.6	26.2	12.7	14.9	21.7	(HexNAc ₂ Hex ₃)Fuc ₂
P7	28.7	124.9	7.3	38.9	0.0	5.9	0.9	3.9	1.4	10.7	1.1	6.0	(HexNAc ₂ Hex ₃)HexNAcFuc, (HexNAc ₂ Hex ₃)FucHex
P8	244.5	260.7	19.1	15.6	0.9	4.7	7.5	8.1	3.6	4.3	23.8	4.7	HexNAc ₂ Hex ₄ HexNAc, HexNAc ₂ Hex ₅
P9	22.3	476.9	14.2	39.3	0.3	20.3	0.7	14.7	2.7	10.8	6.9	20.5	(HexNAc ₂ Hex ₃)Fuc ₂ Hex, (HexNAc ₂ Hex ₃)HexNAcFucHex, (HexNAc ₂ Hex ₃)HexNAcFuc ₂
P10	298.3	283.6	16.5	11.3	0.4	3.9	9.2	8.8	3.1	3.1	11.9	3.9	HexNAc ₂ Hex ₅ HexNAc, HexNAc ₂ Hex ₆
P11	1.0	25.2	0.8	4.4	0.0	2.5	0.0	0.8	0.1	1.2	0.1	2.6	(HexNAc ₂ Hex ₃)HexNAcFuc ₂ Hex
P12	600.1	583.9	40.2	17.5	0.2	3.4	18.5	18.0	7.6	4.8	4.8	3.5	HexNAc ₂ Hex ₇
P13	405.9	403.0	29.6	10.1	0.4	1.9	12.5	12.5	5.6	2.8	12.1	2.0	HexNAc ₂ Hex ₈
P14	186.9	171.7	83.0	25.9	0.2	3.1	5.8	5.3	15.6	7.1	5.7	3.2	HexNAc ₂ Hex ₉
							72.5	71.1	48.5	45.8	60.8	28.5	Oligomannose + Hybrid
							12.9	4.7	20.3	17.6	15.0	19.7	Core fucosylation
							14.6	4.9	31.2	13.9	24.2	22.8	Core difucosylation
							0	19.4	0	22.7	0	29.0	Core galactosylation

Table S2: Quantification of glycosite-specific N-glycosylation profiles of recombinant H11, AC1 and PEP1

H11, AC1 and PEP1 were purified from insect cells co-expressing either an empty vector control (EV, left panel per site) or the galactosyltransferase-1 (GALT-1, right panel per site). Trypsin digested glycopeptides were analyzed by nanoHPLC-HCD MS/MS mass spectrometry. H11 and AC1 have four N-glycosylation sites and PEP1 has five (site1-site5). Glycan structures found on each glycosylation site is listed with their respective percentage. Most abundant (>10 %) glycan structures are marked in grey and depicted in Figure 3. nd: not detected

Glycan structure	H11					AC1					PEP1													
	site 1		site 2		site 3		site 4		site 1		site 2		site 3		site 4		site 5							
	EV	GALT-1	EV	GALT-1	EV	GALT-1	EV	GALT-1	EV	GALT-1	EV	GALT-1	EV	GALT-1	EV	GALT-1	EV	GALT-1						
HexNAc2Hex9	0.04	0.08	nd	nd	nd	nd	nd	nd	0.00	0.00	0.69	0.64	nd	nd	nd	nd	0.00	0.00	nd	nd				
HexNAc2Hex8	0.41	0.47	nd	nd	nd	nd	nd	0.00	0.00	0.14	0.14	0.14	nd	nd	nd	nd	0.00	0.01	0.00	nd	nd			
HexNAc2Hex7	0.25	0.16	0.14	0.15	nd	nd	nd	0.21	0.16	0.04	0.02	0.02	nd	nd	nd	nd	0.01	0.03	0.00	0.01	0.04			
HexNAc2Hex6	0.11	0.09	0.04	0.04	nd	nd	nd	0.07	0.04	0.01	0.01	0.01	nd	nd	nd	nd	0.02	0.03	0.00	0.01	0.03			
HexNAc2Hex5	0.01	0.03	0.04	0.04	nd	nd	nd	0.06	0.05	0.00	0.00	0.00	nd	nd	nd	nd	0.01	0.00	0.00	0.01	0.03			
HexNAc2Hex6HexNAc	nd	nd	nd	nd	nd	nd	nd	0.01	0.01	0.00	0.00	0.00	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd		
HexNAc2Hex5HexNAc	nd	nd	nd	nd	nd	nd	nd	0.02	0.02	0.01	0.00	0.00	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd		
HexNAc2Hex4HexNAc	nd	nd	nd	nd	nd	nd	nd	0.02	0.02	0.02	0.02	0.02	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd		
(HexNAc2Hex3)HexNAc	0.10	0.11	0.05	0.01	nd	nd	nd	0.01	0.02	0.00	0.00	0.12	0.04	0.04	0.02	nd	nd	nd	nd	nd	nd	nd		
(HexNAc2Hex3)HexNAc2	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	0.02	0.02	0.01	0.01	nd	nd	nd	nd	nd	nd	nd		
HexNAc2Hex4	nd	nd	0.11	0.11	nd	nd	nd	0.04	0.05	0.01	0.02	0.02	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd		
HexNAc2Hex3	0.08	0.04	0.41	0.46	nd	nd	nd	0.54	0.58	0.07	0.14	0.04	0.04	0.03	0.01	0.01	0.00	0.01	0.00	0.04	0.00	0.01	0.02	
HexNAc2Hex2	nd	nd	0.06	0.08	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
HexNAc2Hex2Fuc	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	0.07	0.02	0.01	0.00	0.06	0.00	nd	nd	nd	nd	nd	nd
(HexNAc2Hex3)Fuc	nd	nd	nd	nd	0.43	0.38	0.13	0.06	0.02	0.01	0.00	0.00	0.33	0.12	0.16	0.08	0.07	0.06	0.09	0.08	0.30	0.22	0.11	0.10
(HexNAc2Hex3)HexNAcFuc	nd	nd	nd	nd	nd	nd	nd	nd	0.27	0.08	nd	nd	0.10	0.05	0.12	0.04	0.03	0.02	nd	nd	0.04	0.00	0.01	0.01
(HexNAc2Hex3)HexNAc2Fuc	nd	nd	nd	nd	nd	nd	0.04	0.01	nd	nd	nd	nd	0.00	0.00	0.01	0.00	0.00	nd	nd	nd	nd	nd	nd	nd
HexNAc2Hex2Fuc2	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	0.03	0.01	0.02	0.01	0.22	0.00	0.28	0.01	0.02	0.00	0.22	0.02
(HexNAc2Hex3)Fuc2	nd	nd	nd	nd	0.53	0.16	0.45	0.13	nd	nd	nd	nd	0.28	0.15	0.44	0.17	0.58	0.20	0.54	0.19	0.49	0.10	0.52	0.18
(HexNAc2Hex3)HexNAcFuc2	nd	nd	nd	nd	0.04	0.01	0.10	0.01	nd	nd	nd	nd	0.00	0.00	0.05	0.02	0.03	0.00	0.04	0.00	0.04	0.00	0.02	0.00
(HexNAc2Hex3)HexNAc2Fuc2	nd	nd	nd	nd	nd	nd	0.00	0.00	nd	nd	nd	nd	0.07	0.02	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)FucHex	nd	nd	nd	nd	0.00	0.06	0.00	0.07	0.00	0.04	0.00	0.00	0.00	0.21	0.11	0.15	0.00	0.06	0.00	0.00	0.00	0.00	0.56	0.00
(HexNAc2Hex3)HexNAcFucHex	nd	nd	nd	nd	0.00	0.04	0.00	0.22	nd	nd	nd	nd	0.00	0.11	0.01	0.10	0.00	0.05	0.00	0.03	0.00	0.07	0.00	0.08
(HexNAc2Hex3)HexNAc2FucHex	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
(HexNAc2Hex3)Fuc2Hex	nd	nd	nd	nd	0.00	0.35	0.00	0.36	nd	nd	nd	nd	0.00	0.24	0.00	0.35	0.00	0.57	0.00	0.45	nd	nd	nd	0.46
(HexNAc2Hex3)HexNAcFuc2Hex	nd	nd	nd	nd	0.00	0.00	0.00	0.07	nd	nd	nd	nd	0.00	0.00	0.00	0.04	0.00	0.04	0.00	0.06	0.00	0.05	0.00	0.02
(HexNAc2Hex3)HexNAc2Fuc2Hex	nd	nd	nd	nd	nd	nd	nd	0.00	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
High Mannose	0.82	0.84	0.22	0.23	0.00	0.00	0.00	0.34	0.25	0.88	0.80	0.80	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.07	0.04	0.00	0.03	0.10
Hybrid	0.10	0.11	0.20	0.12	0.00	0.00	0.00	0.06	0.07	0.03	0.03	0.14	0.06	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
Paucimannose	0.08	0.04	0.58	0.65	0.00	0.00	0.00	0.58	0.63	0.08	0.16	0.04	0.03	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.04	0.00	0.01
Core fucosylation	0.00	0.00	0.00	0.00	0.43	0.48	0.45	0.42	0.02	0.06	0.01	0.00	0.51	0.43	0.38	0.17	0.18	0.09	0.20	0.34	0.85	0.21	0.19	0.19
Core difucosylation	0.00	0.00	0.00	0.00	0.57	0.52	0.55	0.58	0.00	0.00	0.00	0.31	0.40	0.51	0.58	0.83	0.81	0.86	0.72	0.56	0.15	0.76	0.69	
Core galactosylation	0.00	0.00	0.00	0.00	0.45	0.01	0.72	0.00	0.04	0.00	0.00	0.01	0.56	0.01	0.49	0.00	0.71	0.00	0.64	0.00	0.68	0.00	0.55	

Table S3: Eggs per gram feces (EPG) and worm burden (WB) counted on respective days post oral infection (DPOI) and at necropsy day for each trial group with six animals.

Mean group counts are listed with standard deviation (SD) of the mean.

Counts	DPOI	Native+						Mean	SD
EPG	d0	0	0	0	0	0	0	0	0
	d14	0	0	0	0	0	0	0	0
	d21	0	0	0	0	100	0	17	40.8
	d28	0	50	200	50	150	250	117	98.3
WB	d30	230	360	570	440	890	260	458	245.1
Counts	DPOI	P+						Mean	SD
EPG	d0	0	0	0	0	0	0	0	0
	d14	0	0	0	0	0	0	0	0
	d21	6850	8530	9000	6400	5850	4550	6863	1669.5
	d28	6350	13050	6100	9700	9150	10400	9125	2615.9
WB	d30	2910	3280	2820	2980	3140	3350	3080	211.2
Counts	DPOI	P-						Mean	SD
EPG	d0	0	0	0	0	0	0	0	0
	d14	0	0	0	0	0	0	0	0
	d21	4500	2300	5200	2300	3450	0	2958	1858.1
	d28	7550	7750	5400	5600	8700	3550	6425	1907.6
WB	d30	2870	2490	2250	2930	3120	2510	2695	328.9
Counts	DPOI	AP+						Mean	SD
EPG	d0	0	0	0	0	0	0	0	0
	d14	0	0	0	0	0	0	0	0
	d21	3400	6350	2400	6000	7450	700	4383	2621.2
	d28	5500	8700	7400	5300	18300	6800	8667	4883.7
WB	d30	2800	3070	2240	2700	2690	2510	2668	278.8
Counts	DPOI	HAP+						Mean	SD
EPG	d0	0	0	0	0	0	0	0	0
	d14	0	0	0	0	0	0	0	0
	d21	500	2550	9500	3700	5350	4450	4342	3033.7
	d28	5050	7250	10400	8250	15000	5400	8558	3714.2
WB	d30	1800	1390	2840	2650	3170	2450	2383	667.8
Counts	DPOI	HAP-						Mean	SD
EPG	d0	0	0	0	0	0	0	0	0
	d14	0	0	0	0	0	0	0	0
	d21	1400	3100	3400	4050	1500	3650	2850	1128.7
	d28	6100	9150	6800	9850	5300	3800	6833	2304.7
WB	d30	2290	3100	2650	2590	2770	2180	2597	332.8
Counts	DPOI	Control						Mean	SD
EPG	d0	0	0	0	0	0	0	0	0
	d14	0	0	0	0	0	0	0	0
	d21	5050	6100	3150	1650	7850	8650	5408	2694.5
	d28	7100	8550	8900	5800	13500	10000	8975	2657.4
WB	d30	2880	2370	2340	2570	3080	3290	2755	390.4

Table S4 Plasmids of this study

SP: gp67 secretion signal peptide, PSP: PreScission protease cleavage site, HIS: His₁₀-tag, MBP: Maltose binding protein

Name	Insert/Gene	Primers used	Plasmid	Source
pFastBac1	-	-	-	Thermo Fisher Scientific
pUC57-H11	<i>H. contortus</i> H11	-	pUC57	GenScript
pUC57-AC1	<i>H. contortus</i> AC1	-	pUC57	GenScript
pUC57-PEP1	<i>H. contortus</i> PEP1	-	pUC57	GenScript
pSF10	SP, HIS, PSP, H11 (without TM)	H11-TM-Fw <i>Stu</i> I/SacI and H11-TM-Rev <i>Xho</i> I/new	pRG105	This study
pSF11	SP, HIS, PSP, AC1	-	pRG105	This study
pSF12	SP, HIS, PSP, AC1, HIS	AC1-FwHis_ <i>Xba</i> I and AC1-RvHis_ <i>Xho</i> I	pSF11	This study
pSF14	SP, HIS, PSP, PEP1	-	pRG105	This study
pSF15	SP, HIS, PSP, PEP1, HIS	PEP1-FwHis_ <i>Xba</i> I and PEP1-Rvhis_ <i>Xho</i> I	pSF14	This study
pSF22	SP, MBP, PSP, PEP1, HIS	-	pRG132	This study
pRG84	GST, PSP, PDI1	-	-	This study
pRG105	gp67 SP, His10, PSP, PDI1	-	-	(Hang et al., 2015)
pRG132	gp67 SP, MBP, PSP, cloning site	-	-	(Hang et al., 2015)
pMA148	FLAG, GALT-1	-	-	R. Gauss, unpublished
				(Titiz et al., 2009)

Chapter 3

The extensive surface variability of H11 aminopeptidase glycoproteins from *Haemonchus contortus*

Susanna M. Fleurkens¹, Patrick C. Brunner², Chia-wei Lin¹, Jasmin Frey¹,
Uwe Schmitt³, Robert Gauss¹, Christine Neupert⁴, Irene Schiller⁴, Markus Aebi¹

¹ Institute of Microbiology, Department of Biology,

² Institute of Integrative Biology, Department of Environmental Systems Science,

³ ID Scientific IT Services,

^{1, 2, 3} ETH Zurich, CH-8093 Zurich, Switzerland

⁴ Malcisbo AG, Wagistrasse 27a, CH-8952 Schlieren, Switzerland

Introduction

Haemonchus contortus is a blood-feeding parasitic nematode in sheep and goats. This successful parasite infects a broad range of domestic and wildlife ruminants and has spread from sub-Saharan Africa to almost all regions of the globe [1]. It belongs to the Trichostrongyloidea nematodes, which are characterized by high genetic variability and large population sizes [2, 3]. Five species of this superfamily, including *H. contortus*, display extremely high levels of within-population genetic diversity (96 – 99 %) [3]. *H. contortus* shows not only high genetic variability within, but also between populations within and among continents [4]. Sequence polymorphisms result from high mutation rates, large effective population sizes and migration rates [5]. In the case of *H. contortus*, the high levels of genetic diversity within populations are predominantly due to its large effective population size [6]. A single ruminant host can harbor hundreds of adult female worms with each sexually producing thousands of eggs per day, resulting in the excretion of billions of new genotypes onto pasture every few days [7]. Moreover, *H. contortus* has an up to ten times higher mutation rate than that of vertebrates [3]. This parasite provides a high adaptive capacity to respond to selective pressures, imposed by environmental factors, host immune responses or chemical treatments [6]. Indeed, the rapid development of resistances to anthelmintic drugs within a few years of drug usage is well documented [5, 7]. To overcome the fast spread of drug-resistant strains, the development of a safe and effective vaccine is the best practical alternative.

The most characterized native vaccine against *H. contortus* is the native H11 vaccine purified by lectin-affinity chromatography that contains a mixture of numerous different proteins with the H11 aminopeptidases being the most prevalent [8]. H11 is an integral membrane glycoprotein with a molecular mass of 110 kDa [9], exclusively expressed on the intestinal microvillar surface of the parasitic stages. It shows an aminopeptidase activity and degrades small peptides of host blood proteins [10]. Immunization with native H11 containing extracts induce high levels of protection leading to a reduction of fecal egg counts by up to 95 % and worm burden by 82 % [11]. Protective IgG containing sera inhibit the activity of the H11 enzyme *in vitro*, and levels of enzyme inhibition correlate with levels of protection *in vivo* [12]. Thus, protection is thought to be mediated by antibodies, taken up with the blood meal, inhibiting the enzyme activity

and disrupting nutrient absorption [13]. Therefore, the enzyme H11 is considered to be critical for parasite survival.

Despite the success of the native vaccine, the long term aim is to generate a defined and safe recombinant vaccine using a cost-effective production method. For this reason, there has been a continuous effort to produce recombinant versions of native antigens using various expression systems, like bacteria, yeast, insect cells or *Caenorhabditis elegans* [14]. Although numerous vaccine candidates were successfully expressed, to date, the highest protection was reached with baculovirus-expressed H11 inducing a partial protection of 30 % when tested in a sheep vaccination trial [15-17]. This reduced protective capacity of recombinant vaccines might be due to suboptimal folding, absence of other proteins or lack of post-translational modifications, such as *N*-glycosylation. *H. contortus* possesses a range of unusual *N*-glycan modifications absent from mammals, with up to three core fucose residues and two types of galactosylated fucoses [18, 19]. In a previous study, we tested glycoengineered High five cells, co-expressing the *C. elegans* galactosyltransferase GALT-1, to decorate *H. contortus* digestive proteases with nematode-like *N*-glycans, specifically the galactosylated core α 1,6 fucose and the immunogenic core α 1,3 fucose (S. Fleurkens, manuscript in preparation). Despite a significant anti-glycan and anti-protein immune response stimulated in sheep vaccinated with these recombinant glycoengineered proteins, no protection was observed when challenged with *H. contortus*. We postulated that due to the lack of other proteins, recombinant vaccines are not protective so far.

Interestingly, the native H11 vaccine is not a single protein preparation. It comprises a family of five different H11 isoforms (termed H11, H11-1, H11-2, H11-4 and H11-5), all found in protective H11 preparations purified by lectin-affinity chromatography [17]. The five family members share 62-75 % amino acid identity and are tandemly arranged in the genome, suggesting that the gene family has arisen through recent duplication and divergence [17]. Semi-quantitative RT-PCR showed that H11 and H11-1 are expressed in infective L3 larvae and all isoforms are expressed in adult worms, with H11-4 being the most abundant form and H11-5 being female-specific [17]. Thus different transcriptional patterns of the aminopeptidase H11 were identified, but little is known about the interaction amongst its isoforms (e.g. building isoform heterodimers).

To date, studies tested only single recombinant isoforms of H11, or combinations of two isoforms, e.g. *C. elegans* expressed H11-1/-4 or H11-4/-5, which induced not or only partially a protection of sheep in vaccination trials [17]. H11 isoforms might provide functional redundancy concerning the inhibition of H11 activity by antibodies, but H11 diversity might be further extended considering the high genetic variability detected within and among *H. contortus* populations.

In this study, we identified, characterized and compared the diversity of the aminopeptidase H11 family present in three geographically different *H. contortus* populations. We examined whether the *H11* isoform genes were under selection and correlated the genetic diversity with structural changes of H11 proteins and variable *N*-linked glycans. Our results are relevant for the development of a globally effective *Haemonchus* vaccine and point to possible requirements for vaccine-induced protective immunity in future vaccination approaches, also to be considered for strategies against other parasite induced diseases.

Results

H11 isoform sequences display an extreme genetic diversity in three geographical distant populations of *H. contortus*

To test whether different populations of *H. contortus* express different *H11* alleles, cDNA sequences of three different populations, each from one sheep, were analyzed. Fifty to hundred Swiss (CH), Scottish (SCO) or South African (ZA) mixed adult worms were used for RNA isolation and subsequent cDNA synthesis. cDNA for all *H11* loci, as well as for two other loci encoding intestinal proteins, *AC1* and *PEP1*, and the housekeeping protein *SOD1* were amplified from *H. contortus* cDNA.

Full open reading frames were obtained for all genes with the exception for Swiss worm derived H11 isoforms, where the forward primers were designed to amplify sequences downstream of the N-terminal transmembrane domain encoding cDNA. Sequence analysis revealed high nucleotide polymorphism amongst most of the H11 isoforms, resulting in many different unique sequences among a cDNA length of approximately 2900 base pairs (Table 1). Only the Swiss H11-4 and the South African H11-5 sequences did not show any sequence variation. The genetic diversity indices for the different genes analyzed are summarized in Table 1. Between 176 to 511 segregating sites were identified among the 25 - 29 sequences analyzed for each of the five *H11* isoforms, resulting in 15 - 29 unique nucleotide haplotypes. In contrast, only 10 and 11 haplotypes with 57 and 44 segregating sites were identified among 28 and 29 sequences analyzed for *AC1* and *PEP1*, respectively. *SOD1* revealed 18 haplotypes with 69 segregating sites among the 480 base pairs of coding sequence.

Average nucleotide diversity was higher for *H11* isoforms $\pi = 0.038$, (range 0,025 to 0,059) as compared to the average $\pi = 0.021$ (range 0,01 to 0,032) estimated for *AC1*, *PEP1* and *SOD1* respectively. The difference was marginally not significant, though (t -value = 1.80856, p -value = 0.06). Likewise, average haplotype diversity was higher for *H11* isoforms ($h = 0.957$, range 0,911 to 1,000) as compared to *AC1*, *PEP1* and *SOD1* ($h = 0.833$, range 0,768 to 0,901). This difference was significant ($t = 3.45972$, $p = 0.007$). None of the H11 isoform sequences identified in this study was identical to any previously published H11 sequence.

Table 1: Genetic diversity indices (excluding indels) for the *H11* isoforms, *AC1*, *PEP1* and *SOD1* genes of *H. contortus*.

Gene	<i>n</i>	<i>M</i>	<i>S</i>	π (Sd)	<i>h</i> (Sd)
<i>H11-1</i>	29	29	366	0,033	1,000
<i>H11-2</i>	29	22	310	0,029	0,966
<i>H11-4</i>	25	15	176	0,025	0,927
<i>H11-5</i>	29	21	511	0,059	0,911
<i>H11</i>	25	22	433	0,044	0,980
<i>H11 average</i>	137	21.8	359.2	0.038	0.957
<i>AC1</i>	28	10	57	0,021	0,831
<i>PEP1</i>	29	11	44	0,010	0,768
<i>SOD1</i>	30	18	69	0,032	0,901
<i>control average</i>	87	13	56.7	0.021	0.833

n, number of sequences sampled; *M*, number of haplotypes; *S*, number of segregating sites; π , nucleotide diversity; *h*, haplotypic diversity

We considered it unlikely that the observed nucleotide diversity was a result of amplifying or sequencing artifacts for several reasons: the polymerase used in this study was of high quality and the base-calling from the sequencing chromatograms was unambiguous. Furthermore, if the nucleotide polymorphisms were the result of a methodical error one would expect to observe a similar degree of diversity among all sequences. However, nucleotide diversity was unevenly distributed. For example, we did not observe a single nucleotide polymorphism (SNP) among Swiss samples for *H11-4* as compared to other populations at this locus. Therefore, we are confident that we are looking at true polymorphisms within and among genetic loci and between geographic locations.

We explored the phylogenetic relationships of the *H11* gene family on the amino acid level using all sequences obtained in this study. Additionally, we included the previously published reference sequences for each isoform for comparison (GenBank accession: *H11-1*, AJ249941.1; *H11-2*, AJ249942.2; *H11-4*, AJ311316.1; *H11-5*, KF381362.1; *H11*, X94187.1). The sequences clustered in five distinct phylogenetic groups with maximal statistical bootstrap support. All groups were exclusively

comprised of sequences originating from the same H11 isoform (Figure 1). Within these phylogenetic groups, sequences representing different sampling locations were randomly distributed, hence, no geographic sub-structure according to isolation-by-distance was detected. We conducted specific regression analyses by plotting pairwise geographic against genetic distances to confirm this observation. The Pearson correlation coefficients for all 5 isoform groups were low (ranging from $r = 0.18$ to 0.39) and all were non-significant (Supplementary Table 1).

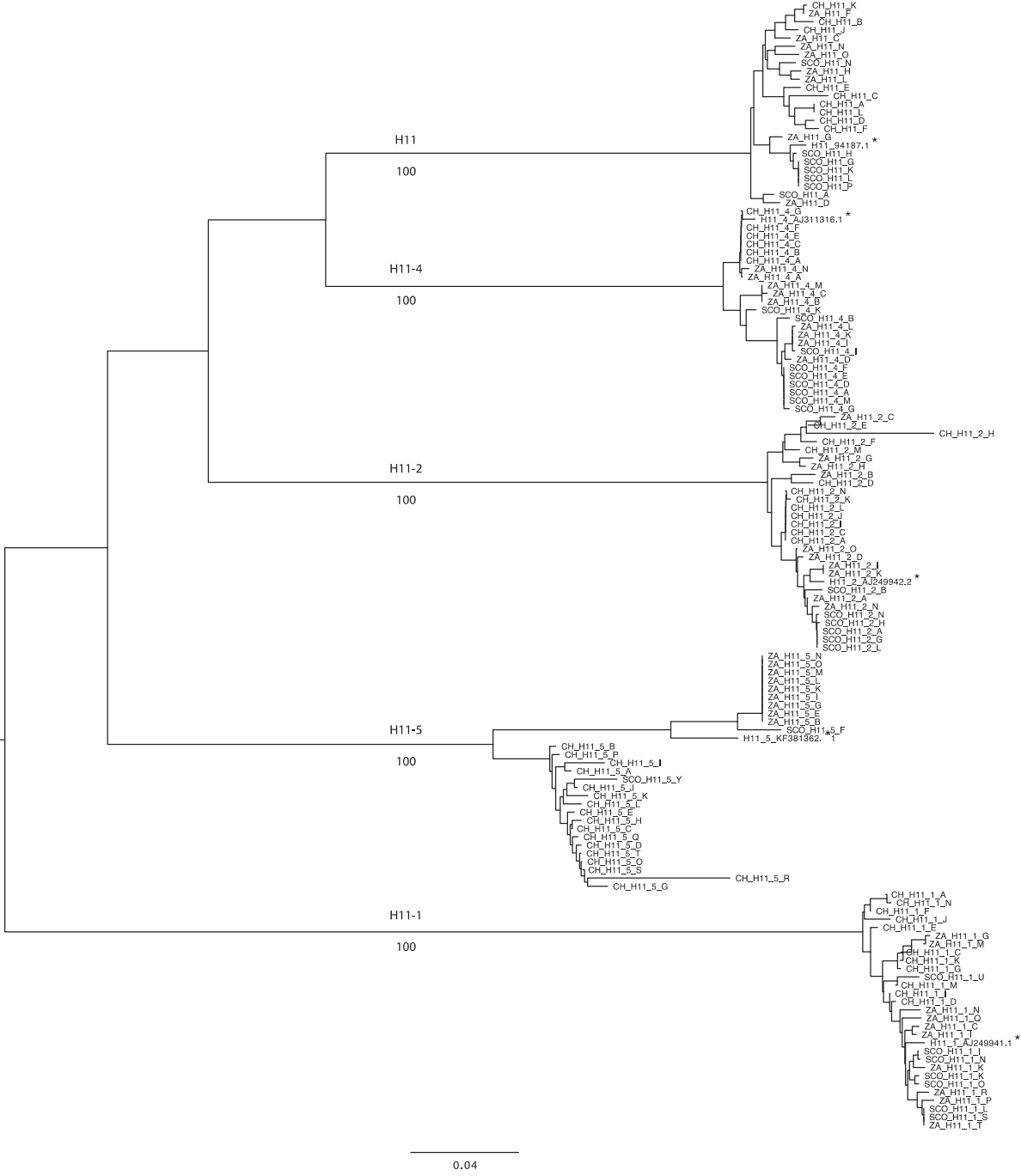


Figure 1: Molecular Phylogenetic analysis by Maximum Likelihood method

The evolutionary relationship was inferred by using the Maximum Likelihood method based on the JTT matrix model [20]. The tree with the highest log likelihood (-11962) is shown. Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.7279)). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 137 amino acid sequences, including the published sequences (marked with a star) and experimentally determined sequences of this study, labelled according to their geographic isolation from Switzerland (CH), Scotland (SCO) or South Africa (ZA). There were a total of 901 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [21].

The H11 amino acid diversity is found on the protein surface

Due to the potential of H11 aminopeptidases as vaccine candidates against *H. contortus*, we were interested in evaluating the potential of sequence diversity to create antigenic variation. We created consensus sequences for each of the five isoforms with amino acid positions showing 100 % identity in alignments with all sequences analyzed for one isoform (Figure 2A). These consensus sequences showed that 78 - 95 % of amino acid positions were conserved within one isoform, with the highest variable site value identified in H11-5 (223 amino acid exchange / 993 total amino acids). The *N*-terminal transmembrane domain was excluded from this analysis, as the sequence information was lacking for the Swiss population due to a different isolation strategy. A few constructs (H11: 3, H11-1: 6, H11-5: 2) had a mutation leading to a stop codon, these truncated sequences were excluded from the diversity calculations.

An alignment of the five consensus sequences revealed that the same variable sites were often found in more than one isoform at the same position (Figure 2A). Diversity occurred throughout all domains of the protein, except the areas surrounding the HEXXH₁₈E zinc-binding and GXMEN catalytic motifs (purple). To identify the location of the amino acid diversity on the protein structure, one sequence of each H11 isoform was modeled on the crystal structure of the homologous human aminopeptidase N (hAPN [22]). As the human aminopeptidase, the *H. contortus* H11 proteins are type-2 membrane glycoproteins. They form dimers and each monomer possesses the characteristic four-domain structure of the M1 metallopeptidases [22]. A model for the orientation of H11 on the cell surface is shown in Figure 2B. The modeled proteins are

shown in the closed form with an internal cavity in domain II (palegreen) surrounding the active sites.

Interestingly, the variable sites of all isoforms were found almost exclusively on the surface of the protein (Figure 2C, Supplementary Figure 1). Among the many amino acid mutations observed in H11, only a few (<12 % on average) were located inside the protein and no mutations were detected within the active sites. The protein surfaces exposed towards the microvilli plasma membrane (Supplementary Figure 1E) and the surface where potential dimerization takes place (Supplementary Figure 1C) were showing the least mutations.

Overall 50 % of all amino acid replacements were shared among the Swiss (CH), Scottish (SCO) and South African (ZA) population for H11-2, H11-4 and H11 (Table 2). In contrast, 93 % of the mutations identified in H11-5 were unique to one of the populations analyzed, leading to distinct surface patches when marked on the modeled protein (Figure 2D). Variability of protein sequences is expected to rise with more H11 sequences being analyzed, as we did not reach saturation in finding identical sequences for many of the subsets. Subsequently, amino acid diversity on the surface might increase as well.

Figure 2: Localization of the amino acid variation on the surface of the modelled H11 protein

(A) Conservation of H11 isoform sequences. Alignment of consensus sequences of H11 isoform show positions with 100 % identity within the sequences analyzed for one isoform (including the published sequence and all sequences of this study). A red dash shows lack of conservation and a black dot lack of an amino acid. HEXXH₁₈E zinc-binding and GXMEN catalytic motifs are conserved and marked in purple. The alignment background is colored according to the different domains of the modeled protein structure. **(B)** Dimeric H11 protein structure modeled by PHYRE2 [23] on the crystal structure of the human aminopeptidase N (hAPN [22]) by using the sequence of H11-5 variant A of the Swiss isolate. The protein forms a dimer and the left monomer is shown in a cartoon and the right monomer in a surface representation. The dimer is depicted in its proposed orientation with respect to the plasma membrane and colored according to domains (I: lightblue, II: palegreen, III: lightpink, IV: paleyellow). The grey arrow indicates the view on the protein surface used for figure C and D. **(C)** Variable amino acid residues found on a given isoform are marked in red on the modeled monomeric protein structures. The monomer is rotated (y, 90° x, 45°) with respect to the dimer in B. **(D)** Amino acid variability found in the different geographical populations of H11 isoforms is marked according to the localization of the diversity in Switzerland (orange), Scotland (cyan) or South Africa (green) or in two / three populations (purple) on the modeled monomeric protein structures. H11-4 and H11-5 represent two populations only as no polymorphism was found in Switzerland for H11-4 and in South Africa for H11-5. The monomer is rotated (y, 90° x, 45°) with respect to the dimer in B. The following sequences were used for the modeling of the protein structure: H11-4 variant A of the South African isolate, H11-1 variant D, H11-2 variant A, H11-5 variant A and H11 variant D of the Swiss isolate.

Table 2: Summary of variable amino acid sites in the H11 proteins

Overall includes all variable amino acid sites found among the published and newly identified amino acid sequences. *Overall (geographic)* depicts the amount of variable amino acid sites found among the new sequences and is the summary of the variable sites that are *Unique* or *Shared (geographic)*.

Protein	Overall	Overall (geographic)	Shared (geographic)	Unique CH	Unique SCO	Unique ZA
H11-1	72	64	23	17	3	21
H11-2	87	79	40	33	1	5
H11-4	49	45	23	-	4	18
H11-5	223	207	15	84	108	-
H11	90	80	40	23	7	10

H11 sequences are under strong diversifying selection

The high number of variable amino acid residues and their localization on the surface of the modeled H11 proteins, suggested that these genes may be under positive diversifying selection. A widely used approach to estimate the degree of selection is to compare dN (the number of non-synonymous changes per non-synonymous site) with dS (the number of synonymous changes per synonymous site). A ratio dN / dS of 1 ($\omega = 1$) indicates neutrality, while $\omega < 1$ is suggesting purifying, and $\omega > 1$ is suggesting diversifying selection, respectively. We estimated dN and dS for all pairwise sequence comparisons using DnaSP [24]. Figure 3 summarizes the results in a box plot graphic. Both, dN and dS values are significantly higher in the H11 family as compared to *AC1*, *PEP1* and *SOD1* (*T-test*, $p < 0.001$), supporting our hypothesis of accelerated evolution in the former group. However, the ratio ω was not > 1 for any of the genes.

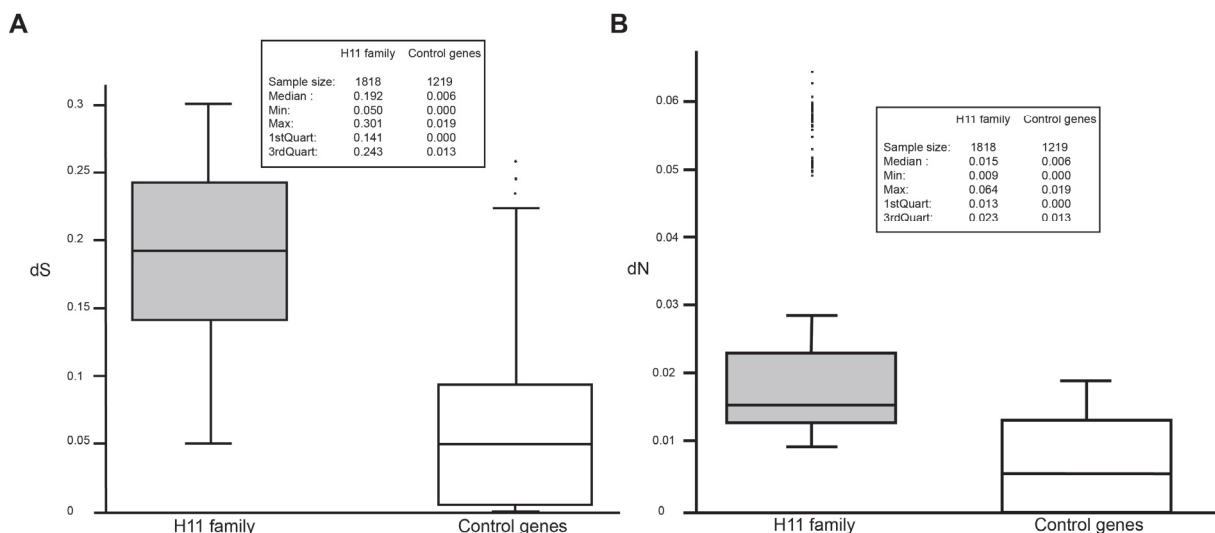


Figure 3: Ratios of synonymous (dS) and non-synonymous changes (dN)

Comparison of dS (**A**) and dN (**B**) estimates between all pairwise gene comparisons within the *H11* family and the genes *AC1*, *PEP1* and *SOD1* (control genes).

However, since diversifying selection is unlikely to affect all sites of a gene over prolonged time, ω averaged over all sites is almost never > 1 . Thus interest has been focused on detecting positive selection that affects only specific codon sites. We therefore applied the maximum-likelihood method CodeML implemented in PAML [25, 26] to test for codon-specific selection in the aminopeptidase genes. The selection model M8 (allowing ω to exceed 1 among codon sites) fitted the data significantly better

than the neutral model M7 (allowing ω to range from 0 to 1) for all of the H11 isoform groups ($p < 0.01$). This suggested that strong diversifying selection is acting on all members of this gene family (Table 3). In contrast, SOD1 and PEP1 showed no significant deviation from neutrality ($p = 0.490$ and $p = 1$). The gene sequences coding for the intestinal protein AC1 also showed evidence of diversifying selection.

The majority of the individual codon sites was under purifying selection, however, up to 17 sites (H11-4) were identified to be under positive diversifying selection (Table 3). The positions for the positively selected codons and the corresponding ω -values are listed in the Supplementary Table 2.

Table 3: Likelihood ratio test for positive selection of amino acid sites

Protein	Model 7 -ln (M7)	Model 8 -ln (M8)	LRT statistics*	<i>P</i> value	positively selected codons**	inference of evolution
H11-1	-9433.26	-9371.37	123.77	< 0.001	8	strong diversifying selection
H11-2	-8152.14	-8132.81	38.65	< 0.001	12	strong diversifying selection
H11-4	-5822.52	-5786.74	71.56	< 0.001	17	strong diversifying selection
H11-5	-9541.13	-9536.82	8.60	0.014	3	diversifying selection
H11	-9581.47	-9567.78	27.39	< 0.001	15	strong diversifying selection
AC1	-1879.84	-1870.71	18.25	0.010	3	diversifying selection
PEP1	-2135.05	-2133.64	2.81	0.490	0	neutral
SOD1	-1248.96	-1248.96	-0.01	1.000	0	neutral

*LRT statistics follow a χ^2 distribution with degrees of freedom of 2.

** Positively selected sites with posterior probability (*P*) values of > 0.95

H11 amino acid diversity rearranges *N*-glycosylation sites on the protein

H11 is a glycoprotein and therefore, this post-translational modification can contribute to the variability of the isoform and allelic proteins. *N*-glycans can change the surface structure of a protein dramatically due to their size and variability in composition.

First, we analyzed whether *N*-glycosylation sites (N-X-S/T) were affected by amino acid replacements. Individually, each H11 isoform harbors between 3 and 8 glycosites, but overall a total of 15 different glycosites, all located on the surface of the modeled protein, were found (Figure 4A). One site (glycosite 7) was conserved among all isoforms, two glycosites (3 and 13) were shared among four isoforms, glycosite 2 was shared among two isoforms and all others were uniquely found in one isoform only. Single nucleotide polymorphism (SNP) in the analyzed *H. contortus* populations led to the loss or gain of 8 *N*-glycosylation sites within the isoforms when compared to the published sequences, as shown in the glycosite sequence alignment (Figure 4B and 3C). In more detail, a SNP led to an aspartic acid replacement by asparagine and creates thereby a new glycosite in H11-1 (DYT → NYT, Figure 4C), found in all three populations analyzed. Like in H11-1, also in H11-2 a SNP introduced a new glycosite in the Swiss population (KVS → NVS, Figure 4G). In contrast, a SNP or a frame shift by three guanine insertions mutated a *N*-glycosylation site in the Swiss H11-5 (Figure 4D). Alternatively, mutations to serine or threonine on the third position of the N-X-S/T sequon generated new glycosites. For example in the Swiss and South African population, NYR was replaced by NYT or NYS in H11 (Figure 4E), NYA was replaced by NYT in Swiss and Scottish H11-5 (Figure 4F), and NGE was replaced by NGT in Scottish H11-2 (Supplementary Figure 2B). In contrast, none of the amino acid replacements found in H11-4 isoform affected glycosites.

In summary, 8 out of 15 glycosites among the isoforms were affected by the amino acid replacements and individually per isoform the following amount of glycosites were variable: H11-1 (1 out of 5), H11-2 (2 out of 4), H11-4 (none of the 5), H11-5 (4 out of 8), and H11 (2 out of 5) (Figure 4 and Supplementary Figure 2).

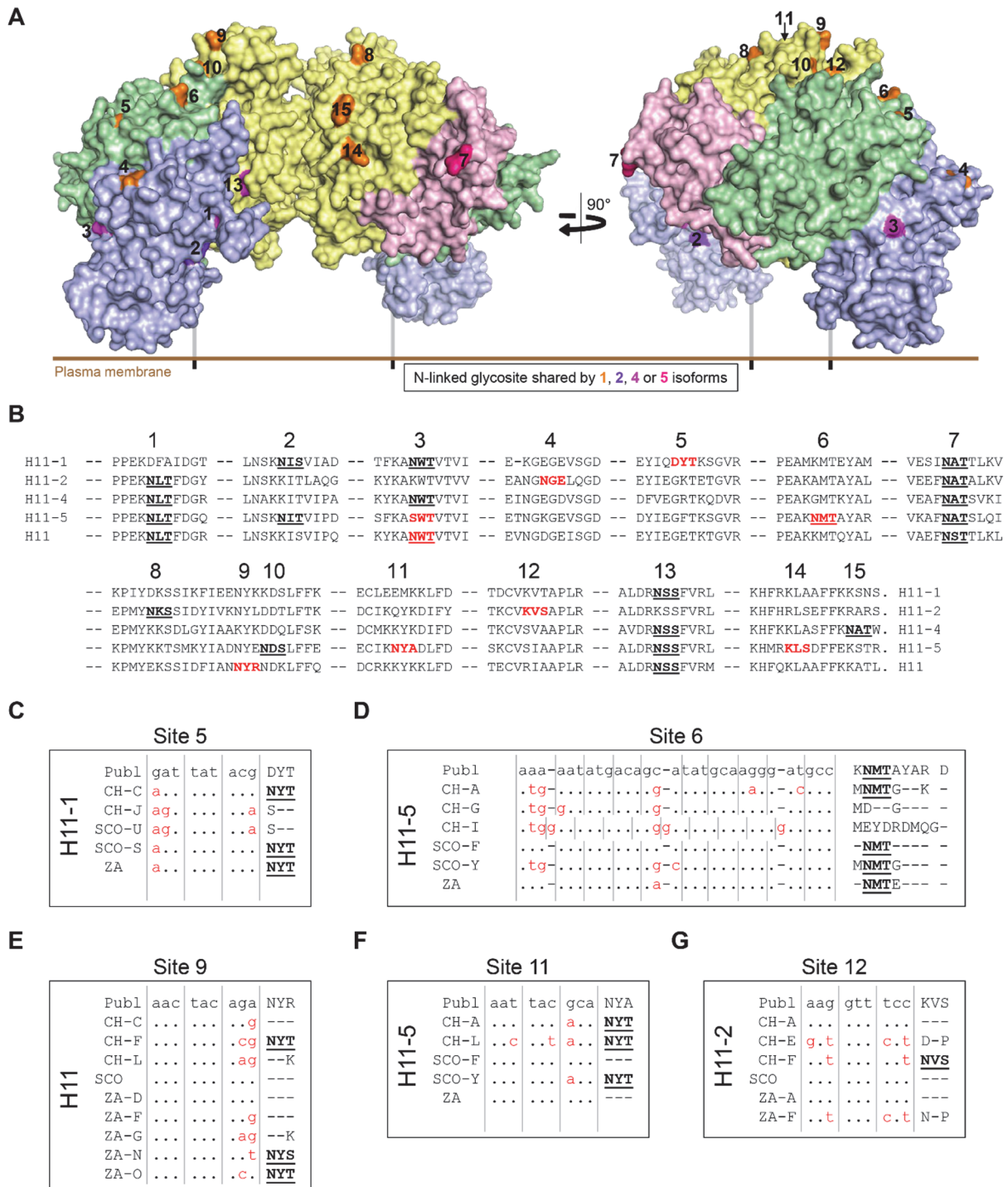


Figure 4: H11 amino acid diversity affects the loss or gain of N-glycosylation sites

(A) Dimeric Swiss H11-5 allele (CH-H11-5A) protein structure modeled by PHYRE2 [23] on the crystal structure of the human aminopeptidase N (hAPN [22]). The Pymol surface model is colored to indicate the different domains (I: lightblue, II: palegreen, III: lightpink, IV: paleyellow). 15 asparagines, or the corresponding aligned amino acid in H11-5A, are highlighted on the protein surface and colored by conservation of the glycosite among isoforms, if shared among all five H11 isoforms: pink (glycosite 7), among four: magenta (glycosite 3 and 13), among two: purple (glycosite 2) or uniquely found in one isoform: orange (glycosite 1, 4, 5, 6, 8-12, 14, 15) (B) Alignment of glycosites of H11-1 (AJ249941.1), H11-2 (AJ249942.2), H11-4 (AJ311316.1), H11-5 (KF381362.1) and H11 (X94187.1). All 15 glycosites

are either found in the aligned published sequences (bold type and underlined) or in sequences of the other populations analyzed (bold type). The *N*-glycosylation site is marked in red, when it is lost or gained compared to the published sequence. **(C) - (G)** Detailed view of alignment of different H11 isoform (vertical) nucleotide sequences of different populations (Swiss, CH; Scottish, SCO; South African, ZA). Nucleotide mutations or insertions (red) causing amino acid replacements in the *N*-glycosylation sequon (bold type and underlined) leading to the loss or gain of a *N*-glycosylation site when compared to the published sequence (Publ).

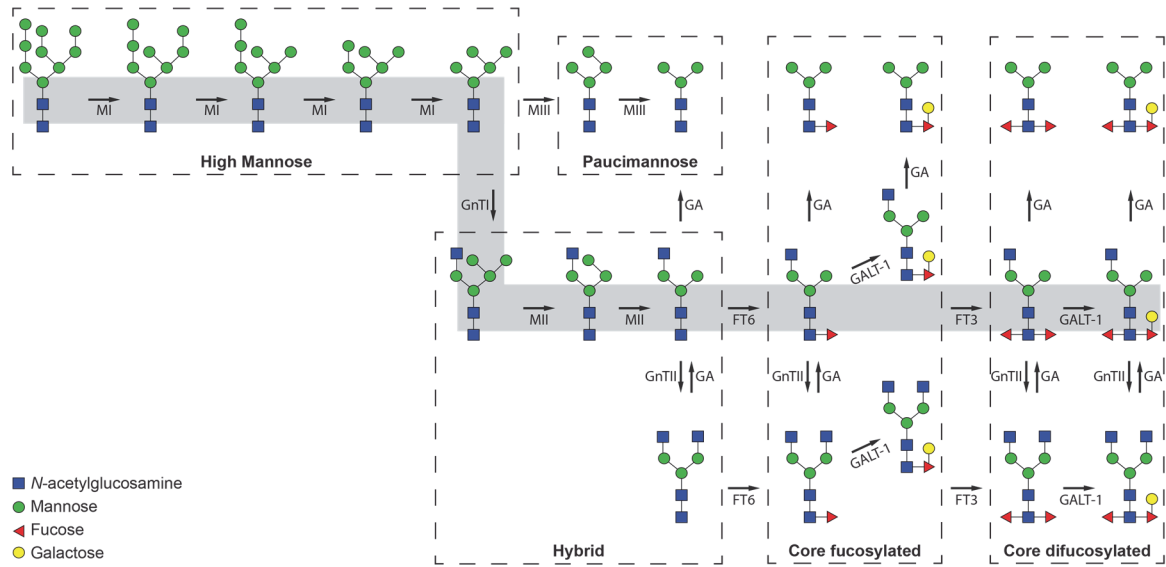
Recombinantly expressed H11 isoforms show site-specific differences in the *N*-glycan modifications

To examine whether variations of the primary amino acid sequence had an effect on the *N*-linked glycan structures of the H11 isoforms we recombinantly expressed H11 proteins in insect cells and assessed their site-specific *N*-glycan modifications. Site-specific glycan heterogeneity is primarily affected by the structure of the protein and the glycosylation potential of the expressing cell [27].

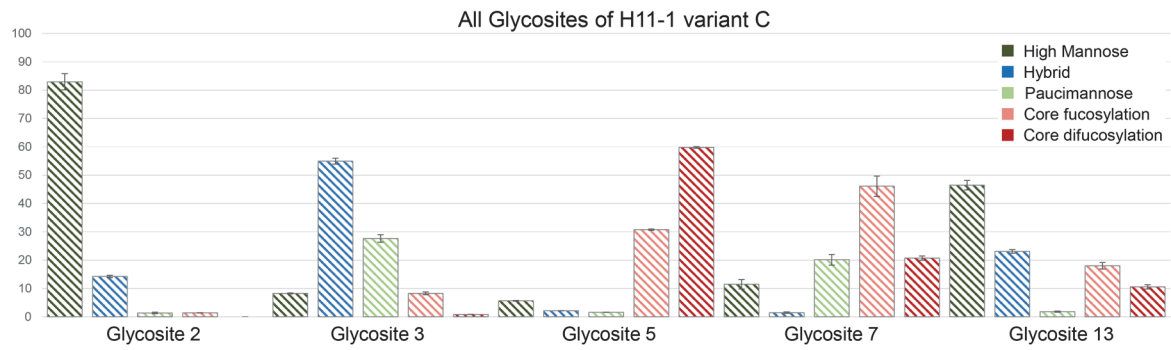
H11 proteins were produced in glycoengineered High five insect cells co-expressing the *C. elegans* galactosyltransferase-1 GALT-1 [28] (Supplementary Figure 3). This enzyme adds a β 1,4-galactose to the α 1,6-linked core fucose residue (GalFuc) in the Golgi apparatus of High five insect cells, thereby creating the glycoepitope GalFuc found on native *H. contortus* proteins. Secreted His₁₀-tagged H11 proteins were affinity purified from the culture supernatant. Proteins were proteolytically cleaved with trypsin and the digested glycopeptides were subsequently analyzed by nanoHPLC-HCD-MS/MS mass spectrometry (Supplementary Table 3) [27]. Due to the complexity of the analysis, this was done for some Swiss variants only. Different glycoforms of the glycosylated peptides were quantified for every *N*-glycosylation site (Supplementary Table 4) and grouped into high mannose, hybrid, paucimannose, core fucosylated and core difucosylated structures according to the level of processing in High five cells (Figure 5A).

Recombinantly produced H11 isoforms co-expressed with GALT-1 were decorated with nematode-like *N*-glycans similar to the ones found on native H11 from *H. contortus*, namely with the galactosylated core α 1,6 fucose (the GalFuc epitope) and the core α 1,3 fucose. As expected, we observed an *N*-glycan heterogeneity at all *N*-glycosylation sites analyzed (Supplementary Table 4). For example, the analysis of the *N*-glycosylation sites of the Swiss variant C of isoform H11-1 revealed differences in the processing of *N*-glycans on all five glycosites (2, 3, 5, 7 and 13) (Figure 5B). Importantly, the processing of glycans was site-specific. Glycosite 2 and 13 carried mainly high mannose structures, whereas glycosite 3 was decorated with hybrid and paucimannose glycans, and glycosite 5 and 7 were processed the most with core mono- and di-fucosylated structures.

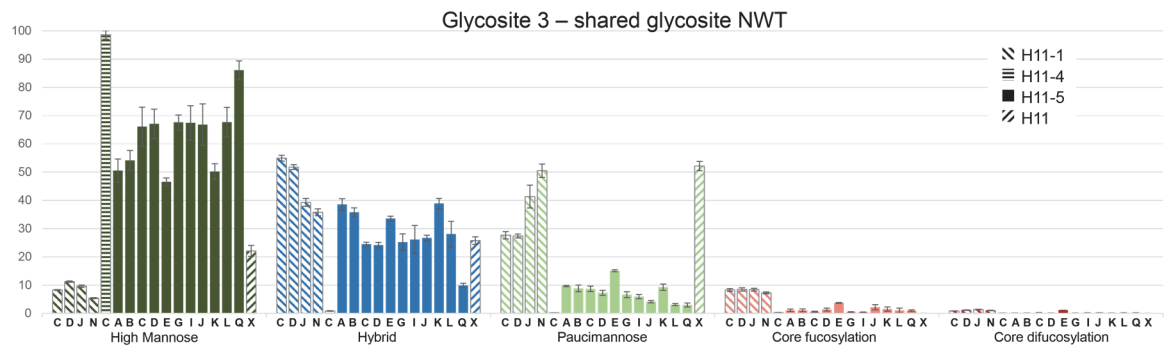
A



B



C



D

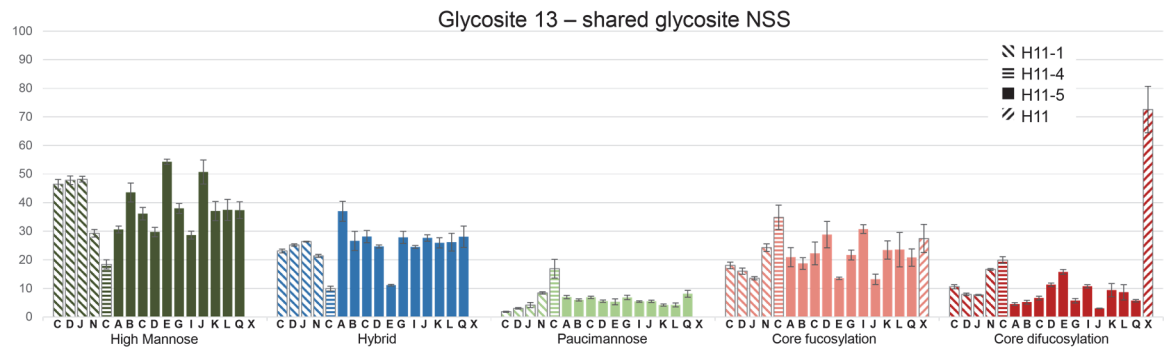


Figure 5: Recombinant H11 isoforms purified from glycoengineered insect cells show differences in the site-specific N-glycan modifications

(A) N-glycan-processing pathway in the ER and Golgi apparatus of glycoengineered High five insect cells overexpressing the *C. elegans* galactosyltransferase. The main pathway is underlined with a grey bar and all glycan structures and enzymes involved are shown (MI, MII, MIII: α -mannosidase I, II and III; GnTI, GnTII: N-acetylglucosamine transferase I and II; GA: N-acetylglucosaminidase; FT6, FT3: α 1,6- and α 1,3-fucosyl transferase, GALT-I: *C. elegans* β 1,4-galactosyltransferase). The glycan structures are grouped according to the level of processing into high mannose, hybrid, paucimannose, core fucosylated and core difucosylated structures. **(B)** Site-specific N-glycan profile of the Swiss variant C of the H11-1 isoform. The relative abundance of the different glycoforms at a given glycosite (2, 3, 5, 7 and 13) is shown. **(C)** Site-specific N-glycan profile of the shared glycosite NWT (Glycosite 3) of isoforms H11-1, H11-4, H11-5 and H11. The relative abundance of the different glycoforms for the different isoform variants (H11-1C, -D, -J, -N; H11-4C; H11-5A, -B, -C, -D, -E, -G, -I, -J, K, -L, -Q and H11-X) is shown. **(D)** Site-specific N-glycan profile of the shared glycosite NSS (Glycosite 13) of isoforms H11-1, H11-4, H11-5 and H11. The relative abundance of the different glycoforms for the different isoform variants (H11-1C, -D, -J, -N; H11-4C; H11-5A, -B, -C, -D, -E, -G, -I, -J, K, -L, -Q and H11-X) is shown. All H11 isoforms were purified from insect cells co-expressing the *C. elegans* galactosyltransferase GALT-1. Site-specific glycan profiles of trypsin digested glycopeptides were analyzed by nanoHPLC-HCD MS/MS mass spectrometry. Data represent the mean values of grouped glycan ratios of two (5E, 5F and 5I) or three (other variants) different experiments with error bars indicating the standard deviation. Detailed data analysis can be found in Supplementary Table 3 and 4.

We further compared the site-specific glycostructures at conserved N-glycosylation sites among the different H11 isoforms and within alleles of one isoform. We observed that all three shared N-glycosylation sites (3, 7 and 13) showed distinct processing patterns among the four isoforms. While isoform H11-4 showed mainly un-processed high mannose structures at glycosite 3, H11-5 exhibited hybrid structures and in contrast, H11-1 and H11 presented further processed paucimannose glycan structures (Figure 5B). An interesting difference in N-glycan processing was detected on the shared glycosite 13 (Figure 5C). Here, the isoform H11 showed 50 % more difucosylated core structures as compared to the other isoforms analyzed. This glycosite also showed differences in the galactosylation of the core fucose with variable rates ranging from 13 to 49 % (Supplementary Table 4). Our results clearly demonstrated that the sequence diversity between and within isoforms strongly affected N-glycan processing, adding an additional level of surface variability to H11 (Figure 5B and 5C).

We evaluated the variability of our experimental approach to determine glycan heterogeneity by expression in an adapted insect-cell expression system and analyzed site-specific *N*-glycan heterogeneity in native H11 preparations. Due to the analytical method (analysis of glycopeptides), we were able to characterize isoform-specific sites in native extracts. We extracted native H11 from a detergent-soluble extract of *H. contortus* adult worms, with or without further enrichment by lectin affinity purification, and analyzed the site-specific *N*-glycosylation of both preparations by mass spectrometry (Supplementary Table 5). Highly similar glycosylation patterns were identified for both proteins, the recombinantly expressed H11 from glycoengineered insect cells and the native H11 present in adult worms (Figure 6). In both samples, glycosite 1 was modified mainly by high-mannose and hybrid *N*-glycan structures, whereas glycosite 3 showed mainly paucimannose structures and glycosite 13 was decorated by highly processed nematode-specific *N*-glycans (GalFuc). We concluded that the recombinant H11 site-specific glycosylation resembled closely the glycosylation pattern of the native H11 protein.

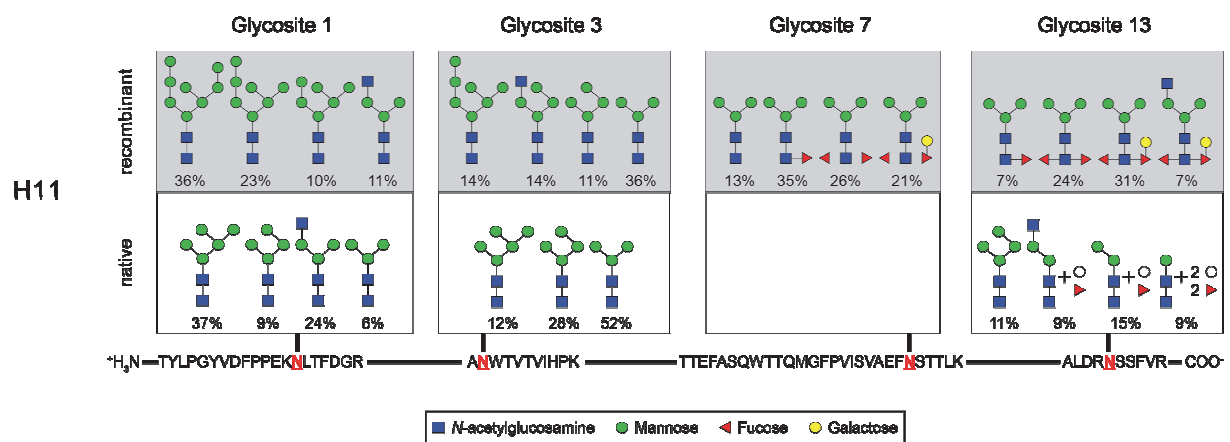


Figure 6: Native H11 and recombinant H11 show site-specific and similar *N*-glycan modifications

Glycosite-specific *N*-glycosylation profile of recombinant H11 purified from insect cells co-expressing the galactosyltransferase GALT-1 (upper panel, grey) and native H11 purified from adult *H. contortus* worms (lower panel, white). Trypsin digested glycopeptides were analyzed by nanoHPLC-HCD MS/MS mass spectrometry. H11 has four *N*-glycosylation sites, their location is marked by red N's representing the asparagine residue where the glycan is attached. Trypsin digested peptide sequences of these four glycosylation sites are shown with the four most abundant carbohydrates attached and their respective percentage. Detailed data analysis can be found in Supplementary Table 4 and 5.

Discussion

As we proceed in combating the anthelmintic resistance problem of the sheep parasite *H. contortus*, developing an effective recombinant vaccine would be the best alternative to effectively control the disease. Over the last decades, numerous protective native antigens have been identified, with the native H11 vaccine being one of the most effective against *H. contortus* infection. Vaccination studies with recombinant H11 used single isoforms or combinations of two isoforms and were unsatisfying in reaching the necessary protection rate. The *Haemonchus* genome encodes five different H11 isoforms [17] and, in combination with the high level of genetic diversity in *H. contortus* [4], this might explain the failure of recombinant H11 protein to elicit a protective immune response. Based on this hypothesis, we examined the distribution and conservation of H11 antigen genes within and between parasite populations, as this might have fundamental implications on the development of a globally or locally active recombinant H11 vaccine.

We identified an extreme genetic diversity specifically within the *H11* aminopeptidase family of three different *H. contortus* populations. Both, average nucleotide diversity as well as haplotype diversity, was higher for *H11* isoforms as compared to the average estimated for *AC1*, *PEP1* and *SOD1* respectively. This level of diversity was comparable to previously reported variation in cysteine proteases of *H. contortus* (range 0.0129 to 0.1024) [29-31]. Specifically, for the cysteine protease *AC1*, also tested in our study, the nucleotide diversity was found to be similar to the one reported in a first study [30], but lower as compared to the one in a more recent publication [31]. Previous studies attributed the high genetic diversity within *H. contortus* populations to a high mutation rate and large effective population sizes [2, 3]. In fact, billions of new genotypes are passed onto pasture every few days in a flock with hundreds of sheep, each harboring hundreds of adult female worms of which each produces thousands of eggs per day [7]. Moreover, enough progeny is created to mutate every single nucleotide position in the genome each day [7], assuming that *H. contortus* has a similar mutation rate as *C. elegans*, 2.1×10^{-8} mutations per site per generation [32]. Furthermore, these high levels of genetic diversity are retained in laboratory isolates passaged for many years [33-35]. These properties of *H. contortus* explain the high levels of genetic diversity found in this study.

One consequence of high nucleotide diversity is, that it might lead to non-neutral modifications of the protein sequence, either through non-synonymous mutations leading to an amino acid change or a frame shift, or introduction of stop codons leading to non-functional proteins. In this study, up to 22 % of the H11 protein sequence was affected by amino acid replacements (223 out of 993 amino acids). This extent of variation is at an extreme level, and up to 10 times higher as compared to other reported nematode proteins and vaccine candidates tested so far [36]. For example, a relatively low level of 2 % diversity was found in the hookworm vaccine antigen, the *Ancylostoma* secreted protein termed ASP-1 (10 out of 424 amino acids) [37].

Importantly, the sequence variation of the H11 proteins was found almost exclusively on the surface of the modeled protein structure (>88 % on average), leaving the internal structure of the protein unaffected. Additionally, no amino acid replacements were found within the catalytic motifs of the proteins. We concluded that the biological activity (hydrolysis of peptides) of the H11 enzymes was not significantly modified by the sequence diversity within one isoform. No estimation on the absolute level of H11 diversity in *H. contortus* was made, as we did not reach saturation for many of the sequence subsets. None of the H11 isoform sequences identified in this study was identical to any of the previously published ones, indicating sequence diversity in different *H. contortus* populations from different geographic locations. Indeed, a minimum of 50 % of the amino acid replacements were specific for one geographical isolate. However, also 50 % of the amino acid variation was found in the two or three populations analyzed in this study. Variability of protein sequences is expected to rise with more H11 sequences and more parasite populations being analyzed. Amino acid diversity might increase further and might enhance diversity of H11 proteins even more.

Surface variability in the H11 population was increased further by post-translational modifications, in particular *N*-linked glycosylation. Genetic changes resulted in the gain or loss of 8 out of 15 *N*-glycosylation sites among the isoforms due to alteration of the N-X-S/T glycan acceptor sequon. As the surface structure of a glycoprotein also directs *N*-glycan processing in the Golgi compartment [27], we expected differentially structured *N*-glycans on H11 isoforms and allelic H11 variants. Indeed, mass spectrometry-based analysis revealed varying *N*-glycan structures between isoforms and between the variants of one single H11 isoform. This site-specific *N*-glycan

processing was not only detected in recombinantly expressed proteins from insect cells, but also in the native H11 isoforms purified from adult *H. contortus* worms.

This high surface variability of H11 asks for the biological factors promoting H11 diversity in *H. contortus*. Using maximum-likelihood models, we showed that positive diversifying selection is acting on the *H11* genes and we hypothesize that the host antibody-mediated immune response exerts the selective pressure. This might be in conflict with the “hidden antigen hypothesis” that explains the protective mode of action of native H11 vaccine [38]. There is evidence that “hidden antigens” are not completely hidden from the host's immune system, as unvaccinated animals after being exposed to *H. contortus* for a long time are able to generate an immune response against the gut antigens [39, 40]. In addition, the immune response against H11 primed by immunization can be boosted, presumably by H11 released from dead or dying parasites in the sheep stomach [38, 41].

Why would an antibody-mediated immune response against the surface of the H11 protease be growth inhibiting for *H. contortus*? The answer might lay in the mode of action of the protease. As suggested for the human homolog, the arch-like structure of the closed dimeric form showed that the internal cavity harboring the catalytic site is not connected to bulk solvent by an appreciable channel or opening [22]. It was suggested that a large conformational change from the closed dimer to an open S-shaped dimer configuration makes the catalytic center accessible to bulk solvent and the substrates, the interconversion of the open to the closed form is essential for enzymatic activity and the reaction cycle [22]. We propose that antibody binding to the monomers of the H11 enzyme prevents such a conformational change, thereby inhibiting the H11 activity, as was shown for sera from vaccinated sheep [12]. Consequently, the broad surface variability might outcompete the adaptive immune system in reoccurring infection cycles and prevents protection in vaccination using recombinant protein. However, priming the immune response with a highly variable mixture of H11 isoforms and allelic variants as achieved by vaccination with native H11 extract is sufficient for prevention of primary infection.

One very well studied example for immense surface variation including rearrangement of *N*-glycans to avoid recognition by the host immune response is the highly variable human immunodeficiency virus 1 (HIV-1). Approximately 50 % of the gp120 surface was shown to have a variability of greater than 10 % [42]. In addition, the acquisition

and rearrangement of *N*-glycans are building a glycan shield allowing HIV-1 to conceal itself from neutralizing antibodies. As a consequence, only 3 % of the surface is still immunoglobulin-domain accessible, when the genetic variation was combined with glycan shielding [42]. Interestingly, there are broadly neutralizing human antibodies against HIV infection [43].

In summary, H11 proteins of *H. contortus* show a massive surface variability due to amino acid alterations on the surface, different positioning of *N*-glycans, and varying *N*-glycan structures between the isoforms and between the variants of one isoform. In our view, the very high structural complexity in epitopes is the result of a selection, where the host immune system inactivates the essential activity of the intestinal H11 protease by inhibiting the catalytic reaction cycle of the enzyme. A detailed analysis of H11 variability might lead the way to a properly designed recombinant vaccine that can counteract the high variability of the target in *H. contortus* populations. To the best of our knowledge, this is the first study that analyzed the variability of a eukaryotic surface glycoprotein at the level of the sequence structure and glycan composition. In the case of the H11 protein family in *H. contortus*, selection seems to act on the surface properties of the glycoprotein, resulting in protein sequence and *N*-glycan variability. The analysis of *N*-glycan variability is therefore essential to understand the selective forces that shape glycoprotein function.

Methods

***H. contortus* parasite material and RNA extraction**

Fifty to hundred Swiss (CH), Scottish (SCO) or South African (ZA) whole, mixed adult worms, amplified from field isolates, were used for total RNA extraction using an RNeasy kit (Qiagen # 74804). cDNA was synthesized using the transcriptor universal cDNA master kit (Roche # 05893151001) following the manufacturer's instructions.

Amplification of *H. contortus* coding genes and construction of plasmids for baculovirus protein expression

Open reading frames of all plasmids constructed in this study were confirmed by nucleotide sequencing.

Plasmids were generated using standard cloning protocols and are listed in Supplementary Table 6 and all primers used in this study are listed in Supplementary Table 7. 0.5-2 µg cDNA was used as a template and the following thermal cycle was used for the gradient Touch Down PCR: 98°C 30 sec; [98°C 10 sec, 72-60°C ($\Delta T(^{\circ}C) = -1$) 30 sec, 72°C *sec] 12 cycles; [98°C 10 sec, gradient (60.3°C, 62.4°C, 65.2°C, 68.2°C, 70.8°C) 30 sec, 72°C *sec] 32 cycles; 72°C 5 min (* dependent on the size of the gene). PCR amplified coding genes were either ligated into pGEM-T- Easy for nucleotide sequencing or further cloned into the insect cell expression vector pFastBac™ Dual.

pSF23-Dual-GALT: The plasmid was constructed to allow for simultaneous expression of *C. elegans* galactosyltransferase GALT-1 and another gene. *C. elegans* FLAG-GALT1 was amplified by PCR using pMA148 as template and primers 'GALT Fw XmaI' and 'GALT Rv NsiI'. The insert was cloned as *XmaI-NsiI* fragment into the respective restriction sites of vector pFastBac™ Dual downstream of the p10 promoter.

pSF25 and pSF34-56: These plasmids were constructed to allow simultaneous expression of *C. elegans* GALT-1 and secretion of N-terminally His-tagged *Haemonchus contortus* H11 isoforms in the pFastBac™ Dual.

After Touch Down PCR amplification from cDNA using primers 'H11-1 fw ApaI' and 'H11-1 rv StuI' for H11-1, 'H11-2 fw ApaI' and 'H11-2 rv StuI' for H11-2, 'H11-4 fw ApaI' and 'H11-4 rv HindIII' for H11-4, 'H11-5 fw ApaI' and 'H11-5 rv StuI' for H11-5, the PCR

products were cloned as either *StuI/ApaI* fragment (H11-1, H11-2, H11-5) or *HindIII/ApaI* fragment (H11-4) into pRG101 downstream of the N-terminal gp67 secretion signal peptide (SP), His₁₀-tag (HIS) and PreScission protease cleavage site (PSP). Finally, the coding genes (including SP, HIS and PSP) were excised by digestion with either *CpoI/XmaI* (H11-1C, -1D; H11-2A, -2D; H11-4C, H11-5A, -5B, -5C, -5D, -5E, -5F, -5J, -5K) or with *PauI/XmaI* (H11-1J, -1N; H11-2F, -2H; H11-5G, -5I, -5L, -5Q, -5R) and introduced into pSF23-Dual-GALT downstream of the Polyhedrin promoter generating pSF34 - pSF56. In contrast, the H11 gene was cloned as *BamHI-HindIII* fragment from a Genscript synthesized plasmid downstream of the Polyhedrin promoter of the pFastBac™ Dual (S. Fleurkens, manuscript in preparation). Then the *C. elegans* GALT-1 was cloned as *XmaI-NsiI* fragment into this vector downstream of the p10 promoter (PCR-generated using pMA148 as template and 'GALT Fw XmaI' and 'GALT Rv NsiI' as primers) generating pSF25.

Genetic diversity and phylogenetic reconstruction

Nucleotide sequences were aligned using MAFFT version 7 [44] and gene diversity estimates were calculated using the program DnaSP version 5.10.1 [24]. Since we did not detect significant geographic structure among sampled populations (Figure 1), diversity parameters were estimated for the pooled samples for each H11 isoform group.

The phylogenetic relationship of the entire H11 gene family was assessed on the entire data set after sequence alignment on the amino acid level with MAFFT. The data file was tested for goodness of fit to 48 models of amino acid evolution to select the most appropriate evolutionary model for the analysis using the “find best model” option implemented in MEGA6 [21] (Supplementary Table 1). The final analysis was conducted in MEGA6 using the Maximum Likelihood method based on the Jones-Taylor–Thornton matrix model with a gamma distribution to model evolutionary rate differences among sites (JTT +G) [20]. Confidence values for the phylogenetic tree were inferred with 100 bootstrap replicates.

Selection analysis using maximum likelihood models

A widely used approach to test for signatures of selection is to compare synonymous (dS) and non-synonymous substitution rates (dN) for the gene of interest. From the ratio $dN / dS = \omega$ the type of selection can be inferred in the following way. If non-synonymous substitutions are deleterious purifying selection will remove them from the population and ω will be < 1 . If non-synonymous mutations are neutral they will be fixed at the same rate as synonymous mutations and ω will be $= 1$. Finally, if non-synonymous substitutions are beneficial they will be retained by (positive) diversifying selection and ω will be > 1 .

We estimated the average values of dN and dS across all amino acid codons and for all pairwise sequence comparisons using the approximate method of Nei and Gojobori [45] as implemented in the DnaSP software package.

However, since diversifying selection is unlikely to affect all sites of a gene over prolonged time, ω averaged over all sites is almost never > 1 . Thus interest has been focused on detecting positive selection that affects only specific codon sites. We therefore applied the maximum-likelihood method CodeML implemented in PAML [25, 26] to test for codon-specific selection in the aminopeptidase genes. Two codon substitution models were compared. In the neutral model "M7" the range of ω variation is restricted between 0 and 1 among codon sites (beta distribution). In the alternative model of diversifying selection "M8", however, ω was allowed to exceed 1 (beta & ω). The resulting likelihood scores for the two model were compared with a likelihood-ratio-test for better fit with the data set.

Expression in insect cells

Trichoplusia ni "High Five" and *Spodoptera frugiperda* Sf21 cells were obtained from Prof. K. Locher (ETH Zürich, Switzerland) and cultivated in Sf-900 II SFM medium (Invitrogen # 10902104) without fetal calf serum at 27 °C in shaker flasks. Recombinant bacmids were generated using DH10EmBacY *E. coli* cells and viruses were produced in Sf21 cells according to standard procedures (Invitrogen).

Proteins were expressed in High Five cells by diluting the culture to 1 million cells per ml and infected with 1:100 (v:v) with recombinant virus. GALT-1 expression levels were detected 72 hours post infection as described in [28]. Briefly, 1 million cells were

pelleted by centrifugation (600 x g, 5 min), washed with PBS and lysed in 125 μ L lysis buffer (PBS, 1 % Triton X-100, 1x Protease Inhibitor cocktail (Roche #11873580001)) by rotating at 4 °C for 10 min. The whole cell lysates were centrifuged (16'000 x g, 5 min), and the supernatant was analyzed by immunoblotting. For secreted proteins, the culture supernatants of cells were harvested 72 hours post infection by centrifugation at 3800 x g for 10 min and filtered through a PES-membrane 0.22 μ M filter (Techno Plastic Products) followed by protein purification.

Protein purification

Affinity purification was performed on gravity flow columns filled with 1 ml Ni-NTA agarose beads (Protino #745400.100) equilibrated with 10 column volumes (CV) of Sf900 II SFM medium. The culture supernatant was loaded and the resin was washed afterwards with 10 to 15 CV washing buffer (10 mM imidazole in PBS, pH 7.4). Proteins were eluted four times with 1 CV of elution buffer (250 mM imidazole in PBS, pH 7.4) Elution fractions were concentrated and buffer exchanged to PBS on Amicon Ultra-4 Centrifugal Filter Devices (Millipore). Native H11 extract was prepared from *H. contortus* adult worms by solubilization of a PBS homogenate pellet in 2 % Triton X-100 (Sigma-Aldrich #T9284) and purified using Concanavalin A Sepharose (Sigma-Aldrich #C9017) according to patent WO1990011086 A1 by Munn and Smith [46].

SDS-PAGE and Immunoblotting

Protein samples were boiled at 95 °C for 5 min in reducing sample buffer (62.5 mM Tris-HCl pH 6.8, 2 % SDS (v/w), 5 % β -mercaptoethanol (v/v), 10 % glycerol (v/v), 0.01 % bromophenol blue (w/v)) and separated by SDS-PAGE (10 % acrylamide, 120 Volt, 150 min). Proteins were visualized with Coomassie blue or immunoblotting on nitrocellulose (GE Healthcare #10600002). Nitrocellulose membranes were blocked with 5 % milk in PBST (PBS, 0.1 % Tween-20) and probed with anti-FLAG antibody M2 (Sigma #F3165, 1:2000 dilution in PBST with 5 % milk) followed by monoclonal anti-mouse IgG-HRP (Santa Cruz #sc-2005; 1:5000 dilution in PBST with 5 % milk). After extensive washing with PBST, the membranes were developed using ECL detection solution (GE Healthcare #RPN2105) and exposure to photographic films (Super RX-N Fujifilm #57164152).

Sample preparation for glycopeptides

Purified proteins (between 50-200 µg) were reduced by 50 mM dithiothreitol in 50 mM AmBic buffer (ammonium bicarbonate buffer, pH 8.5) at 37°C for 1 hour on a filter device (Microcon YM-30, Millipore) and subsequently alkylated by 65 mM iodoacetamide in 50 mM AmBic buffer at 37°C in the dark for 1 hour. After extensive washing of the filter with AmBic buffer, the proteins were digested by trypsin (molar ratio 60:1, Promega #V5111) at 37°C for 16 hours. Trypsin digested peptides and glycopeptides were collected by centrifugation and vacuum-dried in a speedvac. Samples were either desalted by Zip-Tip C18 (Millipore #ZTC18S960) prior to glycosite-specific nanoHPLC-HCD-MS/MS analysis.

Glycopeptide analysis by nanoHPLC-HCD-mass spectrometry

For a glycosite-specific *N*-glycan analysis of intact glycopeptides the nanoHPLC-HCD-MS/MS method was used as previously described [27]. A calibrated LTQ-Orbitrap Fusion mass spectrometer (Thermo Scientific) coupled to an Easy-nLC™ 1000 system (Thermo Scientific) was used for sample analysis. Peptides were resuspended in 2.5% acetonitrile (ACN) with 0.1% formic acid (FA) and loaded on a self-made fritted column (75 µm × 150 mm) packed with reverse phase C18 material (AQ, 3 µm 200 Å, Bishoff GmbH, Germany) and eluted by a gradient (from 5 to 35% of solution B (99.9% ACN, 0.1% FA) for 30 min, 55% of B for 10 min, 97% of B for 10 min, with a flow rate of 300 nl/min). One scan cycle included a full scan MS survey spectrum, followed by up to 10 sequential HCD MS/MS on the most intense signals (>50,000). Full-scan MS spectra (700–2,000 m/z) and HCD MS/MS spectra were recorded in the FT-Orbitrap (resolution of 60,000 at 400 m/z for MS and 30,000 at 400 m/z for MS/MS). HCD was performed with a target value of 1e5 and stepped collision energy rolling from 22 V±10% was applied. For full Fourier Transform MS, AGC target values were 5e5. Dynamic exclusion with a single repeat count, 15 s repeat duration, and 60 s exclusion duration was used for all experiments.

Database search and glycosite-specific *N*-glycosylation analysis

MS and MS/MS data were converted into the Mascot generic format (mgf) files and searched against the SwissProt database (Version 201408) using the Mascot search engine (Version 2.4) with the consideration of carbamidomethylation at cysteine and oxidation at methionine. The monoisotopic masses of charged peptides (2+ or more) were searched with a peptide tolerance of 10 ppm and a MS/MS tolerance of 0.25 Da for fragment ions. XCalibur (Version 2.2 sp1.48) and emzed ([47] Version 2) was used for glycosite-specific *N*-glycosylation analysis and all data were inspected manually. First extractmgf was written in Perl and then glycopeptides were extracted from mgf file by identifying glycan oxonium ions ([HexNAc]⁺ 204.09 and [HexNAc+Hex]⁺ 366.12) and glycopeptides specific for each glycosite by sorting out the Y1 ions individually from extracted mgf files (Supplementary Table 3). For quantification, extracted_ion chromatography (XIC) of each glycoform was plotted by its individual m/z with the mass tolerance of 10 ppm. MS .raw files were converted to .mzML files using the program msconvert frontend version 1. Peak area was integrated using the emzed extension 'glyx' (glycosylation explorer, version glyx-0.2.5-py2-none-any.whl). The extension can be downloaded from '<http://emzed.ethz.ch/downloads/>'. All integrated peak areas were inspected manually. The relative amount of each glycoform sharing the same peptide backbone was calculated with the following equation:

$$\text{Relative amount of each glycoform (\%)} = \frac{\text{Peak area of each glycoform}}{\text{Sum of peak area of all glycoforms}} \times 100\%$$

References

1. Hoberg, E.R., J.R. Lichtenfels, and L. Gibbons, *Phylogeny for species of Haemonchus (Nematoda: Trichostrongyloidea): considerations of their evolutionary history and global biogeography among Camelidae and Pecora (Artiodactyla)*. J Parasitol, 2004. **90**(5): p. 1085-102.
2. Anderson, T.J., M.S. Blouin, and R.N. Beech, *Population biology of parasitic nematodes: applications of genetic markers*. Adv Parasitol, 1998. **41**: p. 219-83.
3. Blouin, M.S., et al., *Host movement and the genetic structure of populations of parasitic nematodes*. Genetics, 1995. **141**(3): p. 1007-14.
4. Troell, K., et al., *Global patterns reveal strong population structure in Haemonchus contortus, a nematode parasite of domesticated ruminants*. Int J Parasitol, 2006. **36**(12): p. 1305-16.
5. Prichard, R., *Genetic variability following selection of Haemonchus contortus with anthelmintics*. Trends Parasitol, 2001. **17**(9): p. 445-53.
6. Gilleard, J.S. and R.N. Beech, *Population genetics of anthelmintic resistance in parasitic nematodes*. Parasitology, 2007. **134**(Pt 8): p. 1133-47.
7. Gilleard, J.S., *Haemonchus contortus as a paradigm and model to study anthelmintic drug resistance*. Parasitology, 2013. **140**(12): p. 1506-22.
8. Dicker, A.J., et al., *Proteomic analysis of Mecistocirrus digitatus and Haemonchus contortus intestinal protein extracts and subsequent efficacy testing in a vaccine trial*. PLoS Negl Trop Dis, 2014. **8**(6): p. e2909.
9. Smith, T.S., et al., *Purification and evaluation of the integral membrane protein H11 as a protective antigen against Haemonchus contortus*. Int J Parasitol, 1993. **23**(2): p. 271-80.
10. Williamson, A.L., et al., *Digestive proteases of blood-feeding nematodes*. Trends Parasitol, 2003. **19**(9): p. 417-23.
11. Newton, S.E. and E.A. Munn, *The development of vaccines against gastrointestinal nematode parasites, particularly Haemonchus contortus*. Parasitol Today, 1999. **15**(3): p. 116-22.

12. Munn, E.A., et al., *Vaccination against Haemonchus contortus with denatured forms of the protective antigen H11*. Parasite Immunol, 1997. **19**(6): p. 243-8.
13. Smith, T.S., et al., *Cloning and characterization of a microsomal aminopeptidase from the intestine of the nematode Haemonchus contortus*. Biochim Biophys Acta, 1997. **1338**(2): p. 295-306.
14. Nisbet, A.J., et al., *Immunity to Haemonchus contortus and Vaccine Development*. Adv Parasitol, 2016. **93**: p. 353-96.
15. Reszka, N., et al., *Haemonchus contortus: characterization of the baculovirus expressed form of aminopeptidase H11*. Exp Parasitol, 2007. **117**(2): p. 208-13.
16. Cachat, E., et al., *Attempts to immunize sheep against Haemonchus contortus using a cocktail of recombinant proteases derived from the protective antigen, H-gal-GP*. Parasite Immunol, 2010. **32**(6): p. 414-9.
17. Roberts, B., et al., *Novel expression of Haemonchus contortus vaccine candidate aminopeptidase H11 using the free-living nematode Caenorhabditis elegans*. Vet Res, 2013. **44**: p. 111.
18. Haslam, S.M., et al., *Haemonchus contortus glycoproteins contain N-linked oligosaccharides with novel highly fucosylated core structures*. J Biol Chem, 1996. **271**(48): p. 30561-70.
19. Paschinger, K. and I.B. Wilson, *Two types of galactosylated fucose motifs are present on N-glycans of Haemonchus contortus*. Glycobiology, 2015. **25**(6): p. 585-90.
20. Jones, D.T., W.R. Taylor, and J.M. Thornton, *The rapid generation of mutation data matrices from protein sequences*. Comput Appl Biosci, 1992. **8**(3): p. 275-82.
21. Tamura, K., et al., *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0*. Mol Biol Evol, 2013. **30**(12): p. 2725-9.
22. Wong, A.H., D. Zhou, and J.M. Rini, *The X-ray crystal structure of human aminopeptidase N reveals a novel dimer and the basis for peptide processing*. J Biol Chem, 2012. **287**(44): p. 36804-13.
23. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis*. Nat Protoc, 2015. **10**(6): p. 845-58.

24. Librado, P. and J. Rozas, *DnaSP v5: a software for comprehensive analysis of DNA polymorphism data*. Bioinformatics, 2009. **25**(11): p. 1451-2.
25. Nielsen, R. and Z. Yang, *Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene*. Genetics, 1998. **148**(3): p. 929-36.
26. Yang, Z., et al., *Codon-substitution models for heterogeneous selection pressure at amino acid sites*. Genetics, 2000. **155**(1): p. 431-49.
27. Hang, I., et al., *Analysis of site-specific N-glycan remodeling in the endoplasmic reticulum and the Golgi*. Glycobiology, 2015. **25**(12): p. 1335-49.
28. Titz, A., et al., *Molecular basis for galactosylation of core fucose residues in invertebrates: identification of caenorhabditis elegans N-glycan core alpha1,6-fucoside beta1,4-galactosyltransferase GALT-1 as a member of a novel glycosyltransferase family*. J Biol Chem, 2009. **284**(52): p. 36223-33.
29. Jasmer, D.P., M.D. Mitreva, and J.P. McCarter, *mRNA sequences for Haemonchus contortus intestinal cathepsin B-like cysteine proteases display an extreme in abundance and diversity compared with other adult mammalian parasitic nematodes*. Mol Biochem Parasitol, 2004. **137**(2): p. 297-305.
30. Ruiz, A., et al., *Genetic variability in cysteine protease genes of Haemonchus contortus*. Parasitology, 2004. **128**(Pt 5): p. 549-59.
31. Martin, S., et al., *Influence of immunoprotection on genetic variability of cysteine proteinases from Haemonchus contortus adult worms*. Int J Parasitol, 2015. **45**(13): p. 831-40.
32. Denver, D.R., et al., *High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome*. Nature, 2004. **430**(7000): p. 679-82.
33. Redman, E., et al., *Microsatellite analysis reveals marked genetic differentiation between Haemonchus contortus laboratory isolates and provides a rapid system of genetic fingerprinting*. Int J Parasitol, 2008. **38**(1): p. 111-22.
34. Hunt, P.W., et al., *Genetic and phenotypic differences between isolates of Haemonchus contortus in Australia*. Int J Parasitol, 2008. **38**(8-9): p. 885-900.

35. Otsen, M., et al., *Amplified fragment length polymorphism analysis of genetic diversity of Haemonchus contortus during selection for drug resistance*. Int J Parasitol, 2001. **31**(10): p. 1138-43.
36. Maizels, R.M. and A. Kurniawan-Atmadja, *Variation and polymorphism in helminth parasites*. Parasitology, 2002. **125 Suppl**: p. S25-37.
37. Qiang, S., et al., *Variation between ASP-1 molecules from Ancylostoma caninum in China and the United States*. J Parasitol, 2000. **86**(1): p. 181-5.
38. Munn, E.A., *Rational design of nematode vaccines: hidden antigens*. Int J Parasitol, 1997. **27**(4): p. 359-66.
39. Smith, W.D., J.A. van Wyk, and M.F. van Strijp, *Preliminary observations on the potential of gut membrane proteins of Haemonchus contortus as candidate vaccine antigens in sheep on naturally infected pasture*. Veterinary Parasitology, 2001. **98**(4): p. 285-297.
40. Meier, L., P.R. Torgerson, and H. Hertzberg, *Vaccination of goats against Haemonchus contortus with the gut membrane proteins H11/H-gal-GP*. Veterinary Parasitology, 2016. **229**: p. 15-21.
41. Andrews, S.J., T.P. Rolph, and E.A. Munn, *Duration of protective immunity against ovine haemonchosis following vaccination with the nematode gut membrane antigen H11*. Res Vet Sci, 1997. **62**(3): p. 223-7.
42. Pancera, M., et al., *Structure and immune recognition of trimeric pre-fusion HIV-1 Env*. Nature, 2014. **514**(7523): p. 455-61.
43. Wu, X., et al., *Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1*. Science (New York, N.Y.), 2010. **329**(5993): p. 856-861.
44. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability*. Mol Biol Evol, 2013. **30**(4): p. 772-80.
45. Nei, M. and T. Gojobori, *Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions*. Mol Biol Evol, 1986. **3**(5): p. 418-26.

46. Munn, E.A. and T.S. Smith, *Production and use of anthelmintic agents and protective immunogens*. 1990, Google Patents.
47. Kiefer, P., U. Schmitt, and J.A. Vorholt, *eMZed: an open source framework in Python for rapid and interactive development of LC/MS data analysis workflows*. *Bioinformatics*, 2013. **29**(7): p. 963-4.

Supplementary Data

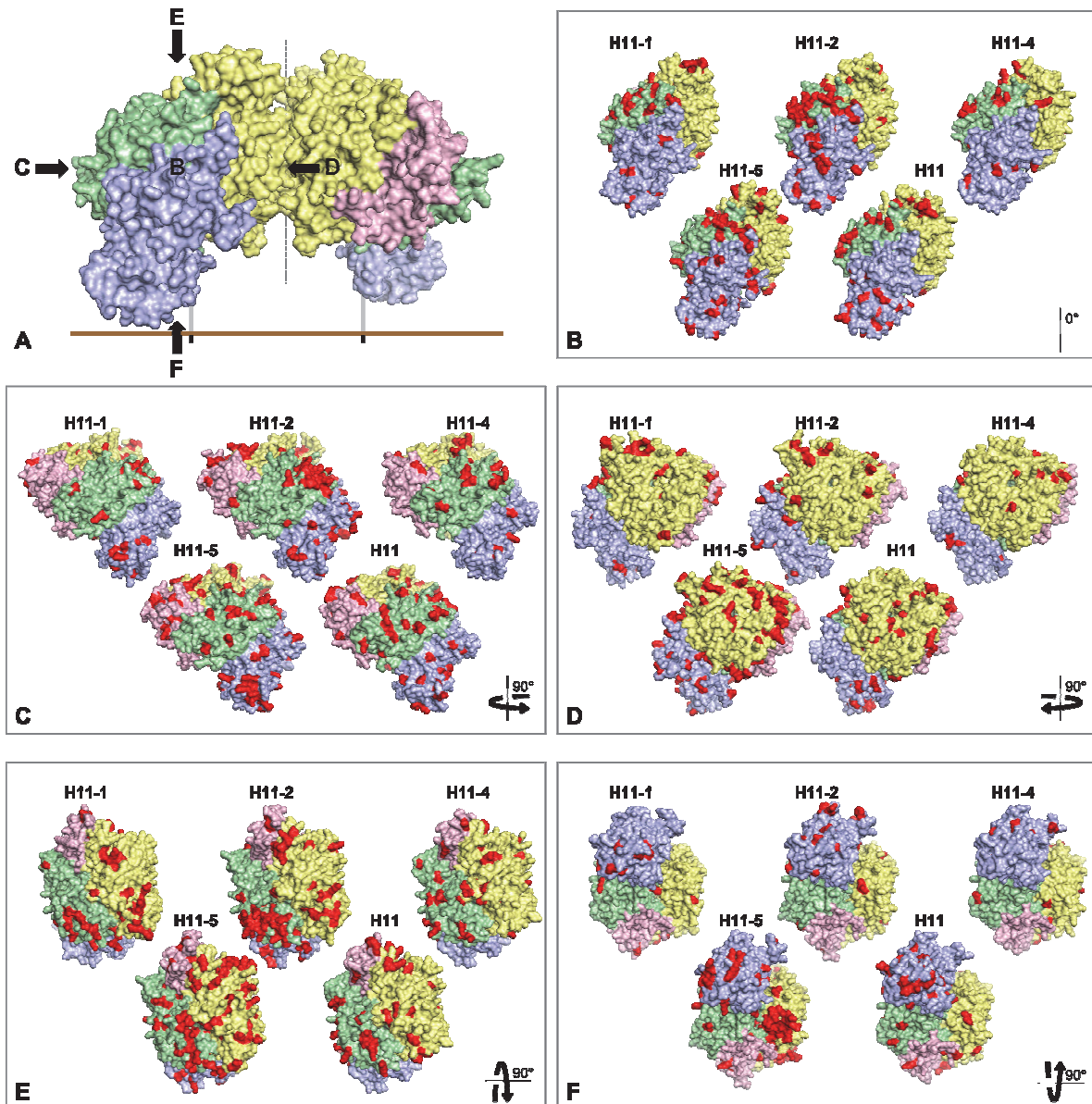


Figure S1: Different views on the H11 amino acid surface diversity

(A) Dimeric CH-H11-5A (Swiss isolate of H11-5 variant A) protein structure modeled by PHYRE2 [23] on the crystal structure of the human aminopeptidase N (hAPN [22]). The dimer is shown in a PYMOL surface representation and colored according to domains (I: lightblue, II: palegreen, III: lightpink, IV: paleyellow). The black arrows indicate the view on the protein surface for figures (B) – (F). Different views on the modeled monomeric protein structures of CH-H11-1D, CH-H11-2A, ZA-H11-4A (South African isolate of H11-4 variant A), CH-H11-5B and CH-H11-D with rotation details indicated in the bottom right corner: (B) no rotation (C) side view (D) side view (dimeric interface) (E) top view (F) bottom view. Amino acids highlighted in red show lack of conservation (<100%) for these positions in alignments with all sequences analyzed for one isoform (see Figure 2A).

A

		Site 3			
H11-5	Publ	agt	tgg	act	SWT
	CH	.a.	<u>NWT</u>
	SCO_F	---
	SCO_Y	.a.	<u>NWT</u>
	ZA	---
H11	Publ	aac	tgg	acc	<u>NWT</u>
	CH-C	<u>NWT</u>
	CH-L	.t	<u>NWT</u>
	SCO_K	<u>NWT</u>
	SCO_N	.g.	S--
	ZA-G	<u>NWT</u>
	ZA-H	.g.	S--

B

		Site 4				
H11-2	Publ	aat	ggg	gag	NGE	
	CH-A	---	
	CH-E	.a	K--	
	SCO-A	---	
	SCO-B	a.c	<u>NGT</u>	
	ZA-A	---	
	ZA-B	.a	K--	

C

		Site 14				
H11-5	Publ	aaa	tta	tcg	KLS	
	CH	---	
	SCO-F	---	
	SCO-Y	.t	<u>NLS</u>	
	ZA	---	

Figure S2: Additional H11 N-glycosylation site diversity

H11 isoform glycosite diversity is created by nucleotide mutations leading to amino acid replacements in the *N*-glycosylation sequon (bold type and underlined) in different isoforms (vertical text) and in different populations (CH, SCO, ZA) are shown here. This single nucleotide polymorphism leads to the loss or gain of a glycosites 3 (A), 4 (B) and 14 (C), when compared to the published sequence (Publ). Localization of these glycosites on the H11 protein surface can be found in Figure 4.

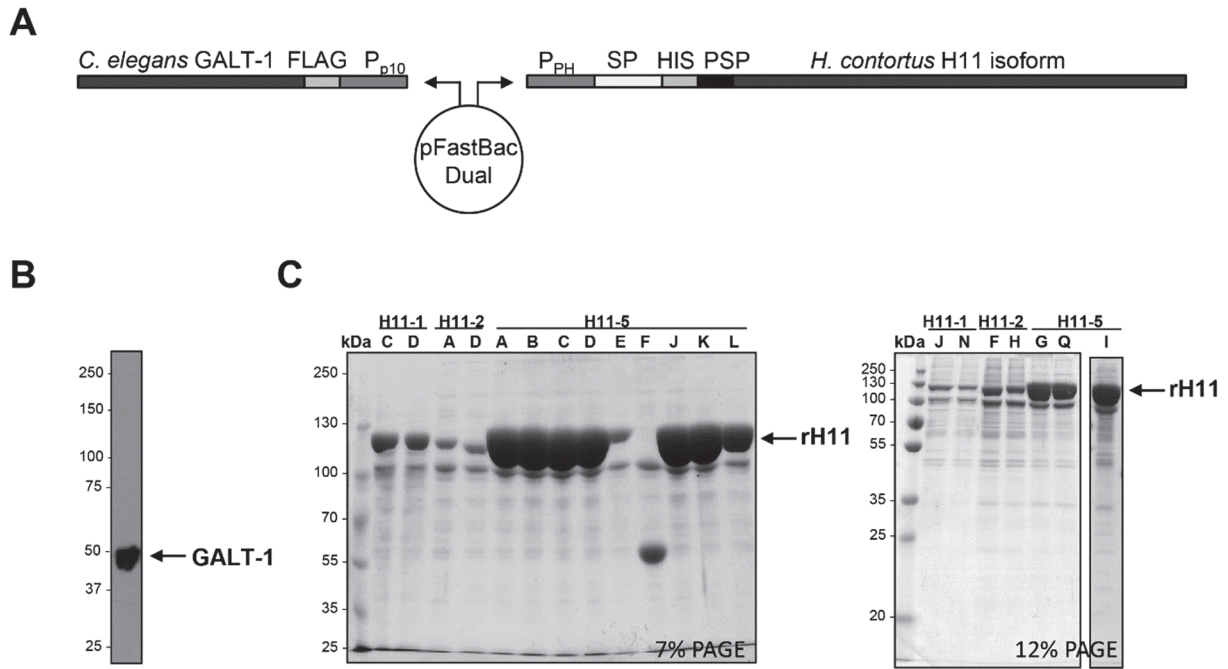


Figure S3: Purification of recombinant *H. contortus* H11 isoforms from glycoengineered High five insect cells

(A) Schematic representation of the H11-GALT-1 expression constructs cloned into pFastBacDual vector. H11 is cloned downstream of the Polyhedrin promoter (P_{PH}) with a gp67 secretion signal peptide (SP), a His₁₀-tag (HIS), and a PreScission Protease cleavage site (PSP) between the N-terminal tag and the respective gene. *C. elegans* galactosyltransferase GALT-1 is cloned downstream of the p10 promoter (P_{p10}) and a FLAG tag. **(B)** A representative GALT-1 expression in High five cells co-expressing GALT-1 and H11, which was confirmed by immunoblotting with anti-FLAG antibody. 1 Mio cells were harvested and lysed 72 hours post co-infection with *H. contortus* proteins. **(C)** Affinity purified secreted recombinant *H. contortus* H11 isoforms from culture supernatant 72 hours post co-expression with GALT-1. Elution samples were analyzed by SDS-PAGE and stained with Coomassie.

Table S1: Maximum Likelihood fits of 48 different amino acid substitution models

Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL) are also presented. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. For estimating ML values, a tree topology was automatically computed. The analysis involved 137 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 901 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [21].

Model	#Param	BIC	AICc	lnL	Invariant	Gamma
JTT+G	272	27320.4931	24676.9107	-12065.8524	n/a	0.7110
JTT+G+I	273	27332.0295	24678.7324	-12065.7589	0.0298	0.7615
WAG+G	272	27339.3570	24695.7746	-12075.2844	n/a	0.7599
WAG+G+I	273	27350.8271	24697.5300	-12075.1577	0.0349	0.8258
LG+G	272	27379.1395	24735.5571	-12095.1756	n/a	0.6761
JTT+G+F	291	27386.3416	24558.1872	-11987.4036	n/a	0.7054
LG+G+I	273	27390.8378	24737.5408	-12095.1630	0.0107	0.6921
JTT+G+I+F	292	27397.8603	24559.9918	-11987.3011	0.0312	0.7578
WAG+G+F	291	27437.4469	24609.2924	-12012.9562	n/a	0.7475
WAG+G+I+F	292	27448.9178	24611.0494	-12012.8299	0.0349	0.8117
LG+G+F	291	27500.0755	24671.9210	-12044.2705	n/a	0.6710
LG+G+I+F	292	27511.7405	24673.8720	-12044.2413	0.0163	0.6954
WAG+I	272	27578.5808	24934.9984	-12194.8963	0.3145	n/a
JTT+I	272	27588.6967	24945.1143	-12199.9542	0.3187	n/a
JTT+I+F	291	27654.8301	24826.6756	-12121.6478	0.3194	n/a
Dayhoff+G	272	27675.3128	25031.7303	-12243.2623	n/a	0.7226
WAG+I+F	291	27681.3693	24853.2149	-12134.9174	0.3167	n/a
Dayhoff+G+I	273	27686.2672	25032.9702	-12242.8777	0.0567	0.8304
LG+I	272	27686.9635	25043.3810	-12249.0876	0.3185	n/a
rtREV+G+F	291	27697.4584	24869.3039	-12142.9619	n/a	0.6536
rtREV+G	272	27700.2916	25056.7092	-12255.7517	n/a	0.6778
rtREV+G+I+F	292	27709.0430	24871.1746	-12142.8925	0.0249	0.6901
rtREV+G+I	273	27711.9350	25058.6379	-12255.7116	0.0190	0.7070
Dayhoff+G+F	291	27714.4456	24886.2912	-12151.4556	n/a	0.7063
cpREV+G	272	27714.8751	25071.2927	-12263.0435	n/a	0.7461
Dayhoff+G+I+F	292	27725.4597	24887.5912	-12151.1009	0.0546	0.8058
cpREV+G+I	273	27725.4741	25072.1770	-12262.4812	0.0707	0.8846
cpREV+G+F	291	27794.5107	24966.3562	-12191.4881	n/a	0.7458
LG+I+F	291	27803.7968	24975.6423	-12196.1311	0.3219	n/a
cpREV+G+I+F	292	27804.7877	24966.9193	-12190.7649	0.0811	0.9108
mtREV24+G+F	291	27920.9232	25092.7688	-12254.6944	n/a	0.6496
WAG	271	27932.4521	25298.5843	-12377.6937	n/a	n/a
mtREV24+G+I+F	292	27932.6468	25094.7784	-12254.6944	0.0000	0.6496
Dayhoff+I	272	27944.6613	25301.0788	-12377.9365	0.3066	n/a
cpREV+I	272	27958.2498	25314.6673	-12384.7308	0.3091	n/a
JTT	271	27967.2183	25333.3505	-12395.0768	n/a	n/a
Dayhoff+I+F	291	27988.6766	25160.5222	-12288.5711	0.3105	n/a
rtREV+I	272	28010.9049	25367.3224	-12411.0583	0.3164	n/a
rtREV+I+F	291	28013.6193	25185.4648	-12301.0424	0.3226	n/a
cpREV+I+F	291	28029.9552	25201.8007	-12309.2104	0.3103	n/a
JTT+F	290	28036.9291	25218.4887	-12318.5590	n/a	n/a
WAG+F	290	28041.8370	25223.3966	-12321.0130	n/a	n/a
LG	271	28067.2329	25433.3651	-12445.0840	n/a	n/a
LG+F	290	28192.0060	25373.5656	-12396.0975	n/a	n/a
mtREV24+I+F	291	28278.3336	25450.1791	-12433.3996	0.2965	n/a
cpREV	271	28328.4239	25694.5561	-12575.6796	n/a	n/a
Dayhoff	271	28340.7116	25706.8438	-12581.8234	n/a	n/a
Dayhoff+F	290	28390.4053	25571.9649	-12495.2972	n/a	n/a
rtREV	271	28402.4356	25768.5678	-12612.6854	n/a	n/a
cpREV+F	290	28410.4876	25592.0472	-12505.3383	n/a	n/a
rtREV+F	290	28423.6037	25605.1633	-12511.8964	n/a	n/a
mtREV24+F	290	28661.8907	25843.4502	-12631.0398	n/a	n/a
mtREV24+G	272	28870.0778	26226.4953	-12840.6448	n/a	0.5626
mtREV24+G+I	273	28881.8010	26228.5040	-12840.6447	0.0000	0.5626
mtREV24+I	272	29376.6387	26733.0563	-13093.9253	0.2922	n/a
mtREV24	271	29797.5749	27163.7071	-13310.2551	n/a	n/a

Table S2: Positively selected amino acid sites with posterior probability (P) values of > 0.95 from the likelihood ratio test (Table 3)

Protein Reference	Total	Positively selected codons
		(Pr($\omega > 1$), *: P>95%; **: P>99%, post mean +- SE for ω)
H11_X94187.1	15	55 T (0.998**, 2.486 +- 0.129), 139 D (0.959*, 2.416 +- 0.374), 146 S (0.983*, 2.459 +- 0.252), 175 G (0.974*, 2.442 +- 0.308), 247 V (0.989*, 2.470 +- 0.211), 286 E (1.000**, 2.489 +- 0.103), 290 G (0.964*, 2.425 +- 0.352), 307 Y (0.992**, 2.475 +- 0.190), 310 Q (0.998**, 2.486 +- 0.132), 487 H (0.955*, 2.407 +- 0.393), 637 K (0.988*, 2.469 +- 0.218), 717 T (1.000**, 2.490 +- 0.101), 740 R (0.997**, 2.485 +- 0.134), 780 N (0.966*, 2.427 +- 0.350), 793 R (0.992**, 2.476 +- 0.187)
H11_1_AJ249941.1	8	317 E (1.000**, 4.597 +- 0.797), 691 K (1.000**, 4.597 +- 0.797), 742 E (0.998**, 4.587 +- 0.817), 749 S (1.000**, 4.597 +- 0.797), 762 D (1.000**, 4.597 +- 0.797), 766 G (1.000**, 4.597 +- 0.797), 782 E (0.974*, 4.481 +- 0.992), 789 R (1.000**, 4.597 +- 0.797)
H11_2_AJ249942.2	12	109 S (0.956*, 2.428 +- 0.396), 182 N (0.974*, 2.461 +- 0.309), 195 T (0.976*, 2.465 +- 0.301), 198 T (1.000**, 2.509 +- 0.103), 286 K (1.000**, 2.509 +- 0.097), 302 K (0.995**, 2.501 +- 0.155), 306 A (0.951*, 2.417 +- 0.418), 645 K (0.979*, 2.471 +- 0.280), 731 S (0.992**, 2.496 +- 0.184), 743 T (0.996**, 2.502 +- 0.148), 781 K (0.999**, 2.508 +- 0.107), 784 A (1.000**, 2.509 +- 0.101)
H11_4_AJ311316.1	17	96 E (0.992**, 6.022 +- 1.098), 419 Y (0.991**, 6.021 +- 1.100), 613 S (0.977*, 5.938 +- 1.275), 636 R (1.000**, 6.066 +- 0.986), 667 V (0.973*, 5.916 +- 1.315), 683 V (1.000**, 6.066 +- 0.986), 684 N (0.966*, 5.880 +- 1.379), 696 S (0.988*, 6.002 +- 1.145), 713 P (0.984*, 5.980 +- 1.193), 732 G (0.967*, 5.884 +- 1.371), 733 Y (1.000**, 6.068 +- 0.983), 739 K (0.986*, 5.991 +- 1.169), 742 Q (1.000**, 6.068 +- 0.983), 743 L (1.000**, 6.068 +- 0.983), 745 S (0.988*, 6.002 +- 1.146), 759 A (0.988*, 6.002 +- 1.145), 801 T (1.000**, 6.068 +- 0.983)
SCO_H11_5_Y	3	550 D (0.998**, 1.921 +- 0.576), 566 V (0.979*, 1.895 +- 0.589), 853 E (0.951*, 1.852 +- 0.607)

Table S3: Overview of the N-glycosylation sites of Swiss H11 isoforms

Trypsin digested glycopeptides with carbamidomethylated cysteines are summarized here (not detected glycopeptides are marked in red)

Overall Glycosite	H11-1	H11-2	H11-4	H11-5	H11	Trypsin digested glycopeptides sequence	Glycosite found in Constructs	peptide [M+H] ⁺	Y1	charge state
1	S1					NLTFDAHVEISMVVVEPTNSIVLNSK	2A	2856.4812	3059.5680	-
						NLTFDAHVEIAMVVVEPTNSIVLNSK	2D, 2H	2840.4863	3043.5731	-
						NLTFDAHVEIAMVVVEPTNSIVLNSK	2F	2854.5020	3057.5888	-
						NLTFDGR	4C	822.4110	1025.4978	2
2	S1				NLTFDGGVGLSLR	5A, 5B, 5C, 5D, 5E, 5F, 5G, 5I, 5K, 5Q	1419.7596	1622.8464	-	
					NLTFDGGVGLSK	5J, 5L	1391.7535	1594.8403	-	
					TYLPGVDFPPEKNLTFDGR	H11 published	2329.1505	2532.2373	3	
3	S2	S3			NISVIADQCELFSSNQK	1C, 1D, 1I, 1N	1979.9282	2183.0150	3	
					NITVIPDK	5A, 5B, 5C, 5D, 5E, 5F, 5G, 5I, 5J, 5K, 5L, 5Q	899.5203	1102.6071	2	
5	S2	S3			ANWTVTVIHPK	1C, 1D, 1I, 1N, 4C, 5A, 5B, 5D, 5F, 5I, 5K, H11 published	1265.7006	1468.7874	2 / 3	
					ANWTVTVIHR	5C, 5E, 5G, 5I, 5L, 5Q	1293.7068	1496.7936	3	
					YIENYTK	1C, 1D	930.4573	1133.5441	2	
6	S4				NIMTYGAR	5B	812.3725	1015.4593	2	
					IWSRPEAMNMTGYAR	5C, 5D, 5E, 5I, 5L, 5Q	1782.8420	1985.9288	3	
					IWSRPEAMNMTGYAK	5A, 5F	1754.8359	1957.9227	3	
					IWSRPEAMNMTGYADGIR	5K	2196.0331	2399.1199	-	
7	S2	S3			MSEFAPQWTTQMGGFPLTVESVNATTLK	1C, 1D, 1J	3113.5328	3316.6196	3	
					MSEFAPQWTTQMGGFPLTVESVNATTLVETQK	1N	3570.7495	3773.8363	4	
					VEEFNATSLK	2A, 2D, 2F, 2H	1137.5787	1340.6655	2	
					TTAFADQWTTQMGGFPLTVAEAFNATSVK	4C	3061.4981	3264.5849	3	
8	S3	S5			AFNATSLQITQR	5A, 5B, 5C, 5D, 5E, 5G, 5I, 5J, 5K, 5L, 5Q	1450.7655	1653.8523	2	
					TTEFASQWTTQMGGFPLTVISVAEFNSTTLK	H11 published	3121.5192	3324.6060	3	
10	S6				AYMMSILEPMYKSSIDYVK	2A	2496.2223	2699.3091	3	
					AYMMSILEPMYKSDIDYVK	2D, 2F, 2H	2524.2173	2727.3041	3	
11	S7				YIADNYKNDLSLFFEMNLQK	5A, 5B, 5C, 5D, 5E, 5G, 5I, 5K, 5L, 5Q	2353.1175	2556.2043	3	
					YIADNYKNDLSLFFEMNLR	5J	2253.0645	2456.1513	3	
					ECIKNYTDLFVEEVMIK	5A, 5B, 5C, 5D, 5E, 5G, 5I, 5J, 5Q	2017.9395	2221.0263	3	
12	S4	S8			ECIKNYTDLFVEQVMK	5K	2016.9555	2220.0423	3	
					ECIKNYTDLFQKEVMIK	5L	2032.9504	2236.0372	-	
13	S5	S4	S8		CVNVSAPLR	2F, 2H	1015.5139	1218.6007	2	
					NSSFVR	1C, 1D, 1I, 1N, 4C, 5A, 5B, 5C, 5D, 5E, 5G, 5I, 5J, 5K, 5L, 5Q	709.3634	912.4502	2	
15	S5	S5			ALDRNSSFVR	H11 published	1164.6126	1367.6994	2	
					LASFFKNATW	4C	1184.6105	1387.6973	2	

Table S4: Quantification of glycosite-specific N-glycosylation profiles of H11 isoforms

H11 isoforms were purified from insect cells co-expressing the galactosyltransferase-1 GALT-1. Trypsin digested glycopeptides were analyzed by nanoHPLC-HCD MS/MS mass spectrometry. Varying N-glycosylation sites are found per isoform: 4 - 5 for H11-1, 3 - 4 for H11-2, 5 for H11-4, 6 - 8 for H11-5, and 3 - 5 for H11. Glycan structures found on each glycosylation site are listed with their respective percentage. Data represent the mean values of glycan ratios of two (5E, 5F and 5I) or three (all others) different experiments. The glycan structures were summarized into the following groups: High Mannose (sum of 1-5), Hybrid (sum of 6-10), Paucimannose (sum of 11-13), Core fucosylation (sum of 14-17 + 22-28), Core difucosylation (sum of 18-21 + 29-31). In addition, the core galactosylation is the sum of glycan structures 22-31. SD indicates the standard deviation that is calculated by the squareroot of the summarized variance.

Glycan structure	H11-1													H11-2							H11															
	2			3			4			5			6			7			8			9			10			11			12			13		
	C	D	J	N	C	D	J	N	C	D	J	N	C	D	J	N	C	D	J	N	C	D	J	N	A	D	F	H	1	3	7	13				
HexNAc2Hex9	1.02	1.14	0.97	0.30	1.23	2.09	1.14	0.45	0.24	0.33	0.00	0.41	0.23	1.35	1.84	1.35	0.13	12.52	14.22	0.28	0.29	0.92	0.65	8.73	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
HexNAc2Hex8	15.10	13.09	10.52	4.41	1.33	2.18	1.38	0.74	0.72	0.99	1.55	2.04	0.80	5.77	6.91	6.64	2.06	27.61	12.92	1.78	1.98	2.05	1.83	2.24	1.87	35.83	0.00	0.00	0.00	0.00	0.00					
HexNAc2Hex7	44.91	39.39	42.55	36.78	1.14	1.66	1.26	0.67	1.31	1.88	3.05	2.33	0.98	7.84	8.44	8.44	3.72	19.71	63.11	4.99	3.85	3.98	3.65	4.10	4.00	22.29	13.52	0.00	0.00	0.00	0.00					
HexNAc2Hex6	20.01	24.91	28.92	43.09	1.86	2.30	2.00	1.32	1.46	1.96	2.79	1.74	9.24	9.99	8.58	4.93	7.31	7.92	4.44	5.09	4.15	3.74	4.33	4.19	10.13	4.08	0.00	0.00	0.00	0.00	0.00					
HexNAc2Hex5	1.91	2.83	2.30	1.87	2.92	3.03	3.85	2.24	1.96	2.39	4.11	1.87	2.09	22.25	20.66	23.19	18.44	3.45	0.42	2.78	7.18	2.54	2.44	2.70	2.41	6.33	4.48	0.00	0.00	0.00	0.00					
HexNAc2Hex4HexNAc	0.00	0.00	0.00	0.00	0.47	0.32	0.24	0.33	0.00	0.00	0.00	1.77	0.76	0.26	0.11	0.11	0.05	1.22	0.07	2.07	0.00	2.72	2.63	3.10	2.88	0.00	0.00	0.00	0.00	0.00	0.00					
HexNAc2Hex5HexNAc	0.00	0.00	0.00	0.00	7.15	8.09	10.71	9.26	0.40	0.53	1.49	1.18	5.42	4.24	5.30	3.48	0.80	0.43	1.92	2.12	3.33	3.34	3.84	3.62	0.63	5.85	0.00	0.00	0.00	0.00	0.00					
HexNAc2Hex4HexNAc	1.97	2.45	2.06	2.22	20.79	18.68	16.18	12.85	0.36	0.39	1.13	1.05	3.30	2.38	1.43	2.34	1.43	0.23	0.90	1.51	2.78	2.92	3.56	3.15	0.00	14.05	0.00	0.00	0.00	0.00	0.00					
(HexNAc2Hex3)HexNAc	12.30	13.06	9.86	9.16	26.53	24.68	12.14	13.35	1.38	1.39	0.00	4.44	2.86	14.13	18.47	17.83	15.52	19.72	1.12	3.60	6.29	3.57	3.95	4.33	4.53	11.08	5.88	0.00	0.00	0.00	0.00					
(HexNAc2Hex3)HexNAc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
HexNAc2Hex4	0.00	0.00	0.00	0.00	2.83	2.60	5.28	3.16	0.00	0.00	0.00	0.94	0.56	0.59	0.13	0.07	0.11	0.00	0.00	0.00	0.00	1.31	1.34	1.59	1.31	0.00	10.67	0.00	0.00	0.00	0.00	0.00				
HexNAc2Hex3	1.37	1.36	1.00	0.82	24.82	24.79	36.07	47.32	1.37	1.50	17.65	11.21	11.21	0.98	1.10	1.05	1.38	6.24	1.14	11.10	9.69	12.76	14.91	14.09	14.33	4.96	35.69	13.12	0.00	0.00	0.00	0.00				
HexNAc2Hex2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.20	1.56	0.80	0.93	0.72	1.75	3.02	6.92	0.00	0.24	0.19	0.23	0.21	0.20	0.00	5.78	0.00	0.00	0.00	0.00	0.00				
HexNAc2Hex2Fuc	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67	1.44	1.48	1.22	0.68	2.56	0.00	0.00	0.21	5.66	0.15	0.13	0.11	0.12	0.00	0.00	4.28	4.48	0.00	0.00	0.00	0.00			
(HexNAc2Hex3)Fuc	1.40	1.77	1.81	1.34	0.77	0.60	0.72	0.75	2.87	2.85	33.12	22.43	23.19	1.47	1.17	1.13	1.02	0.00	0.02	30.53	5.15	29.92	29.24	28.20	29.54	0.00	0.00	35.49	7.15	0.00	0.00	0.00	0.00			
(HexNAc2Hex3)HexNAcFuc	0.00	0.00	0.00	0.00	1.64	1.41	0.82	0.64	0.64	0.74	6.01	5.25	4.61	2.97	3.11	3.08	0.00	0.05	5.85	1.56	3.76	3.73	4.15	4.79	0.00	0.00	0.00	4.50	0.00	0.00	0.00	0.00	0.00			
(HexNAc2Hex3)HexNAc2Fuc	0.00	0.00	0.00	0.00	0.06	0.09	0.07	0.05	0.00	0.00	0.00	0.92	0.60	0.38	0.17	0.11	0.16	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
HexNAc2Hex2Fuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.03	0.00	0.14	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.34	0.38	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
(HexNAc2Hex3)Fuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.18	3.19	3.82	2.06	4.50	0.12	0.09	0.10	0.15	0.00	0.00	1.31	0.90	1.52	1.42	1.12	1.27	0.00	25.80	24.18	0.00	0.00	0.00	0.00	0.00	0.00		
(HexNAc2Hex3)HexNAcFuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.30	0.00	3.27	4.52	0.07	0.05	0.10	0.11	0.00	0.00	0.00	0.03	0.02	0.02	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
(HexNAc2Hex3)HexNAc2Fuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.68	0.46	0.04	0.02	0.04	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
(HexNAc2Hex3)FucHex	0.00	0.00	0.00	0.00	3.31	3.56	4.02	3.12	14.74	18.26	6.98	5.41	7.10	1.30	1.10	0.97	2.35	0.00	0.13	5.72	17.97	10.67	10.60	9.32	8.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
(HexNAc2Hex3)HexNAcFucHex	0.00	0.00	0.00	0.00	2.50	2.79	2.76	2.72	12.04	10.07	0.00	3.47	2.28	9.45	8.72	7.24	14.14	0.00	0.08	2.55	4.53	1.76	1.63	1.77	1.74	0.00	0.00	6.10	0.00	0.00	0.00	0.00	0.00	0.00		
(HexNAc2Hex3)HexNAc2FucHex	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.31	0.00	1.94	1.22	0.48	0.17	0.13	0.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
(HexNAc2Hex3)FucHex2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.05	1.03	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.00	1.01	0.77	0.81	0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
(HexNAc2Hex3)FucHex3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.43	0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.00	0.22	0.17	0.19	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
(HexNAc2Hex3)HexNAcFucHex2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.66	1.74	0.52	0.22	0.17	0.35	0.00	0.00	1.24	1.09	1.20	1.17	1.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)HexNAcFucHex3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.09	3.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.04	0.88	0.93	0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)Fuc2Hex	0.00	0.00	0.00	0.00	0.71	0.90	1.15	0.84	40.34	41.88	15.56	11.20	17.90	2.65	2.47	4.61	0.00	0.07	16.88	17.39	7.91	7.99	6.48	6.79	0.00	0.00	21.30	30.87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)HexNAcFuc2Hex	0.00	0.00	0.00	0.00	0.15	0.23	0.22	0.19	13.65	10.15	1.57	1.07	0.76	5.88	4.47	4.42	9.05	0.04	1.26	1.62	0.32	0.30	0.34	0.30	0.34	0.00	0.00	6.77	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)HexNAc2Fuc2Hex	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.07	0.65	0.00	1.12	0.32	1.83	0.89	0.76	2.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
High Mannose	82.85	81.36	85.26	86.45	8.28	11.25	9.62	5.42	5.69	7.56	11.51	9.43	5.64	46.45	47.85	46.19	29.29	70.59	98.59	14.28	18.39	13.54	12.34	14.28	13.13	83.32	22.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
Hybrid	14.28	15.51	11.93	11.38	54.93	51.76	39.27	35.79	2.15	2.31	1.49	9.12	5.86	23.10	25.20	26.39	21.39	23.17	0.84	8.48	9.92	12.40	12.84	14.83	14.18	11.72	25.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
Paucimannose	1.37	1.36	1.00	0.82	27.66	27.39	41.35	50.48	1.60	1.71	20.14	12.58	12.73	1.83	3.01	4.18	8.41	6.24	0.18	11.95	16.91	14.26	15.89	15.84	4.96	52.14	13.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
Core fucosylation	1.40	1.77	1.81	1.34	8.28	8.46	8.39	7.27	30.77	32.23	46.11	49.33	47.02	18.05	16.02	13.55	24.28	0.00	0.28	45.83	34.88	49.78	48.24	46.69	48.12	0.00	0.00	39.78	27.45	0.00	0.00	0.00	0.00	0.00	0.00	
Core difucosylation	0.00	0.00	0.00	0.00	0.00	0.85	1.14	1.37	1.04	59.78	56.20	20.75	19.54	28.75	10.58	7.92	7.70	16.63	0.00	0.11	19.46	19.11	10.02	10.10	8.31	8.73	0.00	0.00	41.30	27.55	0.00	0.00	0.00	0.00	0.00	0.00

Glycan structure	2												3												
	A	B	C	D	E	F	G	I	J	K	L	Q	A	B	C	D	E	F	G	I	J	K	L	Q	
HexNAc2Hex9	1.45	1.40	1.17	1.00	1.68	2.02	2.43	2.73	2.19	1.60	1.62	1.97	0.95	0.94	0.98	1.19	3.15	3.64	2.22	1.61	1.50	0.90	1.74	3.28	
HexNAc2Hex8	3.66	3.44	2.97	2.32	4.23	4.67	6.37	4.61	6.10	3.55	4.38	5.27	1.29	1.45	1.67	1.95	5.10	7.00	3.68	2.22	2.34	1.31	3.28	5.00	
HexNAc2Hex7	35.00	35.43	34.53	32.74	18.40	21.34	52.54	48.42	48.20	37.96	43.93	55.15	30.91	34.95	45.81	45.10	25.42	12.05	48.14	50.07	43.12	30.65	47.14	64.36	
HexNAc2Hex6	28.69	21.06	22.09	32.67	4.87	5.31	17.62	28.88	19.47	35.58	31.79	17.43	12.97	12.54	10.71	12.94	8.43	6.05	9.04	10.17	14.28	11.71	12.20	10.71	
HexNAc2Hex5	4.77	5.12	4.90	3.66	1.37	2.90	1.58	1.25	4.34	3.79	2.55	1.80	4.41	4.26	6.92	5.89	4.44	4.86	4.50	3.33	5.58	5.62	3.32	2.70	
HexNAc2Hex4	0.96	1.05	1.09	0.50	0.72	0.97	0.90	0.31	1.45	0.55	0.53	0.97	3.60	2.90	2.08	2.04	1.44	1.63	1.99	2.84	2.62	3.90	2.50	0.91	
HexNAc2Hex3	11.76	9.52	9.00	4.39	2.46	4.97	6.43	4.38	4.62	4.96	3.76	4.43	3.82	5.00	10.27	8.94	4.90	4.68	2.91	5.98	3.96	4.27	9.10	3.94	2.14
HexNAc2Hex2	8.64	11.13	10.61	8.34	5.14	8.01	5.22	3.95	3.74	4.43	3.82	5.00	10.27	8.94	4.90	4.68	2.91	5.98	3.96	4.27	4.96	9.10	3.94	2.14	
HexNAc2Hex1	3.02	7.72	8.25	9.63	39.04	29.31	4.14	3.68	5.18	5.01	4.88	4.58	4.32	5.64	2.14	2.14	23.55	34.67	1.72	1.72	4.98	4.56	2.85	1.47	
(HexNAc2Hex3)HexNAc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)HexNAc2Hex4	0.37	0.73	1.62	0.55	0.14	0.49	0.11	0.16	0.57	0.15	0.15	0.09	1.09	1.02	2.94	1.97	0.83	4.40	1.70	1.48	0.49	1.49	0.44	0.69	
(HexNAc2Hex3)HexNAc2Hex3	1.68	3.40	3.78	4.19	21.94	20.01	2.65	1.62	4.14	2.02	2.59	2.39	8.61	7.83	5.73	5.32	14.22	12.39	4.92	4.43	3.65	7.79	2.68	2.28	
(HexNAc2Hex3)HexNAc2Hex2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)HexNAc2Hex1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)Fuc	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.07	0.05	0.07	0.26	0.54	0.04	0.06	0.06	0.04	0.04	0.06	
(HexNAc2Hex3)HexNAcFuc	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.09	0.02	0.32	0.20	0.48	0.02	0.03	0.12	0.09	0.02	0.03	
(HexNAc2Hex3)HexNAc2Fuc	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.52	0.53	0.19	0.52	0.14	0.06	0.10	0.10	1.11	0.87	0.54	0.08	
(HexNAc2Hex3)Fuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)Fuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)HexNAcFuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)HexNAc2Fuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.11	0.22	0.27	1.29	1.18	0.23	0.15	0.33	0.19	0.21	0.48	
(HexNAc2Hex3)FucHex	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.32	0.12	0.19	1.80	2.29	0.11	0.08	0.57	0.34	0.20	0.31	
(HexNAc2Hex3)HexNAcFucHex	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)HexNAc2FucHex	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)FucHex2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)FucHex3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)HexNAcFucHex2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)HexNAcFucHex3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
(HexNAc2Hex3)Fuc2Hex	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.08	0.05	0.09	0.75	0.66	0.06	0.04	0.13	0.06	0.08	0.09	
(HexNAc2Hex3)HexNAcFuc2Hex	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.04	0.11	0.05	0.38	0.18	0.06	0.09	0.06	0.06	0.05	0.08	
(HexNAc2Hex3)HexNAc2Fuc2Hex	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
High Mannose Hybrid	73.56	66.45	65.66	72.39	30.55	36.24	80.55	85.89	80.29	82.48	84.26	81.62	50.53	54.13	66.09	67.07	46.53	33.60	67.58	67.41	66.82	50.19	67.68	86.04	
Paucimannose	24.39	29.42	28.94	22.87	47.37	43.26	16.70	12.33	14.99	14.95	13.00	15.90	38.57	35.77	24.49	24.15	33.59	48.23	25.20	26.13	26.87	38.88	28.06	9.86	
Core fucosylation	2.05	4.13	5.40	4.74	22.09	20.50	2.75	1.78	4.71	2.57	2.73	2.48	9.70	8.85	8.66	7.29	15.05	12.79	6.62	5.91	4.14	9.28	3.12	2.97	
Core difucosylation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.08	1.12	0.60	1.37	3.71	4.55	0.49	0.42	2.18	1.54	1.01	0.97	
Core galactosylation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.13	0.17	0.13	1.13	0.84	0.11	0.13	0.42	0.19	0.12	0.16	
SD (High Mannose)	6.23	8.41	6.91	6.04	1.71	1.64	2.70	8.02	11.19	6.95	4.52	4.53	4.13	3.53	6.92	5.23	1.40	0.69	2.65	6.10	7.29	2.80	5.26	3.34	
SD (Paucimannose)	7.07	8.57	6.84	6.17	3.65	1.94	1.24	1.97	0.45	1.57	1.92	1.06	2.05	1.57	0.70	1.01	0.84	2.63	2.96	4.95	1.03	1.80	4.56	0.82	
SD (Paucimannose)	1.00	2.23	2.35	2.29	0.29	0.32	0.80	0.52	0.93	0.93	0.78	0.48	0.22	1.17	0.96	0.91	0.34	0.67	1.03	0.75	0.42	1.10	0.38	0.70	
SD (Core fucosylation)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.43	0.50	0.20	0.51	0.13	0.17	0.12	0.09	0.90	0.75	0.88	0.28	
SD (Core difucosylation)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.04	0.03	0.04	0.01	0.03	0.03	0.07	0.04	0.04	0.06	0.05	
SD (Core galactosylation)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.23	0.10	0.11	0.12	0.17	0.09	0.08	0.12	0.17	0.12	0.28	

H11-5

Glycan structure

	A	B	C	D	E	F	G	H	I	J	K	L	Q	
HexNAc2Hex8	0.18	0.12	0.10	0.14	1.19	2.72	0.22	0.13	0.20	0.33	0.23	0.22	0.14	0.18
HexNAc2Hex8	1.01	0.51	0.68	0.92	7.83	5.86	1.14	0.82	0.70	0.59	0.97	0.64	0.75	0.85
HexNAc2Hex7	1.17	0.62	0.61	0.77	3.99	3.83	1.32	0.91	1.33	1.70	1.29	1.07	1.51	1.53
HexNAc2Hex6	0.96	0.83	0.77	0.70	2.59	3.90	0.74	0.77	0.55	1.62	2.05	1.66	1.50	1.63
HexNAc2Hex5	0.75	0.88	0.81	0.80	2.55	3.13	0.87	0.93	0.96	1.20	1.41	1.30	1.31	1.25
HexNAc2Hex4	0.48	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	2.29	2.02	2.23	1.72	0.51
HexNAc2Hex3	0.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.52	2.19	2.28	1.49	1.02
HexNAc2Hex2	0.09	0.25	0.00	0.00	0.59	1.11	0.00	0.00	0.00	2.38	2.13	2.29	1.27	1.13
HexNAc2Hex1	0.94	1.78	1.16	1.42	1.91	2.53	1.33	1.04	1.85	1.82	2.31	2.00	1.41	4.64
(HexNAc2Hex3)HexNAc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)HexNAc2	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.20	1.01	1.04	0.68	0.27
(HexNAc2Hex3)HexNAc2	0.00	0.74	0.00	0.00	0.00	0.78	0.00	0.00	0.00	15.12	15.78	15.85	12.96	8.87
(HexNAc2Hex2)HexNAc2	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.60	0.46	0.65	0.67	0.34
(HexNAc2Hex2)Fuc	0.00	0.06	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.62	0.55	0.64	0.91	0.38
(HexNAc2Hex2)Fuc	0.00	1.88	2.52	0.63	1.97	2.82	0.00	0.00	0.00	37.83	38.01	38.31	37.16	21.48
(HexNAc2Hex3)Fuc	0.89	1.04	0.69	0.74	1.71	2.63	0.91	1.01	1.31	1.13	1.40	1.26	0.76	2.33
(HexNAc2Hex3)HexNAcFuc	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)HexNAc2Fuc	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.09	0.11	0.06	0.02	0.07
(HexNAc2Hex3)Fuc2	1.14	1.04	1.65	1.17	0.00	1.60	1.40	1.75	1.17	2.67	2.25	2.53	1.61	1.44
(HexNAc2Hex3)HexNAcFuc2	0.00	0.29	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)HexNAc2Fuc2	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)FucHex	3.02	8.54	4.87	3.94	4.44	3.22	5.54	7.55	6.83	6.33	5.80	6.22	6.35	5.61
(HexNAc2Hex3)HexNAcFucHex	14.40	28.35	17.52	19.45	12.37	8.70	20.82	19.24	23.97	1.59	1.43	1.40	1.09	2.63
(HexNAc2Hex3)HexNAc2FucHex	1.64	0.65	1.01	1.24	0.56	0.22	1.72	1.45	1.72	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)FucHex2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.23	0.83	0.94	0.50	0.07
(HexNAc2Hex3)FucHex3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)HexNAcFucHex2	0.00	1.13	1.12	1.13	0.87	0.00	1.32	1.33	1.19	1.58	1.11	1.25	0.72	0.36
(HexNAc2Hex3)HexNAcFucHex3	0.14	0.29	0.00	0.00	0.00	0.00	0.00	0.00	1.09	0.73	0.89	0.79	0.28	0.69
(HexNAc2Hex3)Fuc2Hex	27.92	27.81	29.50	26.10	32.52	33.04	24.40	31.94	25.52	14.55	14.99	14.53	24.44	15.92
(HexNAc2Hex3)Fuc2Hex	32.79	20.40	29.31	30.86	21.94	19.10	29.64	24.55	26.11	0.39	0.44	0.47	0.57	2.42
(HexNAc2Hex3)HexNAcFuc2Hex	12.06	2.38	7.69	10.00	2.97	1.89	8.61	6.49	7.20	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)HexNAc2Fuc2Hex	4.08	3.05	2.97	3.33	18.15	19.44	4.31	3.65	3.12	4.93	6.47	5.12	4.86	30.27
High Mannose Hybrid	1.93	2.03	1.16	1.42	2.50	5.90	1.33	1.04	1.85	9.01	8.66	8.80	5.88	7.30
Paucimannose	0.00	0.87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	16.91	17.26	17.54	14.31	9.48
Core fucosylation	20.08	42.07	27.72	27.13	21.93	17.75	30.31	30.58	35.02	51.41	49.85	50.91	48.28	33.14
Core difucosylation	73.92	51.97	68.14	66.13	57.42	56.02	64.04	64.72	60.01	17.74	17.77	17.63	26.67	19.81
Core galactosylation	91.96	89.56	91.01	92.72	75.66	66.17	92.05	92.55	92.55	26.76	25.32	25.69	34.46	27.29
SD (High Mannose)	0.33	0.28	0.30	0.35	0.44	0.41	0.25	0.40	0.62	0.24	0.17	0.23	0.25	0.28
SD (Hybrid)	0.66	0.29	0.48	0.12	0.12	0.12	0.34	0.25	1.23	0.54	0.57	0.23	0.21	0.10
SD (Paucimannose)	0.00	0.10	0.00	0.00	0.00	0.01	0.00	0.00	0.00	1.21	1.29	1.36	1.18	0.47
SD (Core fucosylation)	1.96	6.53	3.29	2.50	0.35	0.30	1.98	2.09	8.28	3.79	2.85	1.00	2.68	1.10
SD (Core difucosylation)	2.72	4.97	4.63	2.83	0.90	1.63	2.41	1.85	8.09	2.09	1.63	0.44	2.49	1.04
SD (Core galactosylation)	3.34	8.17	5.19	3.74	0.96	1.62	3.08	2.86	11.53	2.44	1.97	0.83	2.70	1.13

H11-5

Glycan structure	11											13																			
	A	B	C	D	E	G	I	J	K	L	Q	A	B	C	D	E	G	I	J	K	L	Q									
HexNAc2Hex9	0.34	0.39	0.29	0.30	0.98	0.14	0.08	0.28	0.35	0.41	0.34	30.91	26.57	18.47	18.52	21.82	25.13	10.84	36.84	19.36	1.86	3.17	2.04	1.66	3.68	2.44	2.23	4.40	2.71	2.61	2.45
HexNAc2Hex8	0.46	0.52	0.38	0.37	1.53	0.25	0.18	0.32	0.58	0.65	0.60	29.85	28.52	32.77	33.06	33.02	26.07	44.50	35.40	32.47	4.23	7.73	5.27	5.16	16.46	6.22	5.83	9.12	7.06	6.86	5.96
HexNAc2Hex7	0.62	0.86	0.91	0.86	1.79	0.46	0.55	0.51	1.07	1.15	1.11	18.56	20.11	19.99	19.42	20.17	20.42	19.27	16.86	16.04	7.00	12.12	7.91	6.86	16.00	8.97	7.36	12.97	9.05	8.93	9.03
HexNAc2Hex6	0.93	1.14	0.99	0.85	1.79	0.93	1.62	0.72	0.83	0.73	0.86	19.13	22.78	24.93	25.94	22.79	26.81	22.78	10.00	29.16	8.67	11.24	9.73	7.84	9.00	8.92	7.20	12.70	9.37	9.11	8.65
HexNAc2Hex5	1.03	1.07	0.98	0.96	1.05	1.24	2.40	0.75	1.59	2.36	1.34	0.66	0.95	2.69	1.74	1.00	0.73	1.89	0.65	1.35	8.84	9.29	11.19	8.20	8.25	11.44	6.02	11.52	8.88	9.95	11.27
HexNAc2Hex4	0.31	0.23	0.35	0.34	0.45	0.54	0.35	0.19	0.38	0.56	0.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.50	3.20	3.40	2.63	3.14	3.07	2.79	3.70	3.77	3.34	
HexNAc2Hex3HexNAc	1.94	1.89	1.47	1.13	1.07	1.94	1.88	1.15	1.72	1.86	1.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.84	2.82	2.79	2.83	1.90	3.44	3.05	2.90	3.41	3.48	3.03	
HexNAc2Hex3HexNAc	0.85	0.81	0.93	0.93	0.78	1.93	1.57	1.23	1.19	1.49	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.84	2.82	2.79	2.83	1.90	3.44	3.05	2.90	3.41	3.48	3.03	
HexNAc2Hex3HexNAc	7.71	8.56	6.39	6.47	3.67	8.62	5.91	6.12	6.85	5.86	8.28	0.72	0.97	0.86	1.03	0.87	0.76	0.51	0.25	1.09	29.48	20.42	22.20	18.16	6.55	20.65	18.15	21.81	18.72	18.74	21.44
(HexNAc2Hex3)HexNAc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.16	0.13	0.29	0.04	0.61	0.25	0.14	0.12	0.17	0.25	
HexNAc2Hex4	0.46	0.35	0.42	0.58	0.30	0.60	0.22	0.22	0.44	0.49	0.64	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HexNAc2Hex3	17.46	17.11	16.84	16.68	6.57	15.25	16.37	8.63	23.59	28.23	22.32	0.19	0.09	0.29	0.28	0.22	0.06	0.11	0.00	0.52	6.57	5.16	6.41	4.95	1.60	6.37	4.42	4.75	3.82	4.05	7.76
HexNAc2Hex2	3.76	3.18	3.55	2.78	0.82	4.21	3.21	1.67	3.46	5.22	3.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.39	0.79	0.45	0.56	3.71	0.45	0.98	0.72	0.35	0.16	0.36	
HexNAc2Hex2Fuc	1.62	0.57	0.82	0.75	0.34	0.75	0.74	0.87	0.76	0.96	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)Fuc	8.94	10.42	11.07	10.87	11.26	10.97	9.64	15.18	9.12	9.04	10.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.09	2.04	3.08	3.55	2.96	2.17	2.43	1.79	2.11	1.91	2.56	
(HexNAc2Hex3)HexNAcFuc	1.06	1.36	1.40	1.21	3.21	1.16	2.00	2.33	0.60	0.51	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.14	1.87	2.02	2.01	2.65	1.74	2.32	1.57	2.08	1.99	1.84	
(HexNAc2Hex3)HexNAc2Fuc	0.06	0.08	0.07	0.27	0.10	0.07	0.84	0.00	0.08	0.08	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HexNAc2Hex2Fuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)Fuc2	2.96	1.88	1.92	1.84	2.12	1.53	1.05	2.51	1.10	0.86	1.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.13	0.18	0.18	0.34	0.12	0.11	0.08	0.13	0.14	0.14	
(HexNAc2Hex3)HexNAcFuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)HexNAc2Fuc2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)FucHex	21.27	20.56	19.53	20.42	8.85	18.17	17.82	15.09	23.13	22.16	17.65	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)HexNAcFucHex	5.24	6.20	4.49	4.82	4.10	7.11	5.68	6.63	4.34	3.16	4.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	11.63	9.93	11.35	15.75	5.55	10.49	17.41	6.63	13.47	14.24	9.59	9.59
(HexNAc2Hex3)HexNAc2FucHex	17.59	17.20	23.06	23.81	42.27	19.89	17.74	30.16	15.06	11.40	17.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)FucHex2	0.44	0.34	0.42	0.45	0.73	0.45	4.40	1.10	0.46	0.35	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)FucHex3	0.57	0.51	0.36	0.21	0.36	0.26	1.30	0.58	0.50	0.41	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)HexNAcFucHex2	0.43	0.63	0.36	0.37	0.59	0.53	0.57	0.38	0.33	0.21	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)HexNAcFucHex3	2.86	2.72	1.66	1.22	0.69	1.95	3.10	1.70	1.94	1.41	1.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)Fuc2Hex	17.59	17.20	23.06	23.81	42.27	19.89	17.74	30.16	15.06	11.40	17.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(HexNAc2Hex3)HexNAcFuc2Hex	1.08	1.45	1.33	1.48	4.54	1.08	0.79	1.68	0.54	0.45	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.04	2.05	2.45	4.45	5.82	2.11	4.66	1.37	4.18	3.95	1.94	1.94
(HexNAc2Hex3)HexNAc2Fuc2Hex	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
High Mannose Hybrid	3.38	3.97	3.56	3.34	7.15	3.01	4.83	2.59	4.42	5.30	4.25	99.09	98.94	98.84	98.81	99.17	99.38	99.75	98.39	30.60	43.54	26.14	29.73	54.30	37.99	28.64	50.71	37.07	37.46	37.36	37.36
Paucimannose	10.81	11.48	9.14	8.88	5.98	13.02	9.72	8.69	10.13	9.75	11.75	0.72	0.97	0.86	1.03	0.87	0.76	0.51	0.25	1.09	30.00	26.62	36.13	24.68	11.12	27.84	24.52	27.64	25.95	26.16	28.06
Core fucosylation	42.50	43.39	40.18	40.60	30.24	41.40	46.08	33.94	33.94	26.85	0.19	0.09	0.29	0.28	0.22	0.06	0.11	0.00	0.52	6.96	5.96	6.86	5.51	5.31	6.82	5.39	5.47	4.17	4.21	8.12	8.12
Core difucosylation	21.63	20.52	26.31	27.13	48.94	22.51	19.58	34.34	16.71	12.71	20.51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	20.93	18.71	22.27	28.80	13.52	21.65	30.74	13.17	23.42	23.55	20.79	20.79
Core galactosylation	49.48	49.60	51.22	52.79	62.14	49.44	51.39	57.31	46.29	39.55	43.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.52	5.17	6.60	11.28	15.76	5.69	10.71	3.01	9.39	8.62	5.68	5.68
SD (High Mannose)	0.39	0.46	0.30	0.26	0.25	1.27	2.36	0.31	0.86	2.10	1.37	4.92	3.03	5.06	4.53	1.46	6.35	4.06	2.91	7.92	1.20	3.34	2.19	1.64	0.89	1.73	1.37	4.21	3.34	3.69	2.89
SD (Hybrid)	1.56	0.88	0.92	0.70	0.14	3.22	2.63	1.64	0.58	0.46	3.18	0.26	0.17	0.34	0.34	0.18	0.66	0.37	0.23	0.04	3.49	3.37	2.13	0.51	0.28	2.15	0.47	1.13	1.79	3.12	3.73
SD (Paucimannose)	2.15	2.64	1.77	1.40	0.36	1.03	3.64	1.86	1.12	2.33	2.80	0.06	0.08	0.39	0.08	0.13	0.09	0.10	0.00	0.65	0.38	0.38	0.41	1.06	0.79	0.29	0.41	0.37	0.70	1.22	1.22
SD (Core fucosylation)	2.62	1.93	1.49	2.07	0.83	4.79	6.93	1.67	1.70	1.68	3.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.34	2.03	3.97	4.61	0.41	1.74	1.52	1.81		

Table S5: Quantification of glycosite-specific N-glycosylation profile of native H11 isoforms

Native H11 was purified twice from *H. confortus* adult worms by solubilization of a PBS homogenate pellet in 2 % Triton X-100 (Sigma-Aldrich #T9284). Once this extract of native H11 was further enriched by a Concanavalin A Sepharose (Sigma-Aldrich #C9017) affinity chromatography. Site-specific glycan profiles of trypsin digested glycopeptides were analyzed by nanoHPLC-HCD MS/MS mass spectrometry. Glycan structures found on each glycosylation site are listed with their respective percentage, or if quantification was not possible, detection is indicated by detected or nd if not detected. Data represent the mean values of glycan ratios of two different experiments and revealed similar results. The glycan structures were summarized into the following groups: High Mannose (sum of 1-5), Hybrid (sum of 4-6), Paucimannose (sum of 7-9), Mono-fucosylation (sum of 10-19), Di-fucosylation (sum of 20-29) and Tri-fucosylation (sum of 30-33).

Glycan structure	Native H11			Native H11-2			Native H11-5		
	Glycosite 1	Glycosite 3	Glycosite 13	Glycosite 2	Glycosite 7	Glycosite 2	Glycosite 7	Glycosite 2	Glycosite 7
1 HexNAc2Hex7	0.38	nd	nd	nd	nd	detected	detected	nd	nd
2 HexNAc2Hex6	nd	nd	nd	nd	nd	detected	detected	nd	nd
3 HexNAc2Hex5	37.41	12.45	3.52	detected	detected	detected	detected	nd	nd
4 HexNAc2Hex4HexNac	1.12	0.22	nd	detected	detected	nd	nd	nd	nd
5 (HexNAc2Hex3)HexNac	23.94	1.05	7.51	detected	detected	nd	nd	nd	nd
6 (HexNAc2Hex3)HexNAc2	nd	nd	1.39	nd	nd	nd	nd	nd	nd
7 HexNAc2Hex4	9.47	28.37	10.95	detected	detected	detected	detected	detected	detected
8 HexNAc2Hex3	5.76	51.68	3.66	nd	nd	nd	nd	nd	nd
9 HexNAc2Hex2	nd	nd	nd	nd	nd	nd	nd	nd	nd
10 HexNAc2FucHex2	0.86	nd	nd	nd	detected	detected	detected	nd	nd
11 HexNAc2FucHex3	2.78	nd	15.39	nd	detected	detected	detected	nd	nd
12 (HexNAc2Hex2)HexNacFuc	nd	nd	nd	nd	detected	detected	detected	nd	nd
13 (HexNAc2Hex2)HexNacFucHex	0.40	nd	8.60	nd	detected	detected	detected	nd	nd
14 (HexNAc2Hex2)HexNac2FucHex	2.94	3.16	3.99	nd	nd	nd	nd	nd	nd
15 (HexNAc2Hex3)HexNacFucHex	nd	0.31	nd	nd	nd	nd	nd	nd	nd
16 (HexNAc2Hex3)HexNac2FucHex	nd	2.38	1.85	nd	nd	nd	nd	nd	nd
17 (HexNAc2Hex3)HexNac2FucHex2	nd	0.37	nd	nd	nd	nd	nd	nd	nd
18 (HexNAc2Hex2)HexNac2FucHex2	2.44	nd	nd	nd	nd	nd	nd	nd	nd
19 (HexNAc2Hex3)HexNac3Fuc	nd	nd	0.37	nd	nd	nd	nd	nd	nd
20 HexNAc2Fuc2Hex2	nd	nd	7.98	nd	detected	detected	detected	detected	detected
21 HexNAc2Fuc2Hex3	3.91	nd	8.60	nd	detected	detected	detected	detected	detected
22 (HexNAc2Hex2)HexNacFuc2	nd	nd	4.48	nd	detected	detected	detected	nd	nd
23 (HexNAc2Hex2)HexNacFuc2Hex	nd	nd	nd	nd	detected	detected	detected	nd	nd
24 (HexNAc2Hex2)HexNac2Fuc2	nd	nd	2.42	nd	detected	detected	detected	nd	nd
25 (HexNAc2Hex2)HexNac2Fuc2Hex	nd	nd	3.00	nd	detected	detected	detected	nd	nd
26 (HexNAc2Hex2)HexNac2Fuc2Hex2	nd	nd	2.22	nd	nd	nd	nd	nd	nd
27 (HexNAc2Hex3)HexNac2Fuc2Hex2	nd	nd	0.88	nd	nd	nd	nd	nd	nd
28 (HexNAc2Hex3)HexNac3Fuc2	nd	nd	0.42	nd	nd	nd	nd	nd	nd
29 (HexNAc2Hex3)HexNac3Fuc2Hex	nd	nd	0.37	nd	nd	nd	nd	nd	nd
30 HexNAc2Fuc3Hex2	nd	nd	3.03	nd	detected	detected	detected	detected	detected
31 HexNAc2Fuc3Hex3	8.59	nd	6.16	nd	detected	detected	detected	detected	detected
32 (HexNAc2Hex2)HexNac2Fuc3Hex	nd	nd	0.15	nd	nd	nd	nd	nd	nd
33 (HexNAc2Hex2)HexNac2Fuc3Hex2	nd	nd	0.80	nd	nd	nd	nd	nd	nd
High Mannose	37.79	12.45	3.52						
Hybrid	25.06	1.28	8.90						
Paucimannose	15.22	80.05	14.61						
Mono-fucosylation	9.43	6.22	32.45						
Di-fucosylation	3.91	0.00	30.38						
Tri-fucosylation	8.59	0.00	10.14						

Table S6: Plasmids of this study.

SP: gp67 secretion signal peptide, PSP: PreScission protease cleavage site, HIS: His₁₀-tag.

Name	Insert/Gene	Primers used	Source
pFastBacDual	-	-	Thermo Fisher Scientific
pGEM-T-Easy	-	-	Promega
pRG101	SP, HIS, PSP, cloning site	-	(Hang et al., 2015)
pMA148	FLAG GALT-1	-	(Titz et al., 2009)
pSF23	FLAG GALT-1	-	This study
pSF25	FLAG GALT-1; SP, HIS, PSP, H11	GALT Fw Xmal and GALT Rv Nsil	This study
pSF34	FLAG GALT-1; SP, HIS, PSP, CH-H11-1D	H11-1 fw Apal and H11-1 rv Stul	This study
pSF35	FLAG GALT-1; SP, HIS, PSP, CH-H11-2A	H11-2 fw Apal and H11-2 rv Stul	This study
pSF36	FLAG GALT-1; SP, HIS, PSP, CH-H11-4C	H11-4 fw Apal and H11-4 rv HindIII	This study
pSF37	FLAG GALT-1; SP, HIS, PSP, CH-H11-5B	H11-5 fw Apal and H11-5 rv Stul	This study
pSF38	FLAG GALT-1; SP, HIS, PSP, CH-H11-1C	H11-1 fw Apal and H11-1 rv Stul	This study
pSF39	FLAG GALT-1; SP, HIS, PSP, CH-H11-1J	H11-1 fw Apal and H11-1 rv Stul	This study
pSF40	FLAG GALT-1; SP, HIS, PSP, CH-H11-1N	H11-1 fw Apal and H11-1 rv Stul	This study
pSF41	FLAG GALT-1; SP, HIS, PSP, CH-H11-2D	H11-2 fw Apal and H11-2 rv Stul	This study
pSF42	FLAG GALT-1; SP, HIS, PSP, CH-H11-2F	H11-2 fw Apal and H11-2 rv Stul	This study
pSF43	FLAG GALT-1; SP, HIS, PSP, CH-H11-2H	H11-2 fw Apal and H11-2 rv Stul	This study
pSF44	FLAG GALT-1; SP, HIS, PSP, CH-H11-5A	H11-5 fw Apal and H11-5 rv Stul	This study
pSF45	FLAG GALT-1; SP, HIS, PSP, CH-H11-5C	H11-5 fw Apal and H11-5 rv Stul	This study
pSF46	FLAG GALT-1; SP, HIS, PSP, CH-H11-5D	H11-5 fw Apal and H11-5 rv Stul	This study
pSF47	FLAG GALT-1; SP, HIS, PSP, CH-H11-5E	H11-5 fw Apal and H11-5 rv Stul	This study
pSF48	FLAG GALT-1; SP, HIS, PSP, CH-H11-5F	H11-5 fw Apal and H11-5 rv Stul	This study
pSF49	FLAG GALT-1; SP, HIS, PSP, CH-H11-5G	H11-5 fw Apal and H11-5 rv Stul	This study
pSF51	FLAG GALT-1; SP, HIS, PSP, CH-H11-5I	H11-5 fw Apal and H11-5 rv Stul	This study
pSF52	FLAG GALT-1; SP, HIS, PSP, CH-H11-5J	H11-5 fw Apal and H11-5 rv Stul	This study
pSF53	FLAG GALT-1; SP, HIS, PSP, CH-H11-5K	H11-5 fw Apal and H11-5 rv Stul	This study
pSF54	FLAG GALT-1; SP, HIS, PSP, CH-H11-5L	H11-5 fw Apal and H11-5 rv Stul	This study
pSF55	FLAG GALT-1; SP, HIS, PSP, CH-H11-5Q	H11-5 fw Apal and H11-5 rv Stul	This study
pSF56	FLAG GALT-1; SP, HIS, PSP, CH-H11-5R	H11-5 fw Apal and H11-5 rv Stul	This study

Table S7: Primers used in this study.

Name	Sequence
GALT Fw Xmal	5'-AAAAAACCCGGGATGGACTACAAGGACGACGACG-3'
GALT Rv Nsil	5'-AAAAAATGCATCTACAAGTCTAAAAGACCAACAACACTGAATC-3'
AC-1 fw	5'-AAAAATGAAATACTTGGTGCTTGAC-3'
AC-1 rv	5'-AAAAGTAATCCAGCTTTTATCGAATACATTAC-3'
H11-fw new	5'-AAAAATGACGTGCGCAGGGGAG-3'
H11-rv new	5'-AAAATTACAAGGTGGCTTTCTTGAAGAAAG-3'
H11-1 fw new	5'-AAAAATGACAGCAGAGGAGAGTCAGG-3'
H11-1 fw Apal	5'-AAAAAAGGGCCCCTGTATTACTTTACAAGGAAAGCCTTTGATAC-3'
H11-1 rv Stul	5'-AAAAAAAGGCCTTTATGAATTAGATTTTTTGAAGAAAGCTGC-3'
H11-2 fw new	5'-AAAAATGACGGCGGAGTGCC-3'
H11-2 fw Apal	5'-AAAAAAGGGCCCCTGTATTACTTCACTCGTAAAGCATTTCGATAC-3'
H11-2 rv Stul	5'-AAAAAAAGGCCTCTATGATCTTGCTCTCTTGAAGAATTC-3'
H11-4 fw new	5'-AAAAAATGACGGCAGAGAGGAAAC-3'
H11-4 fw Apal	5'-AAAAAAGGGCCCCTGTATTTCTTTACTCGGAAAGCATTGATC-3'
H11-4 rv HindIII	5'-AAAAAAAAGCTTTTACCAGGTAGCGTTCTTAAAGAATG-3'
H11-5 fw new	5'-AAAAAATGACGGTACAGTGGACTAAACG-3'
H11-5 fw Apal	5'-AAAAAAGGGCCCCTGTACTACTTTACCAGGAAAGCTTATGATACTACT-3'
H11-5 rv Stul	5'-AAAAAAAGGCCTTCATCGAGTAGATTTTTTCGAAGAAATC-3'
Hc-sod-1 fw	5'-AAAAATGAGTAACCGTGCTGTTGCTG-3'
Hc-sod-1 rv	5'-AAAATCACTGGGGAGCAGCG-3'
Pep1 fw	5'-AAAAATGCTATATTTATTGCTCTTGGTGAGC-3'
Pep1 rv	5'-AAAATTATTTTCGGCTCCGCAAAG-3'

Chapter 4

Analysis of site-specific N-glycan remodelling in the ER and the Golgi

Ivan Hang^{1*}, Chia-wei Lin^{1*}, Oliver C. Grant², Susanna M. Fleurkens¹,
Thomas K. Villiger³, Miroslav Soos³, Massimo Morbidelli³, Robert J. Woods²,
Robert Gauss¹, Markus Aebi¹

¹ Institute of Microbiology, Department of Biology, ETH Zurich,
CH-8093 Zurich, Switzerland

² Complex Carbohydrate Research Center, University of Georgia, Athens,
Georgia 30602, USA

³ Institute for Chemical and Bioengineering, Department of Chemistry and
Applied Biosciences, ETH Zurich, CH-8093 Zurich, Switzerland

* These authors contributed equally to this work

Glycobiology, 2015. 25(12): p. 1335-49.

Contributions:

Protein expression and purification of secreted Pdi1p with and without GALT-1

Abstract

The hallmark of N-linked protein glycosylation is the generation of diverse glycan structures in the secretory pathway. Dynamic, non-template driven processes of N-glycan remodelling in the ER and the Golgi provide the cellular setting for structural diversity. We applied newly developed mass spectrometry-based analytics to quantify site-specific N-glycan remodelling of the model protein Pdi1p expressed in insect cells. Molecular dynamics simulation, mutational analysis, kinetic studies of *in vitro* processing events and glycan flux analysis supported the defining role of the protein in N-glycan processing.

Introduction

N-linked glycosylation is an abundant posttranslational modification present on secretory proteins in all domains of life (Dell et al. 2010). In eukaryotes, attachment and subsequent modification of N-glycans affect the folding of glycoproteins and regulate their secretion, e.g. by providing signals to the endoplasmic reticulum (ER) quality control machinery (Aebi et al. 2010; Shental-Bechor and Levy 2009). On the final glycoprotein product, N-glycans act as ligands that mediate cell-cell interactions of cell surface glycoproteins and fine-tune the function of mature glycoproteins (van Kooyk and Rabinovich 2008).

Biosynthesis of N-glycoproteins starts with the stepwise assembly of a lipid-linked oligosaccharide precursor at the ER membrane. Oligosaccharyl transferase then transfers the G3M9Gn2 carbohydrate (G: glucose, M: mannose, Gn: N-acetylglucosamine) from the isoprenoid lipid carrier onto the asparagine side-chain of the N-X-S/T consensus sequence of polypeptides in the ER lumen (Kelleher and Gilmore 2006). Initial modifications of the N-glycan in the ER involve successive removal of glucose and α -1,2-mannosyl residues to generate oligomannose type glycans (Hebert et al. 2005; Mast and Moremen 2006). It is in the Golgi where glycosyl hydrolases (GHs) and transferases (GTs) generate the heterogeneous population of carbohydrate structures presented on mature glycoproteins (Hua et al. 2012; Moremen et al. 2012).

GHs, GTs, and nucleotide sugar transporters represent the 'hardware' of the Golgi remodelling pathway that converts oligomannose glycans into hybrid- and complex-type oligosaccharides (Varki 1998). This network of glycan modifying enzymes can vary between different species, tissues and cell types and generates variable N-glycan structures. Glycoproteins expressed in mammalian cells generally carry an array of complex-type oligosaccharides decorated with terminal galactose and sialic acid residues (North et al. 2010). In contrast, the glycosylation machinery of insect cells produces a variety of paucimannose-type products with α 1,6- or α 1,3-linked fucose residues attached to the core Gn (Kato et al. 2010; Shi and Jarvis 2007).

N-glycan diversity is the result of the dynamic and competitive nature of N-glycan synthesis in the Golgi (Stanley 2011). Golgi GTs and GHs usually have different substrate specificities and act sequentially, but they may also compete for substrates (Moremen et al. 2012). Analysis of site-specific N-glycosylation of Sindbis virus

glycoproteins, carboxypeptidase Y, and invertase provided the first evidence that processing of N-glycans is determined by the physical accessibility of the carbohydrates (Hsieh et al. 1983a; Hsieh et al. 1983b; Hubbard 1988; Trimble et al. 1983). Subsequently, site-specific N-glycosylation has been found on different proteins Thy1, lysosomal α -mannosidase, lactoferrin, and HIV envelope glycoproteins (Faid et al. 2006; Go et al. 2008; Heikinheimo et al. 2003; Hua et al. 2012; Nagae and Yamaguchi 2012; Parekh et al. 1987) and alterations in the polypeptide sequence can yield different N-glycan structures (Yu et al. 2013).

To study in detail the factors that affect site-specific N-glycosylation, we expressed a model protein, yeast Pdi1p, in insect cells and employed newly developed mass spectrometry-based analytical techniques to quantify site-specific N-glycan structures produced in the ER and the Golgi. Combining molecular dynamics simulations, mutational analyses, and *in vitro* assays we define the intramolecular interactions of glycans and the protein surface that are major determinants for carbohydrate modification. Implementing glycan flux analysis we identified the rate-limiting processing steps for each glycosite. Our studies confirm the importance of the protein structure in the pathway of N-glycan processing.

Results

Establishing an MS-based workflow to analyze site-specific glycosylation

ER-resident protein disulfide isomerase 1 (Pdi1p) from *Saccharomyces cerevisiae* is composed of four thioredoxin-like domains (a, b, b' and a') (Tian et al. 2006; Wilkinson and Gilbert 2004). Unlike mammalian PDIs, yeast Pdi1p has five N-glycosylation sites (S1-S5) that are localized in three of the four domains (Figure 1A). Tryptic peptides that containing these glycosite were used in our study (Table 1). A C-terminal HDEL sequence retained the protein in the ER and early Golgi, while disruption of this sequence resulted in the secretion of the protein into the medium (sPdi1p), allowing us to analyse ER- and Golgi-located processing separately (Figure 1B). Pdi1p proteins were expressed in *Trichoplusia ni* insect cells using the baculovirus expression system and purified by Ni-NTA affinity chromatography via a N-terminal His₁₀-tag. SDS-PAGE and immunoblot analysis verified that the purity was suitable for the following MS measurements (Figure 1C).

Table 1: Glycosylation site of yeast Pdi1p.

A list of the five tryptic peptides containing a glycosylation sequon monitored in this study. C* represents carbamidomethylated cysteine.

Site	sequence	[M+H] ⁺
1	N ITLAQIDCTENQDLC*MEHNIPGFPSLK	3258.52
2	NSDV N NSIDYEGPR	1579.70
3	QSQPAVAVVADLPAYLA N ETFVTPVIVQSGK	3212.71
4	IDADF N ATFYSMANK	1707.78
5	LAPTYQELADTYA N ATSDVLIAK	2468.25

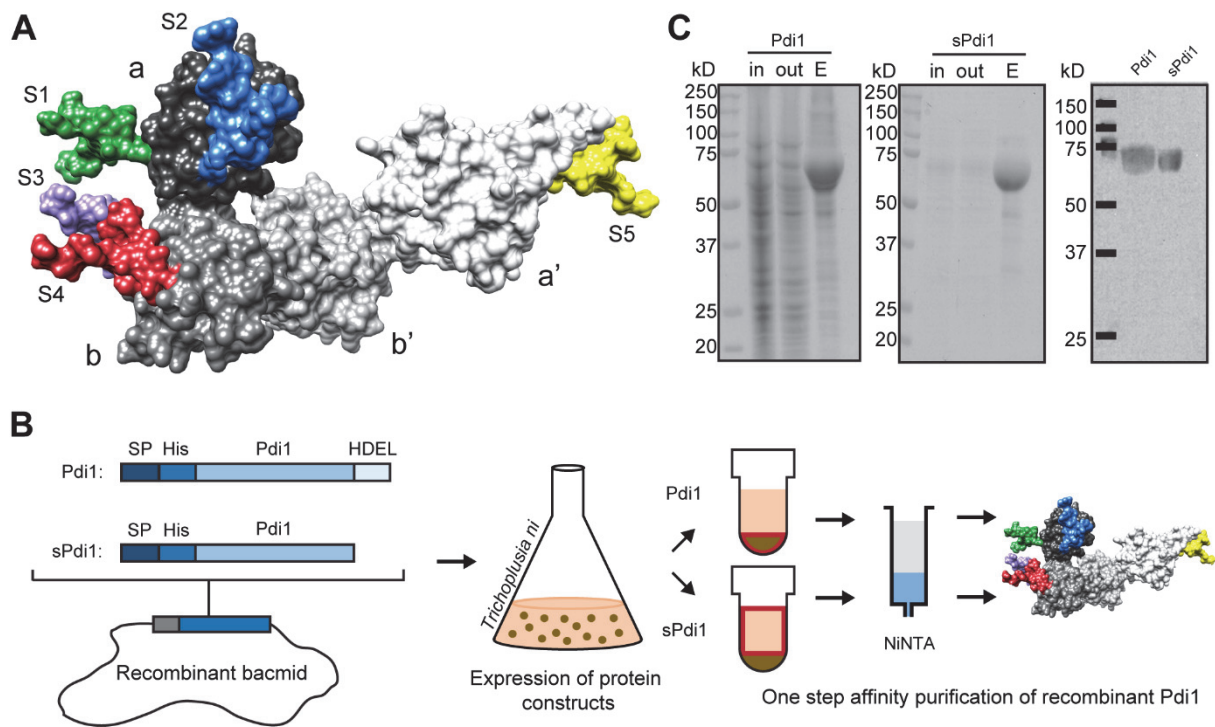


Figure 1. Purification of recombinant Pdi1 and sPdi1p from insect cells.

(A) Model of glycosylated Pdi1p. M9Gn2 N-glycans were modelled onto the five glycosites of the crystal structure of Pdi1p (PDB: 2B5E; www.glycam.org). Glycans are depicted in colors: S1 green, S2 blue, S3 purple, S4 red, S5 yellow. Different domains of Pdi1p are in different shades of gray. **(B)** Schematic representation of Pdi1p and sPdi1p purification. *T. ni* cells were infected with recombinant viruses carrying expression copies of ER retained Pdi1p or secreted Pdi1p (sPdi1p). Pdi1p was purified via NiNTA chromatography from cell lysates; sPdi1p was isolated from culture supernatants. ER retention of full-length sPdi1p was disrupted by the presence of two additional amino acids at the protein's C-terminus (HDELLE) (Raykhel et al. 2007). SP: signal peptide; His: His₁₀ tag, HDEL: ER retrieval signal. **(C)** Purification of Pdi1p and sPdi1p. Samples of input (in), flow-through (out) and eluted (E) fractions were analyzed by SDS-PAGE and stained with Coomassie (left/middle), or analyzed by immunoblot using anti-His5 antibodies (right).

To assess site-specific glycan heterogeneity, we employed the filter-assisted sample preparation (FASP) method (Wisniewski et al. 2009). Peptides were analysed on a nanoLC-LTQ Velos Orbitrap operated in scheduled data-dependent acquisition (DDA) mode, one MS scan followed by ten HCD MS/MS scans. Based on the unique fragmentation of glycopeptides by the HCD method, we developed the ExtractMgf algorithm to speed up manual analyses of MS/MS spectra, as exemplified for the S1 glycopeptide (Segu and Mechref 2010) (Figure 2A). From 3648 measured spectra, the glycan oxonium ions $[\text{HexNAc}]^+$ 204.09 and $[\text{HexNAc+Hex}]^+$ 366.14 were used to identify the MS/MS scans of all glycopeptides from the peak list. Next, the Y1 ion corresponding to m/z of the peptide plus one HexNAc was used to identify the peptide backbone. In the case of S1, 1731.3034 corresponding to $[\text{S1+HexNAc+2H}]^{2+}$ was used to sort out all MS/MS spectra from that glycopeptide. After two sorting runs, the remaining twenty spectra were verified manually. Since $^{0,2}\text{X}$ ring cleavage on the single remaining HexNAc was a frequent event and resulted in neutral loss of 120 Da from the Y1 ion (Figure 2B, Figure S1), we used the triple peaks, Y1, $^{0,2}\text{X}$ and $[\text{peptide+H}]^+$, to manually confirm the identity of the Y1 ion in HCD spectra. One example of a S1 MS/MS spectrum of the M7Gn2 is shown in Figure 2B.

Since glycopeptides containing the same peptide backbone co-elute when applied to reverse phase liquid chromatography, we identified the different glycan structures of one site by grouping MS spectra based on the presence of Y1 ions observed in the original MS/MS spectra. In the example given, the overall glycosylation profile of S1 ranged from M5Gn2 to M9Gn2 (Figure 2C). Finally, we quantified the amount of each glycan structure by its extracted ion chromatogram (Figure 2D). We verified the accuracy of the method by analysing bovine RNase B that carries a known mixture of oligomannose glycans on a single glycosite (Figure S2).

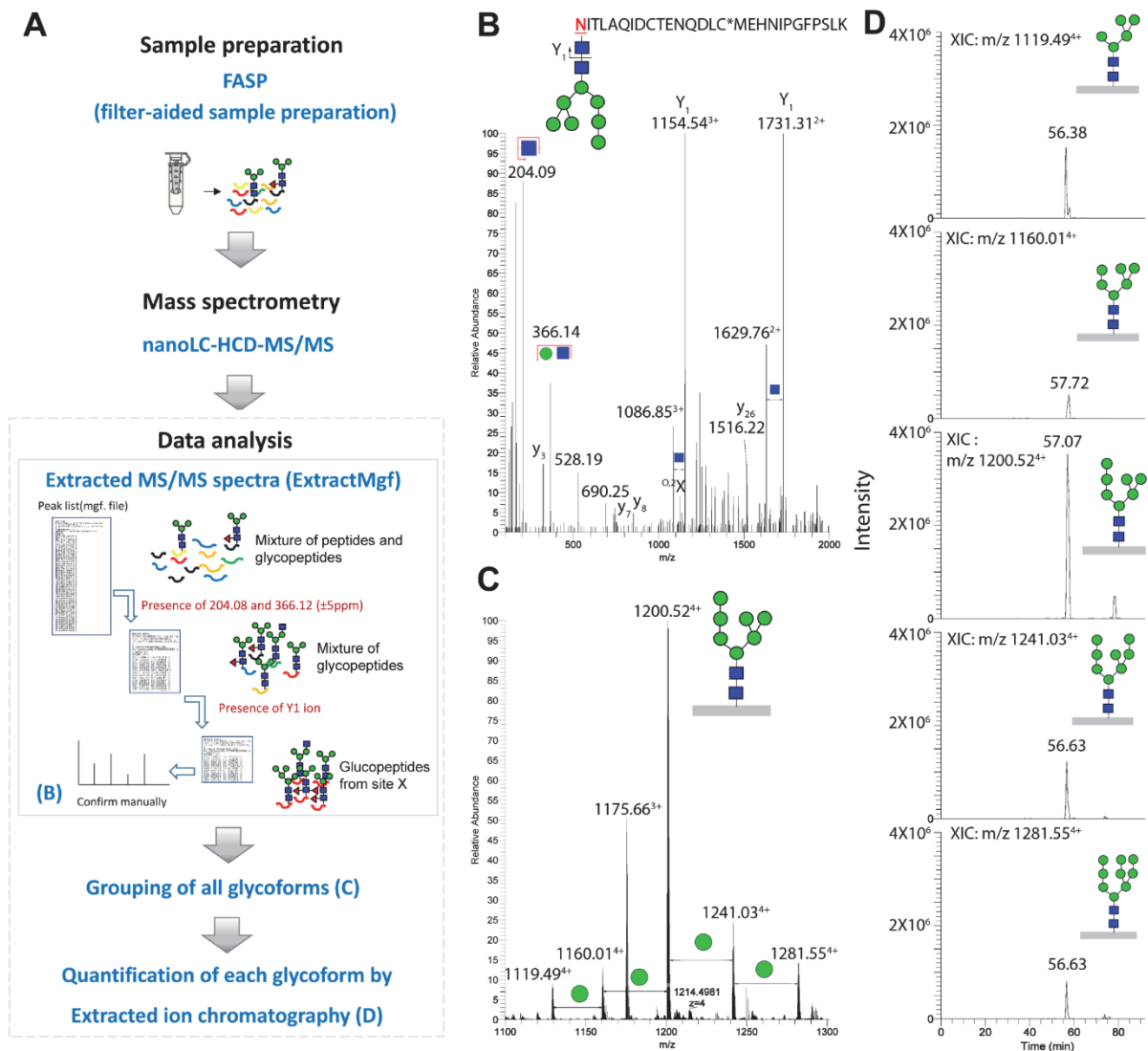


Figure 2. Overall workflow for glycopeptide analysis by mass spectrometry.

(A) Purified proteins were processed by the filter-assisted sample preparation (FASP). The mixture of peptides and glycopeptides was analyzed by LC-HCD mass spectrometry. Raw data was transformed into a peak list and processed by ExtractMgf. For relative quantification, extract ion chromatography (XIC) of each glycoform was plotted and its corresponding peak area was integrated. The relative abundance of each form was calculated. Symbols represent monosaccharides: mannose (green circle), N-acetyl-glucosamine (blue square), fucose (red triangle) (B) One MS/MS spectrum at m/z 1200.54 ($z=4$) was assigned to the first glycosylation site of Pdi1p containing M7Gn2 glycoform. Glycan oxonium ions and doubly/triply charged Y1 at m/z 1154.54 ($z=3$)/1731.31 ($z=2$) are indicated. The nomenclature of peptide fragment ions and glycan fragmentation ions was described previously (Dell et al. 1994; Roepstorff and Fohlman 1984). (C) The overall glycosylation profile of S1 obtained by grouping of MS spectra. Corresponding ions are indicated. (D) XIC of each glycoform sharing the same peptide backbone. See Figure S2 and Table S2 for quantification of RNaseB glycans.

Pdi1p glycans are differentially processed

In an initial experiment, we analyzed the N-linked glycans of Pdi1p retained in the ER and early Golgi. As expected, we found oligomannose structures; yet, we did not observe a defined structure per site, but rather a mixture of glycans processed to varying degrees (Figure 2D). The processing of N-glycans was significantly different for the five sites analyzed: while S2, S3 and S5 mainly carried M7Gn2 to M5Gn2 glycans, S4 contained a higher fraction of M9Gn2 and M8Gn2 glycans (Figure 3A and 3B). S1 showed an intermediate processing pattern with M7Gn2 as the most prevalent structure. By comparing the glycan profiles for each site using Euclidean distance and hierarchical clustering analysis, we confirmed that S4 showed the most distinct glycan pattern (Figure 3C). These results indicated that the location of an N-glycan on Pdi1p was an essential parameter influencing the processing of the glycan by mannosidases of the ER and early Golgi.

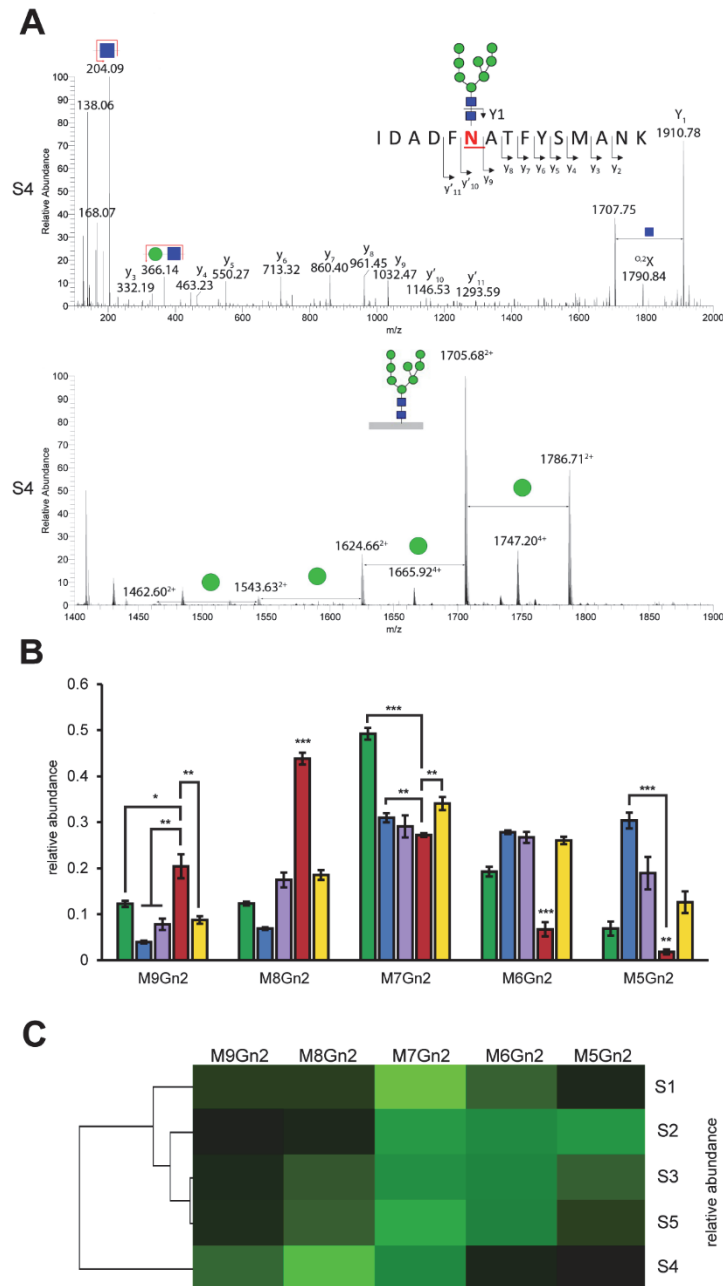


Figure 3. Pdi1p retained in the ER and early Golgi displays site-specific glycan profiles.

(A) Analysis of M8Gn2 on S4. The MS/MS spectrum of m/z 1705.68 ($z=2$) was obtained as described in Figure 2. The peptide was identified by the Y1 ion (S4 peptide plus HexNAc at m/z 1910.78). The peptide sequence is displayed together with the attached M8Gn2 glycan. The underlined N represents the N-glycosite and y' represents y ion without a glycan. **(B)** Site-specific N-glycan profile of Pdi1p. Pdi1p was expressed and purified and site-specific glycan profiles were analyzed. The relative abundance of the different glycoforms at a given site (S1 to S5) is shown. Data represent the mean values of glycan ratios of four independent experiments with error bars indicating the standard deviation. P values were calculated by paired Student's t test: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. **(C)** Similarity representation of five Pdi1p N-glycosylation sites based on the relative glycoform distribution. Similarity between site-specific profiles was calculated using the Euclidean distance and the dendrogram was obtained by Centroid Linkage Clustering. The colour scale represents the relative glycoform abundance from Figure 3B. See Figure S3 and Table S3 for spectra and raw data.

Interactions with the protein surface reduce accessibility of the S4 glycan

In order to test whether glycan-protein interactions influence oligosaccharide processing, we generated a three dimensional model of a truncated version of Pdi1p composed of the a and b domain and with a M9Gn2 glycan attached to S4. To remove any bias from the initial glycan orientation an explicitly-solvated molecular dynamics simulation was performed for 0.5 μ s. Throughout the simulation the B-branch of the S4 glycan remained in close contact with the surface of the b-domain, while the A- and C-branches formed interactions with the a-domain (Figure 4A and Supplemental Movie). Once formed, the interactions between the S4 glycan (located on the b-domain) and the a-domain remained stable over the course of the simulation timescale. The contacts formed with the a-domain reduced the accessibility of the S4 glycan, which correlated with the attenuated processing of the S4 glycan by mannosidases (Figure 4B). Notably, the interactions formed by the S4-glycan with the surface of the a-domain altered the relative orientation and dynamics of the a- and b-domains while the unglycosylated protein fluctuated about the crystal structure conformation (Figure 4C). To confirm these *in silico* predictions we designed truncated variants of Pdi1p, each containing the glycosylated b-domain with an N-terminal His₁₀-tag and a C-terminal HDEL sequence (Figure 5A). The different proteins were expressed, purified, and the peptides were analyzed by the ExtractMgf workflow (Figure 5B & 5C). The glycan profiles on S4 of full-length Pdi1p, the abb'- and the ab-domain glycoprotein were almost identical. However, when we removed the a-domain, the distribution of glycoforms changed and shifted to smaller, more processed glycans (Figure 5C and S3), resulting in a different glycosylation pattern (Figure 5D). These findings supported the hypothesis that the interaction of a glycan with the surface of the covalently linked protein altered the accessibility of the glycan to the processing machinery.

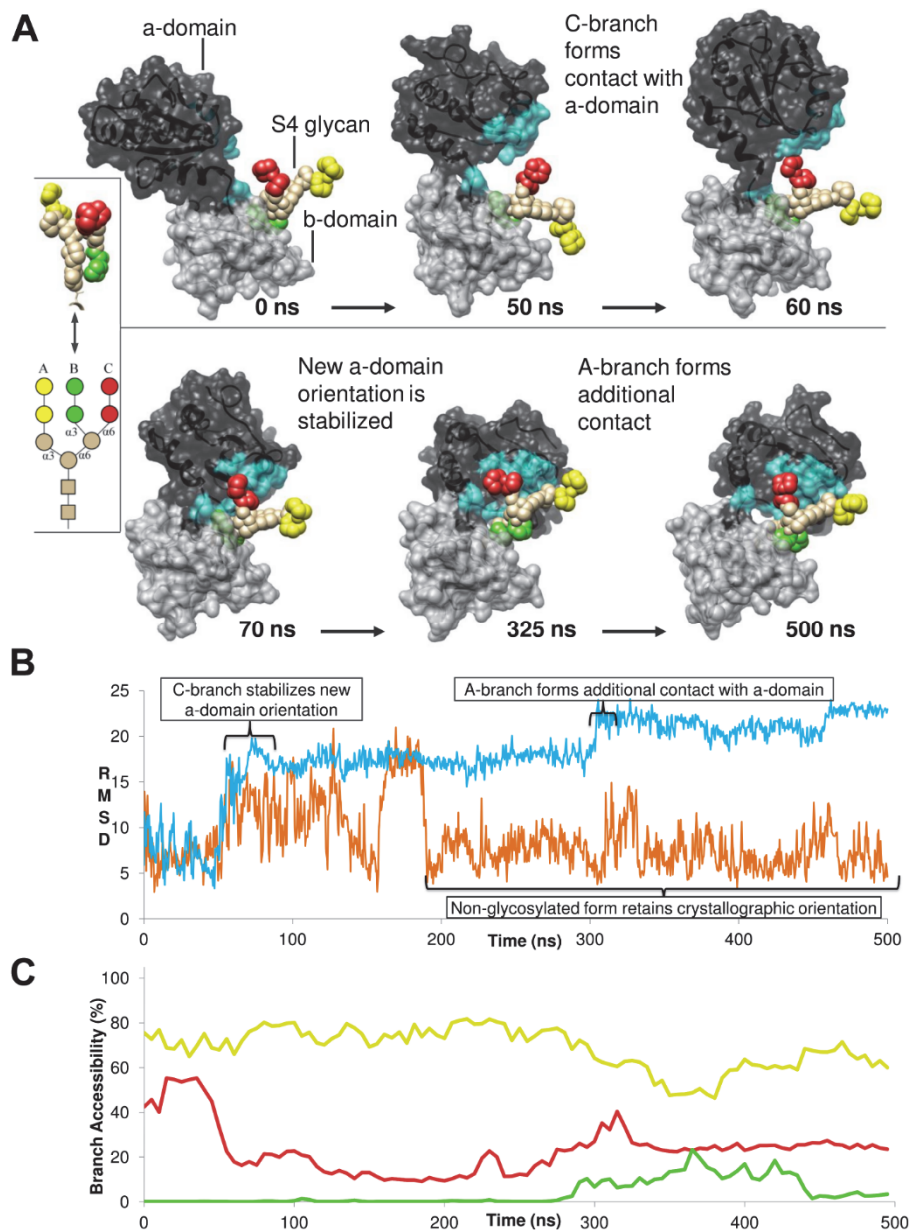


Figure 4. Molecular dynamics simulation of glycosylated Pdi1p-ab.

(A) Six snapshots from the molecular dynamics simulation of the S4 glycosylated ab-domain variant of Pdi1p, showing the contacts formed between the S4 glycan and the a-domain (blue surface). The b-domains of each snapshot are aligned to the crystal structure coordinates. The glycan is shown as van-der-Waals spheres, with the A-branch in yellow, the B-branch in green and the C-branch in red. **(B)** A plot of the RMSD of the a-domain obtained over the course of the simulation relative to the crystal structure orientation (blue line: RMSD of the a-domain with S4 glycosylated, orange line: RMSD of a-domain without S4 glycosylation). **(C)** The relative solvent accessibility of the non-reducing terminal disaccharides of the A- (yellow), B- (green) and C-branches (red) of the S4 glycan over the course of the simulation.

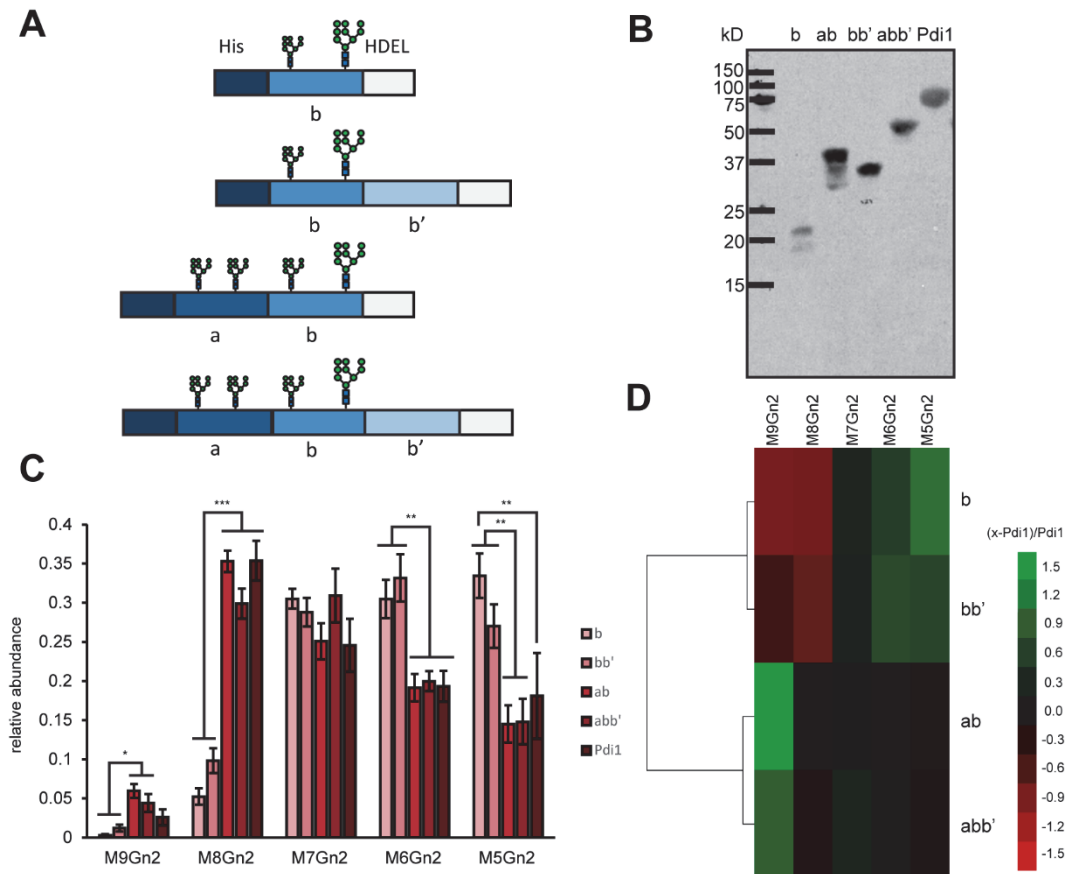


Figure 5. S4 glycan processing of Pdi1p is improved when the a-domain is removed.

(A) Schematic representation of the Pdi1p variants used. His: N-terminal His₁₀ tag, HDEL: C-terminal HDEL ER retention signal. **(B)** Pdi1p variants were expressed and purified. Elution fractions were analyzed by SDS-PAGE and immunoblot using anti-His₄ antibodies. **(C)** Relative abundance of glycoforms on S4 of purified full length and truncated Pdi1p, indicated in different colours. Data represent the mean values of glycan ratios of four independent experiments with error bars indicating the standard deviation * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. **(D)** Similarity representation of Pdi1p variants (indicated at the right) based on the difference in S4 glycoform abundance (given in Fig 6F) to full length Pdi1p. The color scale represents values calculated by the equation: $(x - \text{Pdi1p}) / \text{Pdi1p}$ (x is the average glycoform abundance of a specific variant; Pdi1p is the average glycoform abundance of full length Pdi1p). Similarity between site-specific profiles was calculated as in described in (Figure 3C). See Figure S4 and Table S4 for spectra and raw data.

N-glycans on Pdi1p represent different substrates to Mns1

According to this hypothesis, the five N-linked glycans should display different processing kinetics by a given enzyme. We addressed this by an *in vitro* experiment and analyzed the processing of the glycans of Pdi1p by ER α -mannosidase from *S. cerevisiae*, Mns1. Mns1 recognizes the terminal mannose residues of both, the B- and C-branch of a M9Gn2 glycan and specifically hydrolyses the α -1,2-glycosidic linkage of the terminal mannose of the B-branch (Jelinek-Kelly et al. 1985; Vallee et al. 2000). Pdi1p carrying M9Gn2 glycans was obtained from insect cells that were incubated with the α -mannosidase inhibitor kifunensine during protein production. MS analysis confirmed that Pdi1p was homogeneously glycosylated with M9Gn2 N-glycans on all five sites (Figure 6A and S4). Mns1p was expressed in *T. ni* cells and purified by glutathione affinity chromatography (Figure S4A). *In vitro* assays were performed with a 100x molar excess of glycosylated Pdi1p (Karaveg and Moremen 2005). Aliquots were taken at different time points, and site-specific glycosylation was determined by quantitative MS analysis. Three minutes after initiating the reaction 84%, 80%, 62%, and 54% of M9Gn2 were processed to M8Gn2 on S2, S3, S5, and S1 respectively, whereas the conversion of M9Gn2 on S4 was only 37% (Figure 6B). After ten minutes, all but S4 glycan showed more than 90% processing; conversion to M8Gn2 was completed for all five glycans after 60 min. We concluded that interaction of a glycan with the protein surface competed with enzyme binding and, thus, reduced enzymatic turnover of the glycan: the more favourable this interaction the slower the processing.

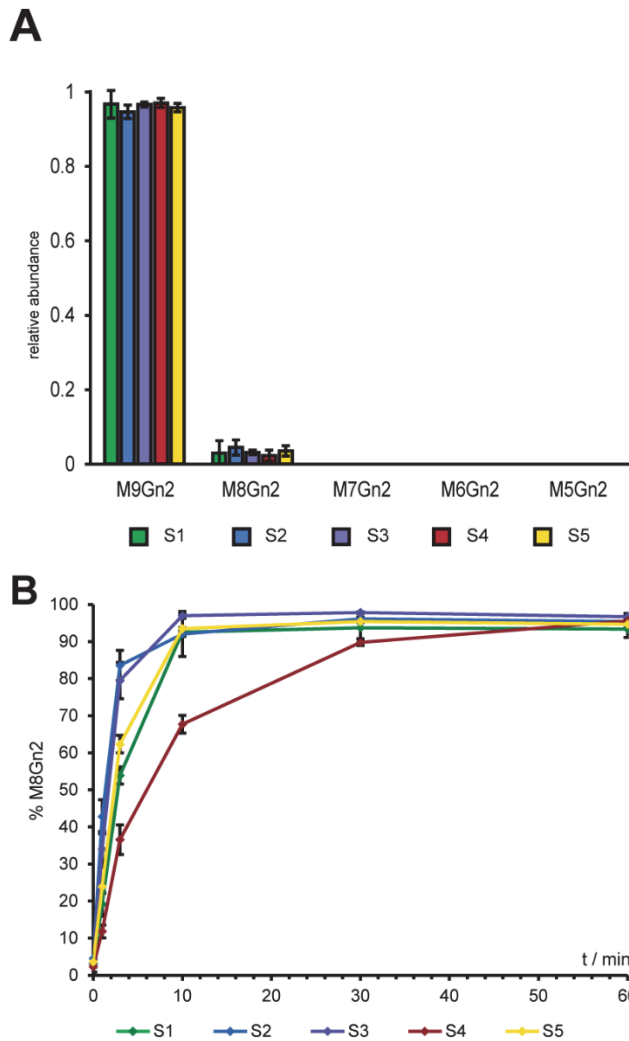


Figure 6. The S4 is processed more slowly by Mns1 *in vitro*.

(A) N-glycan analysis of M9Gn2-Pdi1p. *T. ni* cells were incubated with kifunensine and infected with virus containing an expression copy of GST-tagged Pdi1p. GST-Pdi1p was affinity purified and the GST tag was removed. N-glycan profiles were determined by MS analysis. Data represent the mean values of relative glycan abundances of three independent experiments with error bars indicating the standard deviation of mean. **(B)** M9Gn2-Pdi1p was incubated with Mns1. At the indicated time points aliquots were TCA precipitated and analyzed by MS. Mean values for the relative abundance of M8Gn2 at the different glycosites from three independent experiments were plotted with error bars representing the standard deviation of mean. See Figure S5 and Table S5 for spectra and raw data.

N-glycans of sPdi1p are processed differently in the Golgi

In the absence of the HDEL ER retrieval sequence His₁₀-tagged sPdi1p was secreted into the medium, allowing us to analyse the complete glycan processing pathway. Employing the Extractmfg workflow we found glycan structures reflecting the whole glycosylation machinery of *T. ni* cells, yet, glycans on the five glycosites of Pdi1p were processed differently (Figure 7A and 7B). S1, S2 and S3 glycans showed a similar distribution of oligomannose, paucimannose-type and complex-type structures. However, the ratio of di-fucosylated glycans was significantly higher on S2 and S3 as compared to S1 and oligomannose glycans were almost absent on S5. In contrast to these glycosites, S4 maintained a high ratio of oligomannose oligosaccharides and paucimannose-type glycans that were neither fucosylated nor decorated with a second Gn. Interestingly, one of the abundant glycoforms on S4 was GnM4Gn2, an intermediate product of α -mannosidase II processing (Figure 7B and S5). Hierarchical clustering analysis confirmed that the glycan profiles of S2 and S3, and S1 and S5, respectively, were closely related while the S4 glycan profile was different to all the other sites (Figure 7C). These data confirmed that for the remodelling of N-glycans in the Golgi, the localization of the glycan on the protein surface was a major factor that determined both the structure of an oligosaccharide and the relative abundance of this structure. Consequently, the substrate properties of the five N-linked glycans of sPdi1p differ for a given processing enzyme.

To quantify the *in vivo* activity of different processing enzymes on the five N-glycans we performed a glycan flux analysis using a stoichiometric matrix to describe the connectivity of the network. We calculated the substrate fluxes at each enzyme employing the assumption that the secreted Pdi1p represented a quasi-steady state of the glycan-processing in the cell. Conversion for each reaction were normalized to the incoming fluxes (Figure 7D). This analysis revealed that conversion for enzymes acting early in the pathway were generally higher than those for enzymes acting late in the processing. Substrate conversion by α -mannosidase I and II were significantly lower for the S4 glycan than for all the other glycans. On the other hand, the S2 and S3 glycans were more preferred substrates for α 1,6- and α 1,3-fucosyl transferases than the oligosaccharides on S1 and S5. Thus, altered conversion result in site-specific processing of the N-glycans. Importantly, this flux analysis revealed that only a subset of Golgi processing enzymes was sensitive to the localization of the N-glycan.

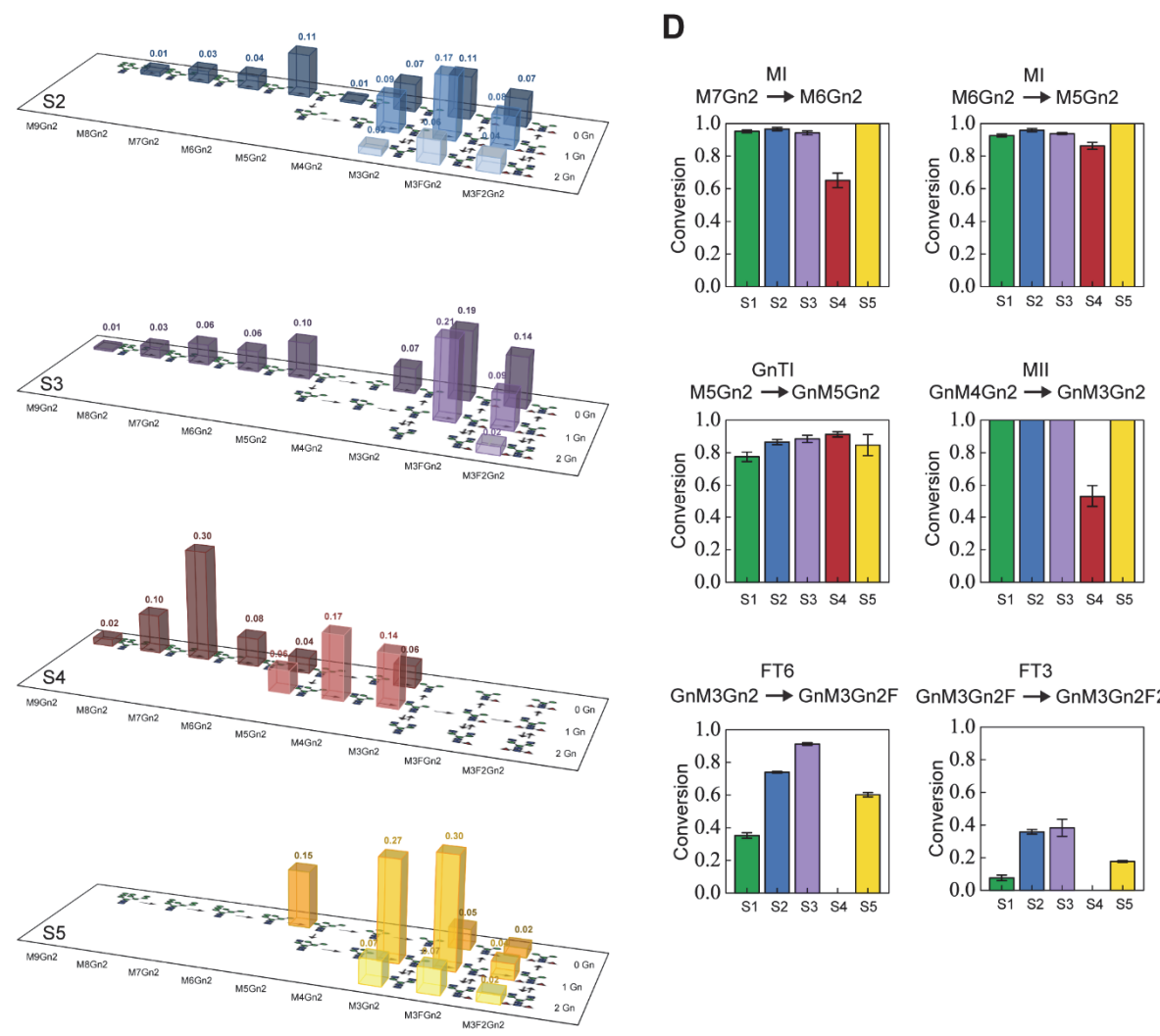
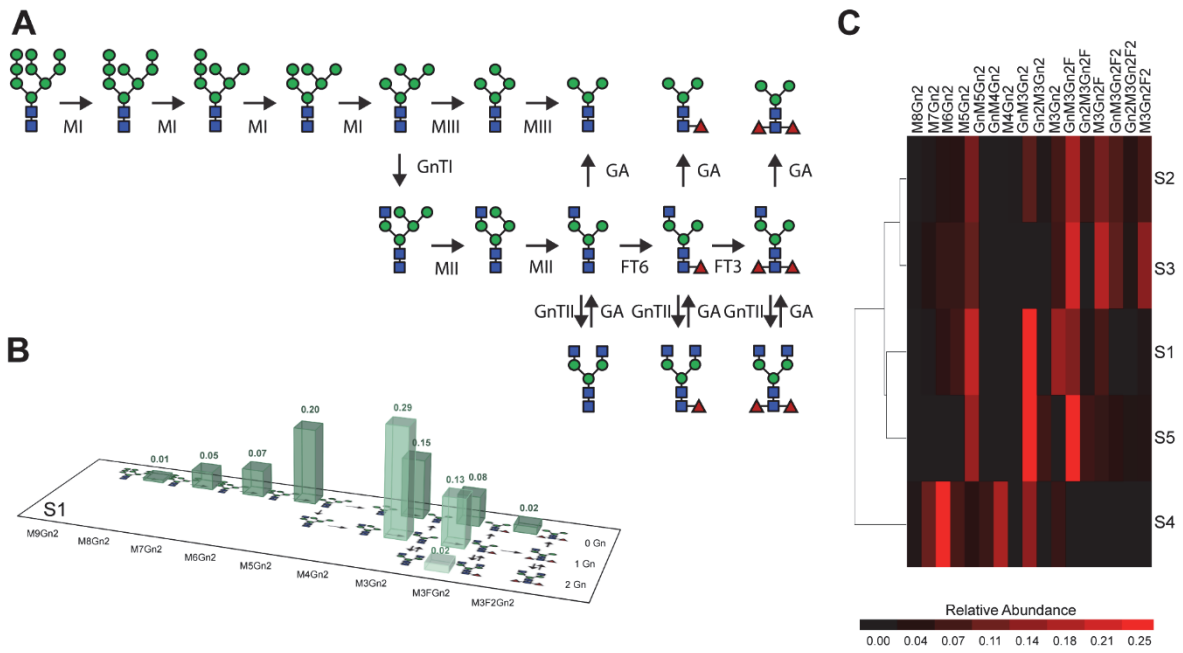


Figure 7. Secreted Pdi1p shows site-specific glycan profiles.

(A) N-glycan processing pathway of *T. ni* showing glycan structures and enzymes (M1, MII, MIII: α -mannosidase I, II and III; GnTI, GnTII: N-acetylglucosamine transferase I and II; GA: N-acetylglucosaminidase; FT6, FT3: α 1,6- and α 1,3-fucosyl transferase). (B) Site-specific glycan profiles of sPdi1p. Data represent the average values calculated from four independent experiments. Relative abundances for each glycoform are depicted above the corresponding column placed at the left of graphic structure representation. The processing pathway from (A) is depicted at the base of the charts. See Table S6 for the complete data set including all glycoforms detected. (C) Pdi1p N-glycosylation sites arranged based on the similarity of their glycoprofiles. The similarity between site-specific glycan profiles and different glycoforms was calculated as described in Figure 3C. (D) Conversions for selected enzymes for each glycosite (S1 to S5). Site-specific enzyme conversion were obtained by glycan flux analysis using the relative abundances of each glycoform from (B). Error bars indicate standard deviation of mean values from four experiments. See Figure S6 and Table S6 for spectra, a complete set of conversions for every enzyme and raw data.

Site-specific processing profile of N-glycans on sPdi1p is dependent upon the protein structure and the Golgi processing machinery

We next analyzed whether protein structure also affect site-specific processing of Pdi1p in the Golgi. Again, we took advantage of the modular structure of Pdi1p and designed truncated variants as in Figure 4D, yet, without the C-terminal HDEL sequence. Constructs were expressed in *T. ni* cells and purified from the culturing media (Figure 8A). After MS-analysis and data processing, we observed that the glycans of the b- and bb'-domain constructs were more processed as compared to full-length sPdi1p and the variants containing the a-domain. The S4 glycan profiles on both b- and bb'-domain constructs contained significantly lower amounts of oligomannose type N-glycans, while the paucimannose-type structure M3Gn2 was more abundant (Figure 8B). Conversion analysis revealed that removal of the a-domain made the S4 glycan a more preferred substrate for α -mannosidases I and II and N-acetylglucosaminidase (Figure 8C). We also noted changes in the processing pattern of the other N-glycans with respect to altered protein structure (data not shown).

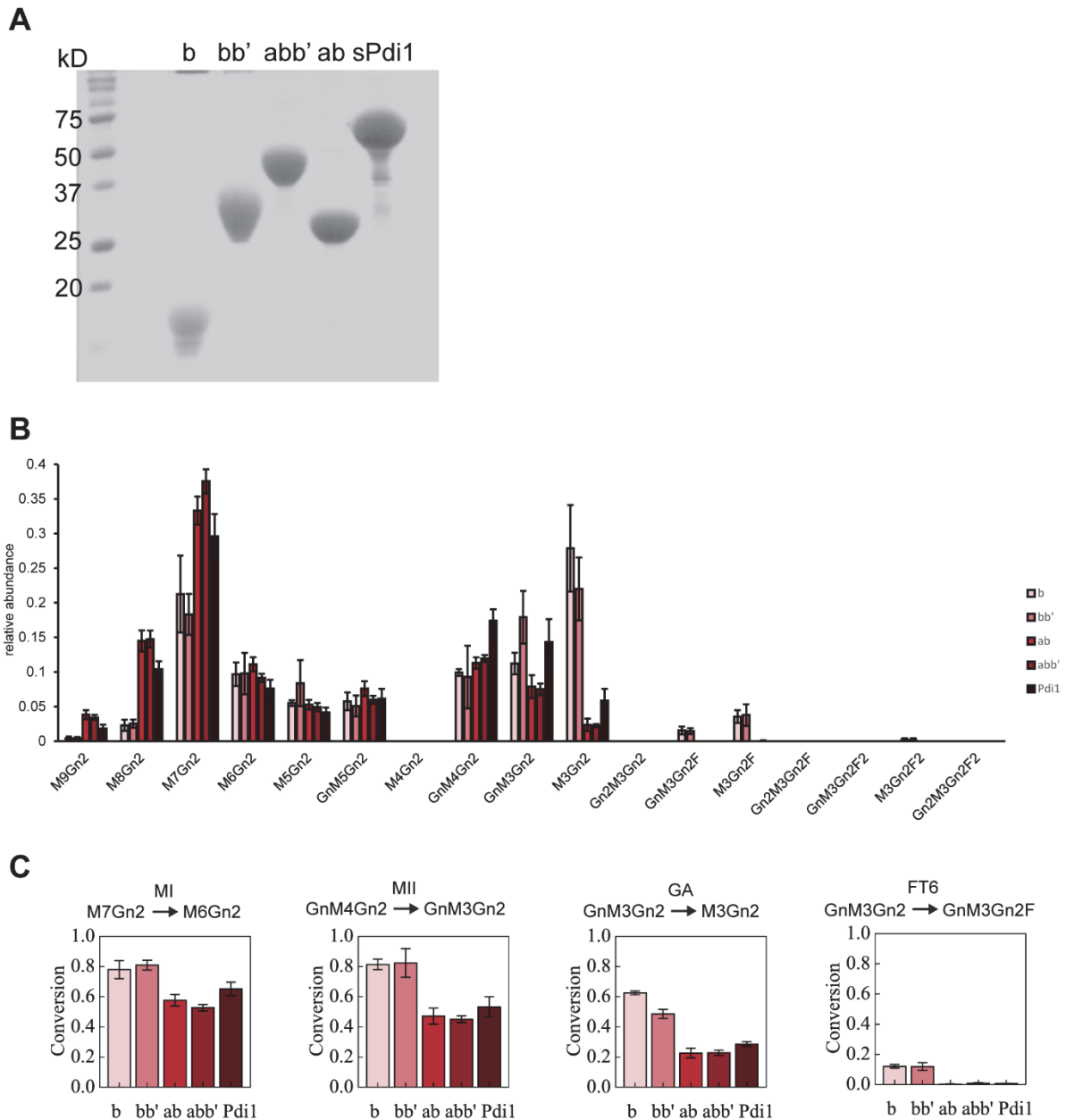


Figure 8. The glycan profile of secreted Pdi1p is altered by changing the protein structure.

(A) Secreted Pdi1p and truncated versions were expressed in insect cells and purified as described in Figure 1. Eluted fractions were analyzed by SDS-PAGE and stained with Coomassie (left/middle). ER retention of sPdi1p was disrupted by the presence of two additional amino acids at the protein's C-terminus (HDELLE) (Raykhel et al. 2007); truncated versions do not contain a HDEL sequence. **(B)** Relative glycan abundances on S4 of purified full length sPdi1p and secreted b-, bb'-, ab-, abb'-variants shown in different colours. Data represent the mean values of glycan ratios of three independent experiments with error bars indicating standard deviation of mean values. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. **(C)** Conversions of S4 N-glycan on sPdi1p variants were calculated as in Figure 6D. Conversions for selected enzymes were plotted for each variant. Error bars indicate standard deviation of mean values from four experiments.

Expression of an additional glycosyltransferase increases site-specific glycan diversity

Microheterogeneity of N-glycan structures on sPdi1p is the result of incomplete conversion by glycan-processing enzymes. In addition, site-specific processing by a subset of these enzymes resulted in site-specific glycan profiles. It is evident that increasing the number of processing steps in such a system will elevate the probability of site-specific N-glycan processing. We therefore expressed β 1,4-galactosyltransferase (GALT-1) from *Caenorhabditis elegans* in insect cells (Figure 9A). GALT-1 acts late in the pathway and adds a β 1,4-galactose to the α 1,6-linked core fucose residue (Titz et al. 2009). We analyzed site-specific glycosylation profiles of sPdi1p purified from cells co-infected with GALT-1 recombinant viruses (Figure 9B). Indeed, we observed site-specific processing: in contrast to S2, S3 and S5 oligosaccharides, GALT-1 did not process the S1 glycan, even though all of the sites presented suitable GALT-1 substrates. Thus, introducing an additional late acting enzyme in the pathway allowed us to differentiate the glycans of Pdi1p further; the glycan profile of S1 deviated from the profile of S5 (Figure 9C). Taken together, the location of the N-linked glycan on a protein in a given setting of glycan-processing enzymes in the secretory pathway defines its processing.

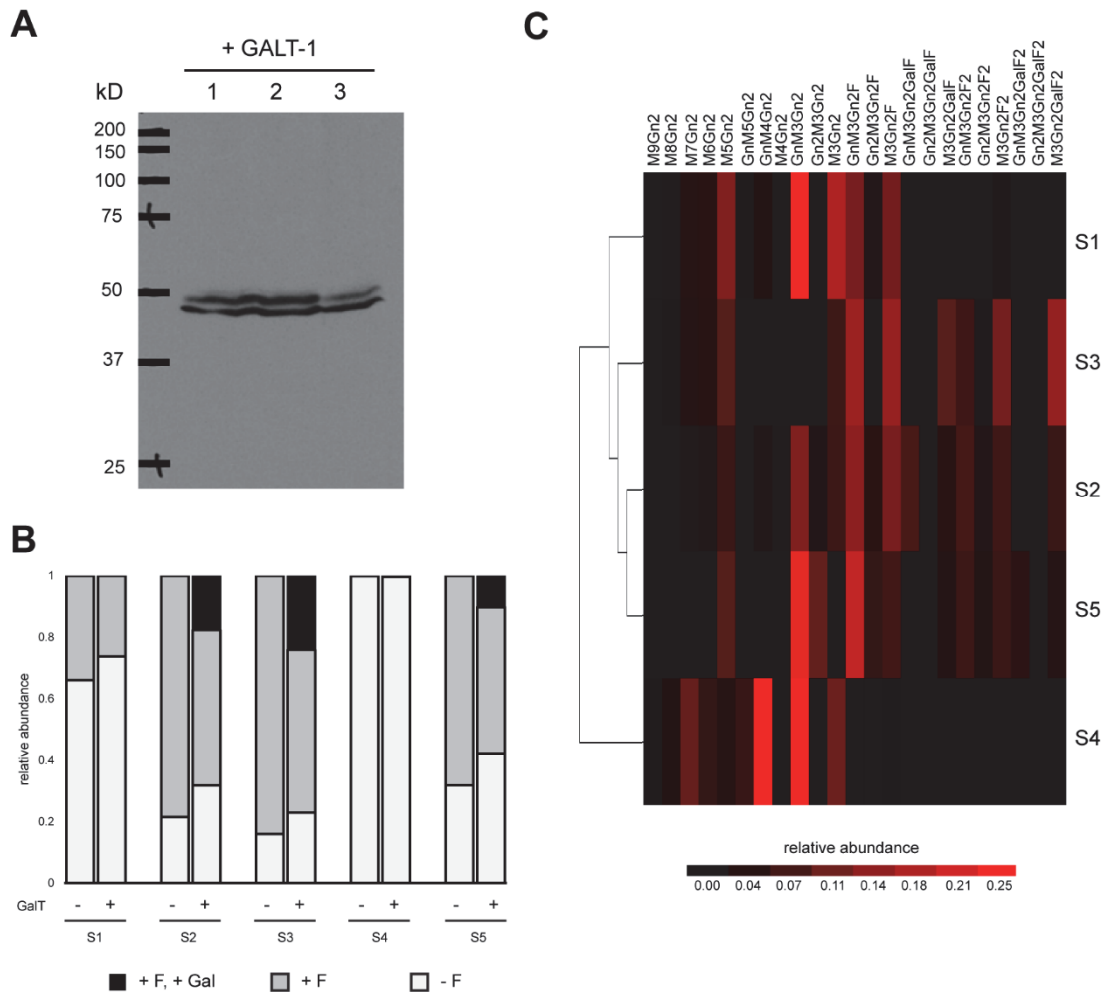


Figure 9. Altering the processing machinery changes the glycan profile of secreted Pdi1p.

(A) GALT-1 expression levels were confirmed by immunoblotting using anti-FLAG antibody in three independent experiments. **(B)** sPdi1p was purified from cells infected with either an unmodified bacmid (-GalT) or a recombinant bacmid carrying an expression copy of GALT-1 (+GalT). Relative glycan abundances without core fucoses (-F), with core fucoses but without galactose (+F), and with galactose attached to the core fucose (+F, +Gal) were calculated for each site. The data represent mean values from three independent experiments. See Table S7 for data on individual glycan structures. **(C)** Similarity of the five glycosites was calculated from relative glycan abundances as described in Figure 3C. See Figure S7 for glycan profiles of S1, S2 and S3, and immunoblot of GALT-1 expression, and Table S7 for raw data.

Discussion

In this study, we exploited baculovirus-driven expression in insect cells to produce model glycoproteins that were retained in the ER and early Golgi (Pdi1p), or secreted into the medium (sPdi1p), respectively (Figure 1). Since yeast Pdi1p carries five N-glycans distributed over the molecule, we could observe remodeling of individual N-linked oligosaccharides at different locations on the same protein while they were presented to the identical cellular glycan-processing machinery for the same amount of time. In addition, the protein has a modular composition of four thioredoxin-like domains that fold independently; hence, we were able to express truncated variants of Pdi1p lacking one or multiple domains without disturbing the structure of the remaining protein (Kemink et al. 1996; Kemink et al. 1999). This set-up allowed us to assess the influence of the protein on glycan remodeling while keeping the cellular environment constant.

Mammalian glycans contain sialic acid residues that are easily lost during ionization in mass spectrometric analyses; consequently, it is difficult to quantify carbohydrates on mammalian glycoproteins (Powell and Harvey 1996). In contrast, insect cells produce fucosylated paucimannose structures without galactose or sialic acid, allowing us to quantify oligosaccharide structures linked to specific glycosites. To assess site-specific glycan composition, we developed an analytic workflow based on mass spectrometry operated in HCD fragmentation mode (Figure 2). Detection of Y1, $^{0,2}X$ and peptide peaks are frequently reported in CID MS/MS of glycopeptides by Q-TOF or TOF-TOF MS, but this peak triplet has not been exploited for identification in the HCD fragmentation method (Krokhin et al. 2004; Sparbier et al. 2007). Our data demonstrated that using the peak triplet and characteristic peptide fragment ions is a suitable tool for glycopeptide identification without further tandem mass spectrometry. Moreover, we developed and evaluated a reliable quantification method of each glycoform that allows to partly automate and speed up data analysis.

Although the glycan-processing machinery of insect cells used in this study is less complex than in mammalian cells, it implements the same underlying principles. According to the current view, a glycoprotein may transit through the Golgi too quickly to be processed completely by every possible enzyme (Moremen et al. 2012; Stanley 2011). In addition, many glycosyl hydrolases and transferases compete for the same oligosaccharide substrate producing additional heterogeneity of glycan structures

(Geisler and Jarvis 2012; Schachter 1991). Earlier studies suggested that the local protein surface influences enzyme accessibility to individual glycans (Hsieh et al. 1983a; Hsieh et al. 1983b; Hubbard 1988; Trimble et al. 1983); however, these studies did not provide direct evidence to support this hypothesis. Analysing the glycan profiles of Pdi1p and sPdi1p, we found that the five glycosites displayed individual carbohydrate profiles throughout the secretory pathway (Figures 3 & 7). In particular, the S4 glycan was less processed in both, the ER and Golgi compartments relative to the other glycans. Molecular dynamics simulations showed that the terminal mannose residues of an M9Gn2 glycan on S4 made long lasting contacts with specific amino acid residues of the a- and b-domain (Figure 4A). These interactions might reduce enzyme accessibility and, thus, processing of the S4 glycan. Along this line, ER mannosidase processed *in vitro* the S4 glycan more slowly than its counterparts on the other glycosites. Consequently, when we removed the a-domain from the protein *in vivo*, the glycan on S4 became more accessible and the resulting glycan profile shifted to more processed glycan structures (Figures 4 & 8). In addition, implementing glycan flux analysis we identified the rate-limiting steps for each glycosite. Interestingly, it is only a subset of processing enzymes that are kinetically controlled (thereby generating microheterogeneity of the glycan structures) and some of these enzymes are affected by the location of the N-glycan on the protein (resulting in site-specific processing). Taken together, our data show that the structure of a glycoprotein can control the processing of some N-linked oligosaccharides at different stages of the glycosylation pathway.

Site-specific N-glycans can affect the function of glycoproteins, e.g. immunoglobulins (Ferrara et al. 2011; Niwa et al. 2004; Sazinsky et al. 2008; Scallon et al. 2007), but for most glycoproteins, this has not been addressed experimentally. It is evident that the quantitative analysis of site-specific glycans will be essential for future research in this direction. The MS-based, quantitative glycoproteomics method presented in this report will become a valuable tool to define more precisely the role of the carbohydrate in glycoprotein function. Similarly, the mathematical modelling of glycan processing based on quantitative glycoproteomics data will be essential to experimentally approach and understand the function of the Golgi and will affect the biotechnological production of glycoproteins.

Materials and Methods

Reagents

The monoclonal anti-His antibody was purchased from Qiagen AG (Cat. No. 34670). Monoclonal anti-GST antibody was purchased from Rockland (Cat. No. 600-101-200). *Spodoptera frugiperda* Sf21 and *Trichoplusia ni* Hi5 cells were obtained from K. Locher and cultivated in Sf-900 II SFM medium (Invitrogen, Cat.No 10902104).

Construction of plasmids for baculovirus-based protein expression

Plasmids used in this study are summarized in Supplementary Table 1. Plasmids were constructed using standard cloning protocols. Open reading frames of all constructs were confirmed by nucleotide sequencing.

pRG105: The plasmid pRG105 was constructed to allow for baculovirus-based expression of secreted proteins fused to an N-terminal affinity tag. gp67 secretion signal peptide followed by GST was PCR-amplified from pAcSecG2T baculovirus transfer vector (Fischer Scientific, Cat.No. BDB554797) using primers RG395 (5'-TGG GCG CGC ATG CTA CTA GTA AAT CAG TCA CAC-3') and RG396 (5'-CCC AAG CTT TTA CTC GAG CTG CAG AGG CCT GAG CTC GGA TCC ACG CGG AAC CAGA-3'). The resulting PCR fragment was cloned into the BssHII and HindIII restriction sites of pRG74 (Gauss et al. 2011) to generate pRG100. Next pRG100 was used as a template to introduce His₁₀ tag following a cleavable signal sequence by PCR using RG394 (5'- AAG TGG TTC GCA TCC TCG GTTT-3') and RG398 (5'- CCC GAG CTC CAG GGG CCC CTG GAA CAG AAC TTC CAG ATG GTG ATG GTG ATG GTG ATG GTG ATG GTG GGT ACC CGC AAA GGC AGA ATG CGC-3') primers producing pRG101. Finally, the *PDI1* gene (without the signal peptide) was excised from pRG84 (Gauss et al. 2011) by digestion with SacI and XhoI and introduced into pRG101 generating pRG105. ER retention of sPdi1p is disrupted by the presence of two additional amino acids at the protein's C-terminus (HDELLE).

pRG143, pRG144, pRG145, pRG146: Truncated versions of the *PDI1* gene were PCR-amplified from pRG84 using the following primer pairs: RG476 (5'- CCG AGC TCA CTA GTC CGG CTG TCG CCG TTG TTG CTG ATC-3') and RG477 (5'- CCC TGC AGC TCG AGG TAG GGC AAG GCT TCC ACT TGC AACC-3') (b-domain);

RG475 (5'- CCG AGC TCA CTA GTG AGG CTG TGG CCC CTG AAG ACT CC-3') and RG477 (ab-domains); RG475 and RG478 (5'- CCC TGC AGC TCG AGA CCT TTC AAG AAG TCC TTA ACC AAAG-3') (abb'-domains); RG476 and RG478 (bb'-domains). To allow for expression of ER-retained proteins, the respective PCR products were cloned into the *SacI* and *XhoI* restriction sites of pRG85 upstream of the HDEL sequence thus replacing the *PDI1* gene and yielding pRG143 (b), pRG144 (ab), pRG145 (abb') and pRG146 (bb').

pRG135, pRG136, pRG137, pRG138: To allow for expression and secretion of truncated Pdi1p variants, the coding sequences of pRG143, pRG144, pRG145, and pRG146 were excised by *SacI* and *XhoI*. Fragments were ligated into the respective restriction sites of pRG105 downstream of the gp67 signal sequence yielding pRG135 (b), pRG136 (ab), pRG137 (abb') and pRG138 (bb').

Expression and purification of recombinant proteins in insect cells

Standard procedures were used to generate recombinant baculoviruses in Sf21 cells (Murhammer, 2007). Recombinant viruses were collected after 72 h. Viruses were amplified by infecting Sf21 cell cultures with virus solution in a 1:10 and 1:100 (v:v) ratio. Expression of intracellular or secreted recombinant proteins was monitored by SDS-PAGE and immunoblotting after each virus amplification step. For protein expression, *T. ni* Hi5 cells were diluted to 1 million cells per ml in shaker flasks and infected 1:100 (v:v) with recombinant viruses. For the analysis of site-specific processing of sPdi1p by GALT-1, sPdi1p was expressed in Hi5 cells co-infected with GALT-1 recombinant viruses.

T. ni Hi5 cells were diluted to 1 million cells per ml in shaker flasks and infected 1:100 (v:v) with recombinant viruses. Infected cells were incubated at 27°C for 48 h. To purify His₁₀-tagged ER retained proteins, cells were pelleted by centrifugation (600 rcf, 5 min), washed with 1xPBS and lysed with lysis buffer (1% TritonX-100, 1x Protease Inhibitor cocktail (Roche, 11873580001) in 1xPBS). The lysates were centrifuged at 3500 rcf for 10 min and the soluble fraction was filtered using PES-membrane 0.2 µM filters (TPP). For the purification of secreted proteins, the culture supernatant was cleared by centrifugation at 3500 rcf for 10 min and filtered. Affinity purification was performed on gravity flow columns filled with 1 ml NiNTA beads (Protino, 745400.100) equilibrated either with 10 column volumes (CV) of lysis buffer or medium. The bound

fraction was washed with 15 CV washing buffer (30 mM imidazole in 1xPBS) and proteins were eluted with 4 CV of elution buffer (250 mM imidazole in 1xPBS). Eluted fractions were concentrated on Amicon Ultra-4 Centrifugal Filter Devices. For MS analysis samples were prepared as described below. Proteins that were stored after buffer exchange to 1xPBS and kept frozen at -20°C.

To purify GST-tagged M9Gn2-Pdi1p, Hi5 cells were infected with recombinant virus in the presence of 10 µM kifunensine (Sigma Aldrich, K1140-1MG) and incubated for 48 h. Mns1p was purified from Hi5 cells 48 h after infection with recombinant virus. Following cell lysis, proteins were bound to Glutathione Sepharose 4B matrix (GE Healthcare, 17-0756-01). Bound proteins were washed with 15 CV of 1xPBS and subsequently with 15 CV of PreScission Protease (PSP; VWR, 27-0843-01) activity buffer (150 mM NaCl, 1 mM EDTA, 1 mM DTT in 50 mM Tris pH 7.0). On column cleavage was performed by incubating bound proteins with 20 U of PSP in 1 ml of PSP activity buffer for 16 hours. Released proteins were collected by gravity flow. Sample were concentrated and the buffer was exchanged to Mns1p activity buffer (150 mM NaCl, 5 mM CaCl₂ in 20 mM MES pH 7.0) (Karaveg and Moremen 2005).

***In vitro* Mns1p activity assay**

Fifty µM Pdi1p were mixed with 0.5 µM Mns1p in total volume of 85 µl and incubated at 37°C. Aliquots of 17 µl were taken at different time points. Each aliquot was mixed with 3 µl of 100% TCA, vortexed, and incubated on ice for 5 min. After centrifugation at 20,000xg and 4°C for 5 min, the pellet was washed three times with 500 µl of ice cold acetone. The pellet was air dried and stored at -20°C. For MS analysis protein pellets were resuspended in 50 µl of 8M urea in 50mM ammonium bicarbonate (pH 8.5).

Sample preparation and glycopeptide analysis by nanoHPLC-HCD-MS/MS

Purified proteins were loaded onto a filter device (Microcon YM-30, Millipore) and washed three times with water. Usually, 50 µg were reduced by 50 mM dithiothreitol in 50 mM ammonium bicarbonate buffer (pH 8.5) at 37°C for 1 h, following by alkylation with 65 mM iodoacetamide at 37°C in the dark for 1 h. After four washing steps with ammonium bicarbonate buffer, proteins were digested by trypsin (molar ratio 50:1,

Promega, Cat. No. V5111) at 37°C for 16 h. Digested peptides and glycopeptides were collected by centrifugation and dried in a speedvac. Samples were desalted by Zip-Tip C18 (Millipore) prior to nanoLC-MS/MS analysis. Samples were analyzed on a calibrated LTQ-Orbitrap Velos mass spectrometer (Thermo Fischer Scientific) coupled to an Eksigent-Nano-HPLC system (Eksigent Technologies). Peptides were resuspended in 2.5% acetonitrile with 0.1% formic acid and loaded onto a self-made tip column (75 µm × 80 mm) packed with reverse phase C18 material (AQ, 3 µm 200 Å, Bischoff GmbH). Peptides were eluted with a flow rate of 200 nl per min by a gradient from 3 to 30% of solution B (99.9% ACN, 0.1% FA) applied for 22 min, 50% B applied for 25 min, 97% B applied for 27 min. One scan cycle comprised of a full scan MS survey spectrum, followed by up to 10 sequential HCD MS/MS on the most intense signals above a threshold of 2000. Full-scan MS spectra (800–2000 m/z) were acquired in the FT-Orbitrap at a resolution of 60,000 at 400 m/z, while HCD MS/MS spectra were recorded in the FT-Orbitrap at a resolution of 15,000 at 400 m/z. HCD was performed with a target value of 1e5 and stepped collision energy rolling from 35, 40 and 45 V was applied. AGC target values were 5e5 for full FTMS. For all experiments, dynamic exclusion was used with a single repeat count, 15 s repeat duration, and 60 s exclusion duration.

Identification and quantification of different glycoforms sharing the same peptide backbone

MS and MS/MS data were processed into the Mascot generic format (mgf) file. Extractmgf was written in Perl to perform the following steps: glycopeptides were initially extracted from mgf file by identifying glycan oxonium ions, [HexNAc]⁺ 204.09 and [HexNAc+Hex]⁺ 366.12. Glycopeptides corresponding to each site were further obtained by sorting out the corresponding Y1 ion individually from extracted mgf file. All MS/MS spectra were confirmed manually. Here, XCalibur 2.2 sp1.48 was used for data processing. For quantification, extracted_ion chromatography (XIC) of each glycoform was plotted by its individual m/z with the mass tolerance of 5 ppm. Peak area was defined manually and integrated by the program. The relative amount of each glycoform sharing same peptide backbone was calculated as following equation:

$$\text{Relative amount of each glycoform (\%)} = \frac{\text{Peak area of each glycoform}}{\text{Sum of peak area of all glycoforms}} \times 100\%$$

Molecular modelling

Three dimensional structure preparation

Initial starting coordinates for the a- and b- subdomains of Pdi1p were taken from PDB: 2B5E. Models were generated for the unglycosylated domains and for a glycoform with M9Gn2 at site 4 (S4) using the online glycoprotein builder available on GLYCAM-Web (Kirschner et al. 2008; Woods Group (2005-2014)). Using tleap (Case et al. 2014) the structures were placed in a truncated octahedron of TIP3P (Jorgensen et al. 1983) water with an 8 Å distance buffer from the solute to the edge of the periodic box and an 0.5 Å spacing between solute and water molecules. Prior to solvation, all waters of crystallization were removed, and the systems were neutralized with sodium counter-ions.

Energy minimization and molecular dynamics simulations

Energy minimization (10,000 steps of steepest descent, followed by 10,000 steps of conjugate gradient) was performed in a stepwise fashion with initially a 5 kcal mol⁻¹ Å⁻² restraint applied to the solute (protein and glycan) heavy atoms. This was followed by minimization with restraints only on the glycan heavy atoms, before a final minimization with no restraints on any atoms. Minimizations were performed with sander.MPI in AMBER14 (Case et al. 2014). Molecular dynamics simulations were performed with the CUDA implementation of PMEMD (Gotz et al. 2012; Salomon-Ferrer et al. 2013) in AMBER14 using constant pressure (nPT) conditions. A Berendsen barostat with a time constant of 1 ps was employed for pressure regulation, while a Langevin thermostat with a collision frequency of 2 ps⁻¹ was employed for temperature regulation. A nonbonded interaction cutoff of 8 Å was employed. Long-range electrostatics were treated with the particle-mesh Ewald (PME) method (Darden et al. 1993). Covalent bonds involving hydrogen were constrained with the SHAKE algorithm allowing a time step of 2 fs (Ryckaert et al. 1977). Cartesian restraints were applied to the protein C α atoms (5 kcal mol⁻¹ Å⁻²) during a 100 ps heating stage from 5 to 300 K, which was followed by a 100 ps equilibration phase. All restraints were then removed for a 1 ns equilibration phase prior to a 500 ns production phase.

Data analysis

RMSD of the a-domain: The cpptraj (Roe and Cheatham 2013) module was used to extract 1,000 snapshots from each simulation at 500 ps intervals, these frames were aligned on the b-domain of the crystal structure, and the root mean squared deviation (RMSD) computed for the a-domain. *Glycan branch accessibility:* NACCESS (Hubbard and Thornton 1993) was employed with a 6 Å radius probe to compute the accessibility of the non-reducing terminal disaccharides in the A-, B- and C-branches of the glycan over 100 snapshots from the simulation; the approximate end-to-end distance of the disaccharide being 12 Å. The values were normalized relative to the accessibility of a Man α 1-2Man α disaccharide.

Glycan flux analysis of *T. ni* Golgi glycosylation network

The connectivity of the network was described by the stoichiometric matrix \mathbf{S} where the low activity of α -ManIII from M4Gn2 to M3Gn2 was neglected to obtain an acyclic system. Assuming a pseudo steady state of the system, the fluxes \mathbf{v} were constrained to (Antoniewicz 2013):

$$S\mathbf{v} = 0 \quad (1)$$

The measured glycosylation species were considered as constant fluxes, \mathbf{v}_m , out of the Golgi apparatus and were used to calculate the fluxes, \mathbf{v}_c , inside the network. M9Gn2 was assumed as the only structure entering the late ER and Golgi apparatus (Krambeck et al. 2009). Consequently, equation (1) was split accordingly:

$$S_m\mathbf{v}_m + S_c\mathbf{v}_c = 0 \quad (2)$$

where \mathbf{S}_m was the stoichiometric matrix for the measured fluxes and \mathbf{S}_c for the calculated internal fluxes. The inverse $(\mathbf{S}_c)^\#$ was used to solve equation (2):

$$\mathbf{v}_c = -(\mathbf{S}_c)^\# S_m\mathbf{v}_m \quad (3)$$

The measured structures fully determined the 19 internal fluxes and the result of equation (3) was unique. The conversion for every enzymatic step was calculated by normalizing the outgoing to the incoming flux.

Acknowledgments

We thank Emma Marriott for excellent technical assistance. We thank Tsung-Hsien Pu and Prof. Kay-Hooi Khoo of the Core Facility for Protein structural Analysis, Academia Sinica, Taipei, Taiwan for their help in the early stage of the software development. We thank the Functional Genomics Centre Zurich for help with the mass spectrometry analysis. R.J.W. would like to thank the National Institutes of Health (P41GM103390, R01GM094919 (EUREKA), as well as the Science Foundation of Ireland (08/IN.1/B2070) for support. This work was supported by ETH Research Grant No. 07 10-1 to M.A. and R.G and Swiss National Science Foundation Grant No. 31003A_127098 to M.A.

References

- Aebi M, Bernasconi R, Clerc S, Molinari M. 2010. N-glycan structures: recognition and processing in the ER. *Trends Biochem Sci*, 35:74-82.
- Antoniewicz MR. 2013. Dynamic metabolic flux analysis--tools for probing transient states of metabolic networks. *Curr Opin Biotechnol*, 24:973-978.
- Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, T.E. Cheatham I, Darden TA, Duke RE, Gohlke H, *et al.* 2014. AMBER 14. University of California, San Francisco.
- Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An N · log (N) method for Ewald sums in large systems. *The Journal of chemical physics*, 98:10089.
- Dell A, Galadari A, Sastre F, Hitchen P. 2010. Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int J Microbiol*, 2010:148178.
- Dell A, Reason AJ, Khoo KH, Panico M, McDowell RA, Morris HR. 1994. Mass spectrometry of carbohydrate-containing biopolymers. *Methods Enzymol*, 230:108-132.
- Faid V, Evjen G, Tollersrud OK, Michalski JC, Morelle W. 2006. Site-specific glycosylation analysis of the bovine lysosomal alpha-mannosidase. *Glycobiology*, 16:440-461.
- Ferrara C, Grau S, Jager C, Sondermann P, Brunker P, Waldhauer I, Hennig M, Ruf A, Rufer AC, Stihle M, *et al.* 2011. Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose. *Proceedings of the National Academy of Sciences of the United States of America*, 108:12669-12674.
- Gauss R, Kanehara K, Carvalho P, Ng DT, Aebi M. 2011. A complex of Pdi1p and the mannosidase Htm1p initiates clearance of unfolded glycoproteins from the endoplasmic reticulum. *Molecular cell*, 42:782-793.

Geisler C, Jarvis DL. 2012. Substrate specificities and intracellular distributions of three N-glycan processing enzymes functioning at a key branch point in the insect N-glycosylation pathway. *The Journal of biological chemistry*, 287:7084-7097.

Go EP, Irungu J, Zhang Y, Dalpathado DS, Liao HX, Sutherland LL, Alam SM, Haynes BF, Desaire H. 2008. Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. *J Proteome Res*, 7:1660-1674.

Gotz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. 2012. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput*, 8:1542-1555.

Hebert DN, Garman SC, Molinari M. 2005. The glycan code of the endoplasmic reticulum: asparagine-linked carbohydrates as protein maturation and quality-control tags. *Trends Cell Biol*, 15:364-370.

Heikinheimo P, Helland R, Leiros HK, Leiros I, Karlsen S, Evjen G, Ravelli R, Schoehn G, Ruigrok R, Tollersrud OK, *et al.* 2003. The structure of bovine lysosomal alpha-mannosidase suggests a novel mechanism for low-pH activation. *J Mol Biol*, 327:631-644.

Hsieh P, Rosner MR, Robbins PW. 1983a. Host-dependent variation of asparagine-linked oligosaccharides at individual glycosylation sites of Sindbis virus glycoproteins. *The Journal of biological chemistry*, 258:2548-2554.

Hsieh P, Rosner MR, Robbins PW. 1983b. Selective cleavage by endo-beta-N-acetylglucosaminidase H at individual glycosylation sites of Sindbis virion envelope glycoproteins. *The Journal of biological chemistry*, 258:2555-2561.

Hua S, Nwosu CC, Strum JS, Seipert RR, An HJ, Zivkovic AM, German JB, Lebrilla CB. 2012. Site-specific protein glycosylation analysis with glycan isomer differentiation. *Anal Bioanal Chem*, 403:1291-1302.

Hubbard SC. 1988. Regulation of glycosylation. The influence of protein structure on N-linked oligosaccharide processing. *The Journal of biological chemistry*, 263:19303-19317.

Hubbard SJ, Thornton JM. 1993. "NACCESS", Computer Program. Department of Biochemistry and Molecular Biology, University College London.

Jelinek-Kelly S, Akiyama T, Saunier B, Tkacz JS, Herscovics A. 1985. Characterization of a specific alpha-mannosidase involved in oligosaccharide processing in *Saccharomyces cerevisiae*. *The Journal of biological chemistry*, 260:2253-2257.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926-935.

Karaveg K, Moremen KW. 2005. Energetics of substrate binding and catalysis by class 1 (glycosylhydrolase family 47) alpha-mannosidases involved in N-glycan processing and endoplasmic reticulum quality control. *The Journal of biological chemistry*, 280:29837-29848.

Kato T, Kajikawa M, Maenaka K, Park EY. 2010. Silkworm expression system as a platform technology in life science. *Appl Microbiol Biotechnol*, 85:459-470.

Kelleher DJ, Gilmore R. 2006. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology*, 16:47R-62R.

Kemmink J, Darby NJ, Dijkstra K, Nilges M, Creighton TE. 1996. Structure determination of the N-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear ¹³C/¹⁵N NMR spectroscopy. *Biochemistry*, 35:7684-7691.

Kemmink J, Dijkstra K, Mariani M, Scheek RM, Penka E, Nilges M, Darby NJ. 1999. The structure in solution of the b domain of protein disulfide isomerase. *J Biomol NMR*, 13:357-368.

Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ. 2008. GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.*, 29:622-655.

Krambeck FJ, Bennun SV, Narang S, Choi S, Yarema KJ, Betenbaugh MJ. 2009. A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data. *Glycobiology*, 19:1163-1175.

Krokhin O, Ens W, Standing KG, Wilkins J, Perreault H. 2004. Site-specific N-glycosylation analysis: matrix-assisted laser desorption/ionization quadrupole-quadrupole time-of-flight tandem mass spectral signatures for recognition and identification of glycopeptides. *Rapid Commun Mass Spectrom*, 18:2020-2030.

Mast SW, Moremen KW. 2006. Family 47 alpha-mannosidases in N-glycan processing. *Methods Enzymol*, 415:31-46.

Moremen KW, Tiemeyer M, Nairn AV. 2012. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*, 13:448-462.

Nagae M, Yamaguchi Y. 2012. Function and 3D Structure of the N-Glycans on Glycoproteins. *Int J Mol Sci*, 13:8398-8429.

Niwa R, Shoji-Hosaka E, Sakurada M, Shinkawa T, Uchida K, Nakamura K, Matsushima K, Ueda R, Hanai N, Shitara K. 2004. Defucosylated chimeric anti-CC chemokine receptor 4 IgG1 with enhanced antibody-dependent cellular cytotoxicity shows potent therapeutic activity to T-cell leukemia and lymphoma. *Cancer research*, 64:2127-2133.

North SJ, Huang HH, Sundaram S, Jang-Lee J, Etienne AT, Trollope A, Chalabi S, Dell A, Stanley P, Haslam SM. 2010. Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. *The Journal of biological chemistry*, 285:5759-5775.

Parekh RB, Tse AG, Dwek RA, Williams AF, Rademacher TW. 1987. Tissue-specific N-glycosylation, site-specific oligosaccharide patterns and lentil lectin recognition of rat Thy-1. *The EMBO journal*, 6:1233-1244.

Powell AK, Harvey DJ. 1996. Stabilization of sialic acids in N-linked oligosaccharides and gangliosides for analysis by positive ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom*, 10:1027-1032.

Raykhel I, Alanen H, Salo K, Jurvansuu J, Nguyen VD, Latva-Ranta M, Ruddock L. 2007. A molecular specificity code for the three mammalian KDEL receptors. *The Journal of cell biology*, 179:1193-1204.

Roe DR, Cheatham TE. 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.*, 9:3084-3095.

Roepstorff P, Fohlman J. 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom*, 11:601.

Ryckaert J-P, Ciccotti G, Berendsen HJC. 1977. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.*, 23:327-341.

Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. 2013. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.*, 9:3878-3888.

Sazinsky SL, Ott RG, Silver NW, Tidor B, Ravetch JV, Wittrup KD. 2008. Aglycosylated immunoglobulin G1 variants productively engage activating Fc receptors. *Proceedings of the National Academy of Sciences of the United States of America*, 105:20167-20172.

Scallon BJ, Tam SH, McCarthy SG, Cai AN, Raju TS. 2007. Higher levels of sialylated Fc glycans in immunoglobulin G molecules can adversely impact functionality. *Molecular immunology*, 44:1524-1534.

Schachter H. 1991. The 'yellow brick road' to branched complex N-glycans. *Glycobiology*, 1:453-461.

Segu ZM, Mechref Y. 2010. Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Commun Mass Spectrom*, 24:1217-1225.

Shental-Bechor D, Levy Y. 2009. Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Curr Opin Struct Biol*, 19:524-533.

Shi X, Jarvis DL. 2007. Protein N-glycosylation in the baculovirus-insect cell system. *Curr Drug Targets*, 8:1116-1125.

Sparbier K, Asperger A, Resemann A, Kessler I, Koch S, Wenzel T, Stein G, Vorweg L, Suckau D, Kostrzewa M. 2007. Analysis of glycoproteins in human serum by means of glycospecific magnetic bead separation and LC-MALDI-TOF/TOF analysis with automated glycopeptide detection. *J Biomol Tech*, 18:252-258.

Stanley P. 2011. Golgi glycosylation. *Cold Spring Harb Perspect Biol*, 3.

Tian G, Xiang S, Noiva R, Lennarz WJ, Schindelin H. 2006. The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell*, 124:61-73.

Titz A, Butschi A, Henrissat B, Fan YY, Hennet T, Razzazi-Fazeli E, Hengartner MO, Wilson IB, Kunzler M, Aebi M. 2009. Molecular basis for galactosylation of core fucose residues in invertebrates: identification of *Caenorhabditis elegans* N-glycan core alpha1,6-fucoside beta1,4-galactosyltransferase GALT-1 as a member of a novel glycosyltransferase family. *The Journal of biological chemistry*, 284:36223-36233.

Trimble RB, Maley F, Chu FK. 1983. GlycoProtein biosynthesis in yeast. protein conformation affects processing of high mannose oligosaccharides on carboxypeptidase Y and invertase. *The Journal of biological chemistry*, 258:2562-2567.

Vallee F, Lipari F, Yip P, Sleno B, Herscovics A, Howell PL. 2000. Crystal structure of a class I alpha1,2-mannosidase involved in N-glycan processing and endoplasmic reticulum quality control. *The EMBO journal*, 19:581-588.

van Kooyk Y, Rabinovich GA. 2008. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat Immunol*, 9:593-601.

Varki A. 1998. Factors controlling the glycosylation potential of the Golgi apparatus. *Trends Cell Biol*, 8:34-40.

Wilkinson B, Gilbert HF. 2004. Protein disulfide isomerase. *Biochim Biophys Acta*, 1699:35-44.

Wisniewski JR, Zougman A, Nagaraj N, Mann M. 2009. Universal sample preparation method for proteome analysis. *Nat Methods*, 6:359-362.

Woods Group. (2005-2014). GLYCAM Web. Athens, GA: Complex Carbohydrate Research Center, University of Georgia (<http://www.glycam.com>).

Yu X, Baruah K, Harvey DJ, Vasiljevic S, Alonzi DS, Song BD, Higgins MK, Bowden TA, Scanlan CN, Crispin M. 2013. Engineering hydrophobic protein-carbohydrate interactions to fine-tune monoclonal antibodies. *Journal of the American Chemical Society*, 135:9723-9732.

Supplementary Data

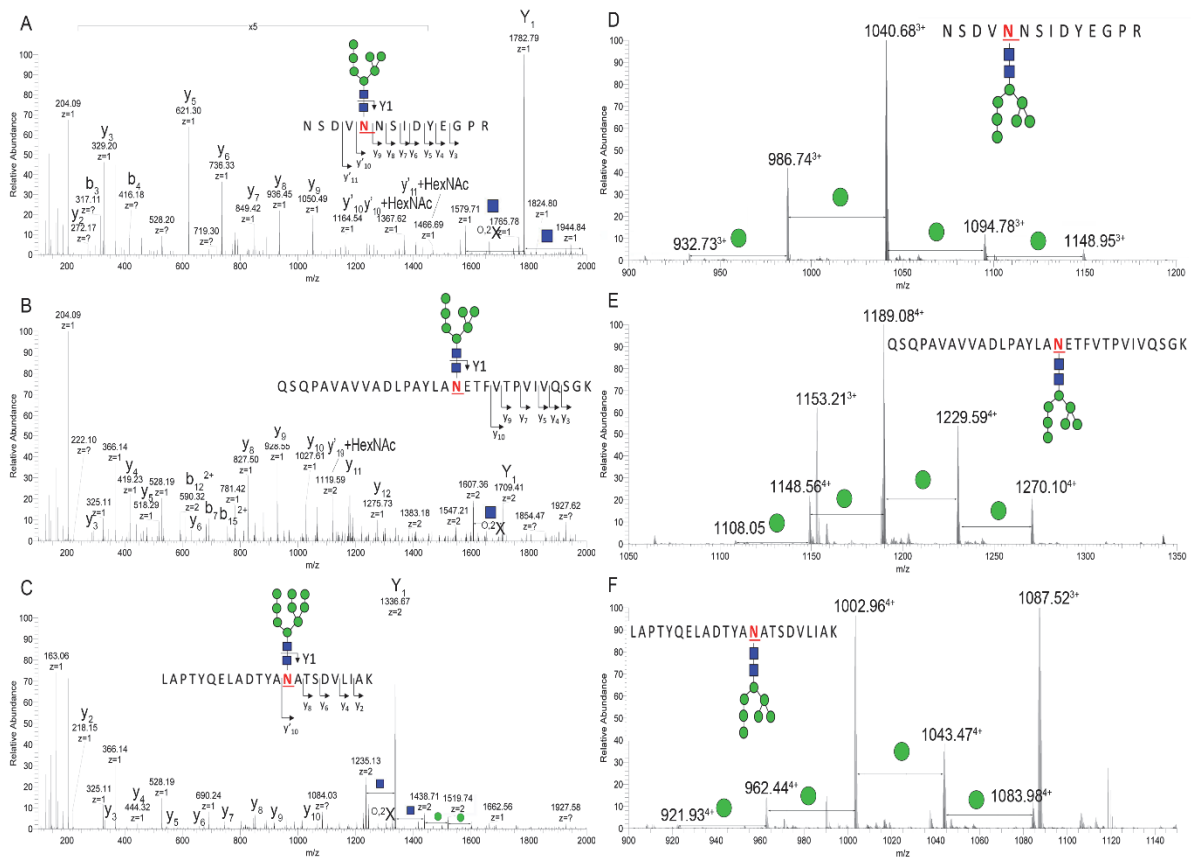
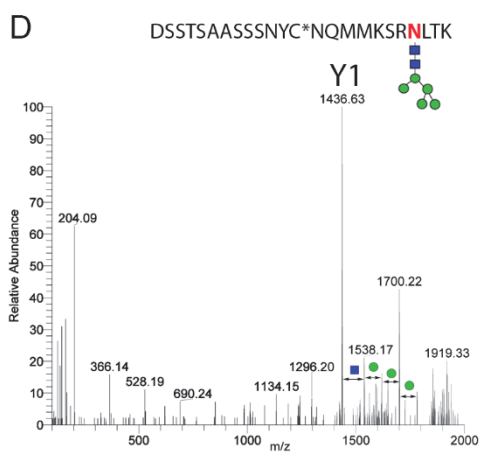
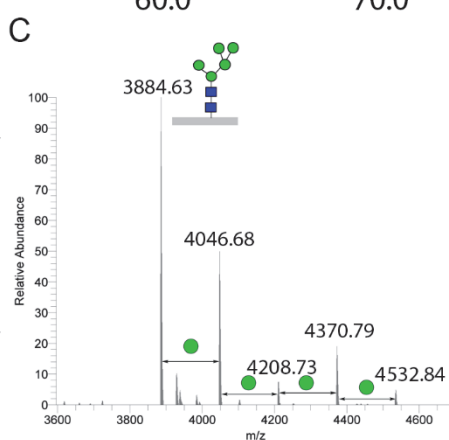
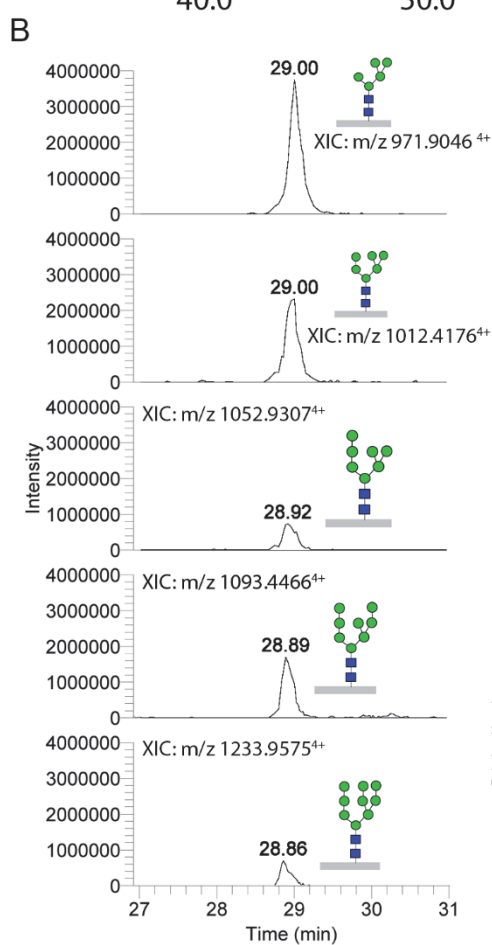
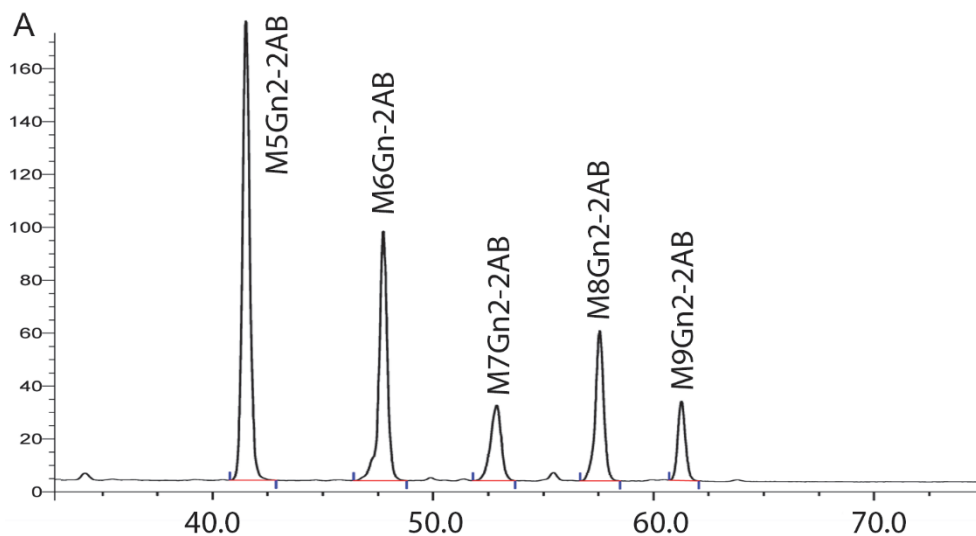


Figure S1. MS/MS and MS spectra depicting microheterogeneity of the S2, S3 and S5 N-glycosylation sites. Related to Figure 3.

(A) MS/MS spectra of m/z 1040.68 (+3) from S2, (B) m/z 1189.08 (+4) from S3 and (C) m/z 1083.98 (+4) from S5 showed respective glycopeptide identity. Corresponding MS spectra of the (D) S2, (E) S3 and (F) S5 depicted the overall glycoform heterogeneity on each site.



E

	2-AB labeled glycan		Glycopeptide					
	HPLC-FD ¹ Average	HPLC-FD SD ²	Peak height ³ Average	Peak height SD	Peak area for z=4 ⁴ Average	Peak area SD	Peak area for z=3 ⁵ Average	Peak area SD
M5Gn2	0.42	0.004	0.39	0.012	0.40	0.007	0.55	0.007
M6Gn2	0.25	0.003	0.26	0.08	0.28	0.006	0.25	0.004
M7Gn2	0.10	0.002	0.08	0.07	0.09	0.001	0.06	0.001
M8Gn2	0.16	0.001	0.19	0.02	0.18	0.005	0.11	0.008
M9Gn2	0.07	0.005	0.08	0.01	0.06	0.004	0.03	0.003

Figure S2. Evaluation of different quantification methods using glycopeptides of bovine RNaseB. Related to Figure 2.

(A) Chromatography of 2-aminobenzoic acid labeled glycan released from RNase B showed the relative amount of each glycoforms. For quantification of glycopeptides, RNase B was digested by AspN-peptidase and the peptides were analysed by nanoLC-MS/MS. Two common quantification methods were evaluated. First, extracted ion chromatography (XIC) of each glycoform was presented by its corresponding m/z (z=4), shown in **(B)**. The peak area of each precursor ion was integrated for relative quantification as described in the materials and methods section. **(C)** Peak height of deconvoluted MS spectra was used for the evaluation of relative quantification. **(D)** MS/MS spectrum showed the identity of this glycopeptide with M5Gn2. **(E)** The statistic results of different methods for relative quantification of each glycoform on the same peptide backbone.

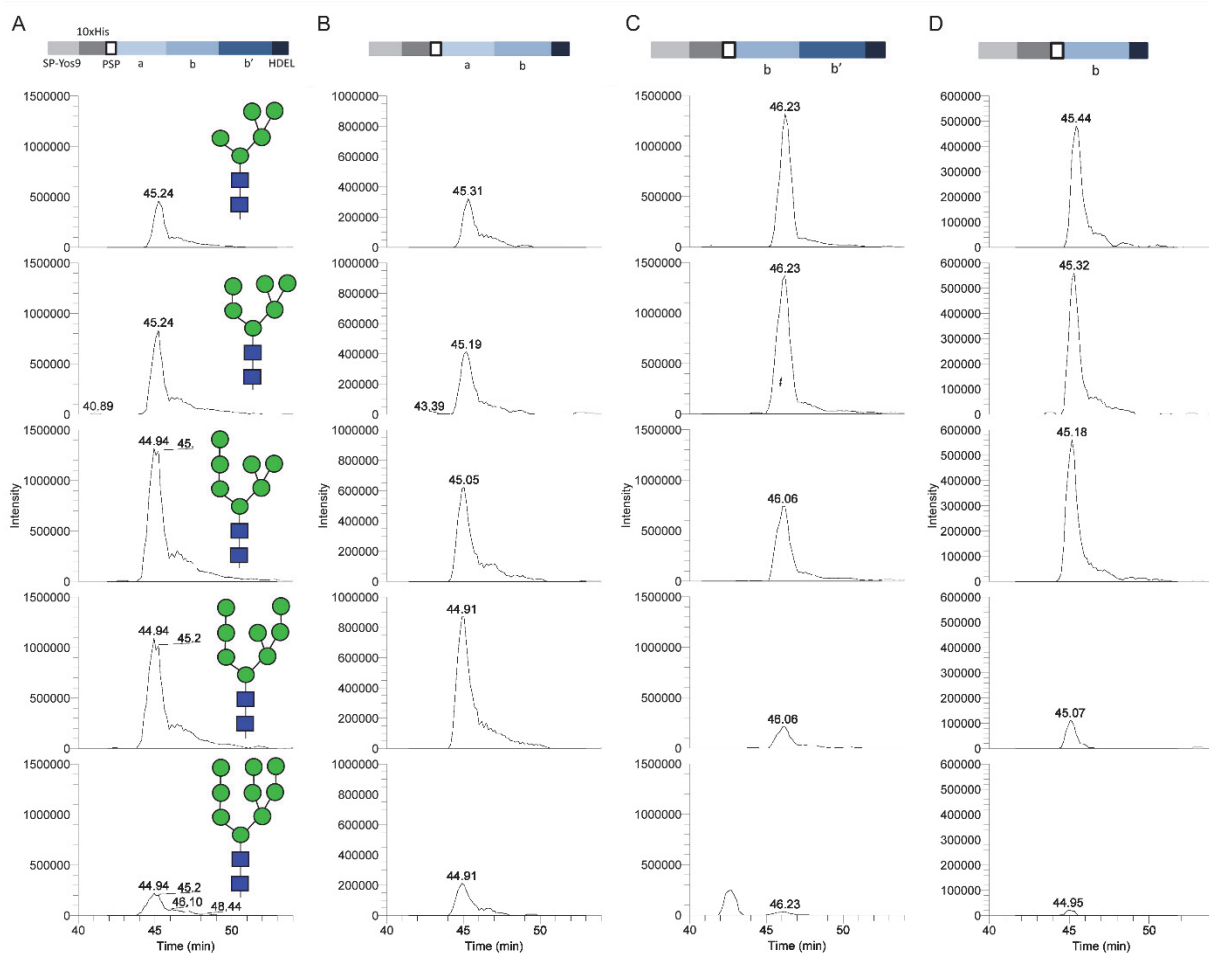


Figure S3. XIC of all glycoforms on S4 from different Pdi1p domain variants.

Related to Figure 4. Peaks corresponding to the glycoforms in the M9Gn2-M5Gn2 range after XIC on **(A)** abb'-, **(B)** ab-, **(C)** bb'- and **(D)** b-domain variant of Pdi1.

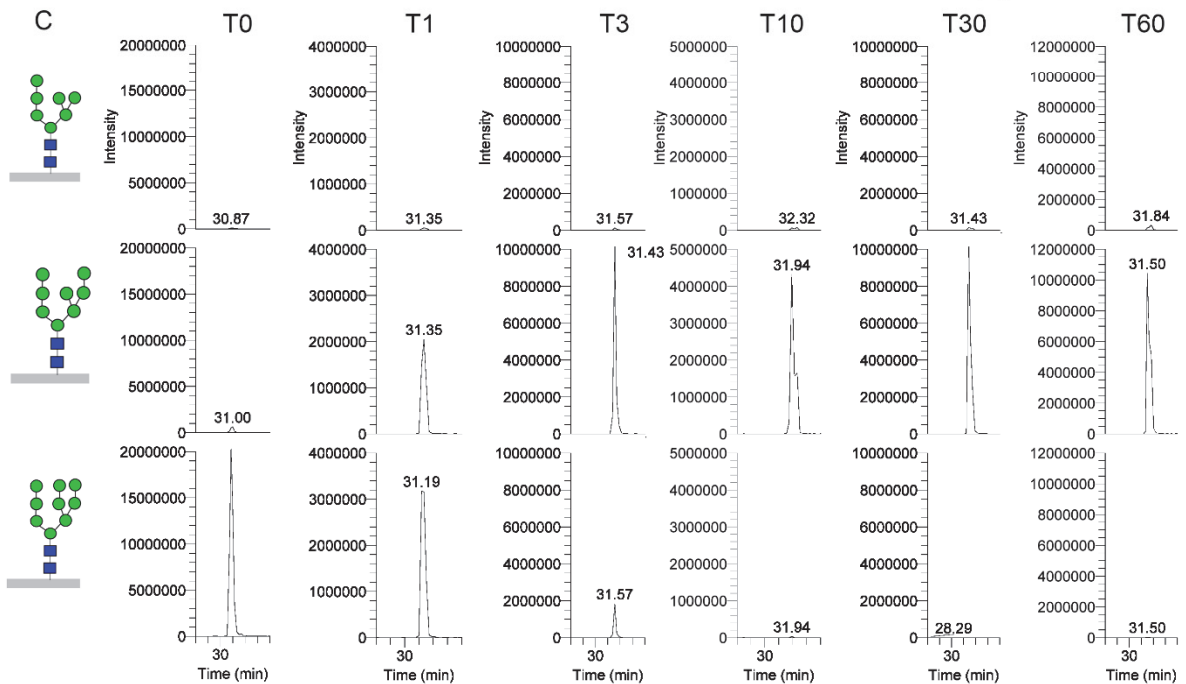
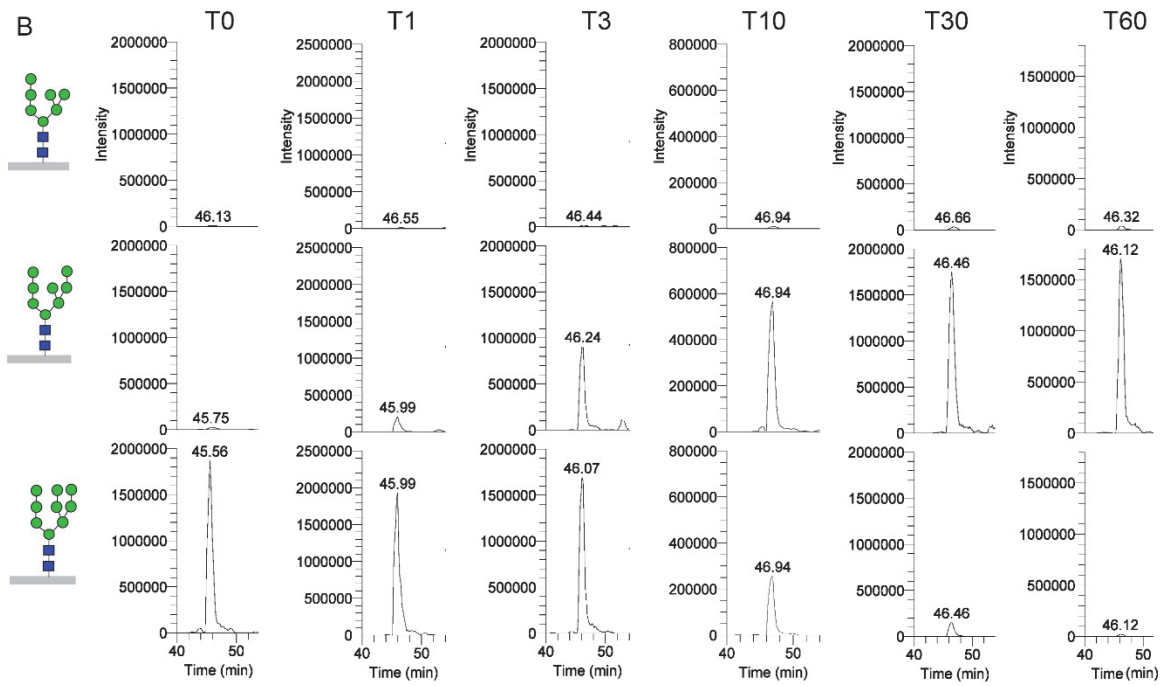
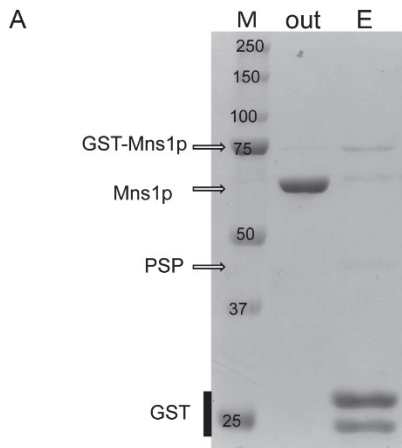


Figure S4. *In vitro* site-specific processing of Pdi1p glycans by Mns1p. Related to Figure 5.

(A) Purification of Mns1p was followed by on column PreScission protease tag cleavage. After 16 h incubation of the glutathione sepharose bound GST-tagged Pdi1p with the protease the flow-through (out) and eluted (E) fraction was analyzed by SDS-PAGE and stained with Coomassie. A time dependent increase of M8Gn2 N-glycans on **(B)** S4 and **(C)** S2 after incubation of kifunensine treated Pdi1p with Mns1p. The reaction was terminated by TCA at respective time points. After digestion, resulting glycopeptides were separated by XIC.

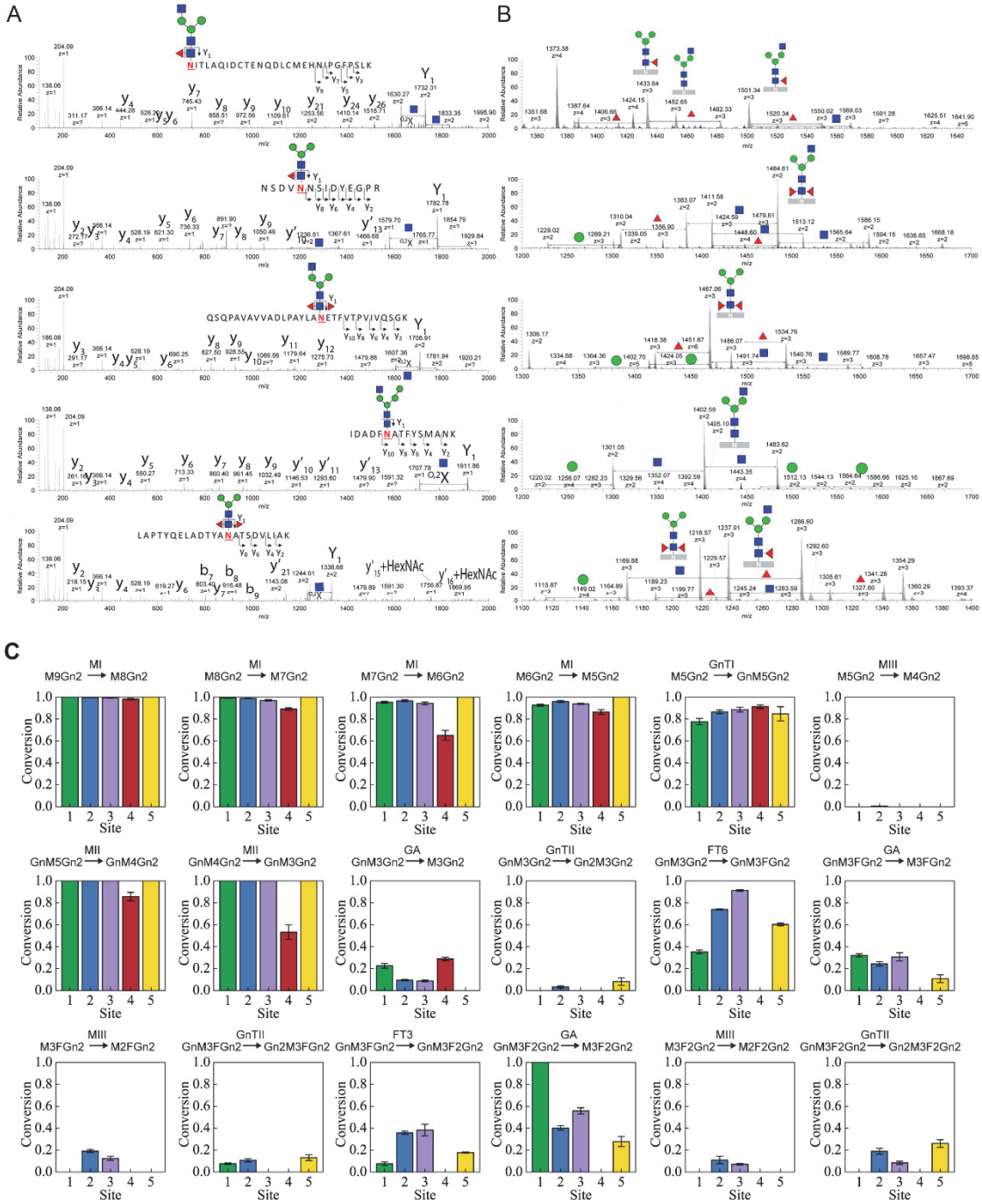


Figure S5. MS/MS spectra from N-glycans of sPDI1.

S1-S5 N- (**A**) and corresponding MS spectra (**B**) showing glycopeptide identity and heterogeneity of all sites from sPDI. (**C**) Site-specific (S1-S5) conversions for all enzymes in the glycosylation network of *T. ni*. Compared to the network depicted in Figure 6A second reaction step of MIII using M4Gn2 as a substrate to generate M3Gn2 was omitted and two additional reactions performed by the same enzyme by using M3FGn2 and M3F2Gn2 to generate M2FGn2 and M2F2Gn2 respectively, were taken into account. Site-specific enzyme conversions were obtained by glycan flux analysis using the relative abundances of each glycoform from (Figure 7B). Error bars indicate standard deviation of mean values from four experiments. Related to Figure 7.

Table S2: Quantification of glycopeptides from RNase B by mass spectrometry and by HPLC-FD

Quantification of glycopeptides from RNase B by mass spectrometry

Quantification by peak area										
observed m/z	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	interday (N)	intraday (N)	average	SD	
1295.540	3	3884.626	2668.155	1216.471	M5Gn2	3	3	0.55	0.007	
1349.557	3	4046.679	2668.155	1378.524	M6Gn2	3	3	0.25	0.004	
1403.574	3	4208.731	2668.155	1540.576	M7Gn2	3	3	0.06	0.001	
1457.593	3	4370.786	2668.155	1702.631	M8Gn2	3	3	0.11	0.008	
1511.610	3	4532.838	2668.155	1864.683	M9Gn2	3	3	0.03	0.003	

observed m/z	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	interday (N)	intraday (N)	average	SD	
971.905	4	3884.626	2668.155	1216.471	M5Gn2	3	3	0.40	0.007	
1012.418	4	4046.678	2668.155	1378.523	M6Gn2	3	3	0.28	0.006	
1052.931	4	4208.731	2668.155	1540.576	M7Gn2	3	3	0.09	0.001	
1093.447	4	4370.794	2668.155	1702.639	M8Gn2	3	3	0.18	0.005	
1133.958	4	4532.838	2668.155	1864.683	M9Gn2	3	3	0.06	0.004	

Quantification by peak height

[M+H] ⁺	glycan structure	average	SD	interday (N)	intraday (N)
3884.626	M5Gn2	0.391	0.012	3	3
4046.678	M6Gn2	0.265	0.080	3	3
4208.731	M7Gn2	0.080	0.070	3	3
4370.794	M8Gn2	0.189	0.020	3	3
4532.838	M9Gn2	0.075	0.010	3	3

Quantification of glycans from RNase B by HPLC-FD

Average	SD	interday (N)	intraday (N)
M5Gn2	0.42	0.004	3
M6Gn2	0.25	0.003	3
M7Gn2	0.10	0.002	3
M8Gn2	0.16	0.001	3
M9Gn2	0.07	0.005	3

Table S3: Quantification results for glycopeptides from ePdp by mass spectrometry

Site 1

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	Average	SD
1492.325	3	4474.981	3258.520	1216.461	M5Gn2	NITLAQIDCTENODLCMEHNIPGFP SLK	0.07	0.015
1546.338	3	4637.023	3258.520	1378.503	M6Gn2	NITLAQIDCTENODLCMEHNIPGFP SLK	0.19	0.011
1600.357	3	4799.080	3258.520	1540.560	M7Gn2	NITLAQIDCTENODLCMEHNIPGFP SLK	0.49	0.013
1654.376	3	4961.135	3258.520	1702.615	M8Gn2	NITLAQIDCTENODLCMEHNIPGFP SLK	0.12	0.004
1708.394	3	5123.190	3258.520	1864.670	M9Gn2	NITLAQIDCTENODLCMEHNIPGFP SLK	0.12	0.007

Site 2

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	Average	SD
1398.581	2	2796.170	1579.704	1216.466	M5Gn2	NSDV NS NSIDYEGPR	0.30	0.017
1479.609	2	2958.226	1579.704	1378.522	M6Gn2	NSDV NS NSIDYEGPR	0.28	0.004
1560.637	2	3120.282	1579.704	1540.578	M7Gn2	NSDV NS NSIDYEGPR	0.31	0.010
1641.666	2	3282.339	1579.704	1702.635	M8Gn2	NSDV NS NSIDYEGPR	0.07	0.003
1722.693	2	3444.393	1579.704	1864.689	M9Gn2	NSDV NS NSIDYEGPR	0.04	0.003

Site 3

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	Average	SD
1477.063	3	4429.196	3212.710	1216.486	M5Gn2	QSQPAVAVVADLPAYLANET FT VPVIVQSGK	0.19	0.035
1531.079	3	4591.244	3212.710	1378.534	M6Gn2	QSQPAVAVVADLPAYLANET FT VPVIVQSGK	0.27	0.012
1585.098	3	4753.303	3212.710	1540.593	M7Gn2	QSQPAVAVVADLPAYLANET FT VPVIVQSGK	0.29	0.024
1639.115	3	4915.353	3212.710	1702.643	M8Gn2	QSQPAVAVVADLPAYLANET FT VPVIVQSGK	0.17	0.016
1693.131	3	5077.399	3212.710	1864.689	M9Gn2	QSQPAVAVVADLPAYLANET FT VPVIVQSGK	0.08	0.012

Site 4

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	Average	SD
1462.599	2	2924.206	1707.775	1216.431	M5Gn2	IDADF NAT FYSMANK	0.02	0.005
1543.627	2	3086.262	1707.775	1378.487	M6Gn2	IDADF NAT FYSMANK	0.07	0.015
1624.655	2	3248.318	1707.775	1540.543	M7Gn2	IDADF NAT FYSMANK	0.27	0.004
1705.683	2	3410.373	1707.775	1702.598	M8Gn2	IDADF NAT FYSMANK	0.44	0.013
1786.707	2	3572.422	1707.775	1864.647	M9Gn2	IDADF NAT FYSMANK	0.20	0.026

Site 5

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	Average	SD
1228.900	3	3684.701	2468.250	1216.451	M5Gn2	LAPTYQELADTYANATSDVLI AK	0.13	0.023
1282.919	3	3846.758	2468.250	1378.508	M6Gn2	LAPTYQELADTYANATSDVLI AK	0.26	0.008
1336.940	3	4008.819	2468.250	1540.569	M7Gn2	LAPTYQELADTYANATSDVLI AK	0.34	0.014
1390.956	3	4170.868	2468.250	1702.618	M8Gn2	LAPTYQELADTYANATSDVLI AK	0.19	0.011
1444.973	3	4332.918	2468.250	1864.668	M9Gn2	LAPTYQELADTYANATSDVLI AK	0.09	0.008

Table S4: Quantification results for glycopeptides from different domains of Pdiip by mass spectrometry

abb' domain

Site 1

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	Average	SD
1492.325	3	4474.981	3258.520	1216.461	M5Gn2	NITLAQIDCTENQDQLCMEHINIPGPPSLK	0.30	0.061
1546.338	3	4637.023	3258.520	1378.503	M6Gn2	NITLAQIDCTENQDQLCMEHINIPGPPSLK	0.22	0.010
1600.357	3	4799.080	3258.520	1540.560	M7Gn2	NITLAQIDCTENQDQLCMEHINIPGPPSLK	0.26	0.032
1654.376	3	4961.135	3258.520	1702.615	M8Gn2	NITLAQIDCTENQDQLCMEHINIPGPPSLK	0.11	0.010
1708.394	3	5123.190	3258.520	1864.670	M9Gn2	NITLAQIDCTENQDQLCMEHINIPGPPSLK	0.11	0.026

Site 2

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	Average	SD
1398.581	2	2796.170	1579.704	1216.466	M5Gn2	NSDVNNSIDYEGPR	0.65	0.065
1479.609	2	2958.226	1579.704	1378.522	M6Gn2	NSDVNNSIDYEGPR	0.19	0.028
1560.637	2	3120.282	1579.704	1540.578	M7Gn2	NSDVNNSIDYEGPR	0.11	0.023
1641.666	2	3282.339	1579.704	1702.635	M8Gn2	NSDVNNSIDYEGPR	0.04	0.013
1722.693	2	3444.393	1579.704	1864.689	M9Gn2	NSDVNNSIDYEGPR	0.01	0.005

Site 3

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	Average	SD
1477.063	3	4429.196	3212.710	1216.486	M5Gn2	QSQPAVVAVVDLPAYLANETFTVPVIVQSGK	0.42	0.093
1531.079	3	4591.244	3212.710	1378.534	M6Gn2	QSQPAVVAVVDLPAYLANETFTVPVIVQSGK	0.20	0.020
1585.098	3	4753.303	3212.710	1540.593	M7Gn2	QSQPAVVAVVDLPAYLANETFTVPVIVQSGK	0.15	0.032
1639.115	3	4915.353	3212.710	1702.643	M8Gn2	QSQPAVVAVVDLPAYLANETFTVPVIVQSGK	0.16	0.026
1693.131	3	5077.399	3212.710	1864.689	M9Gn2	QSQPAVVAVVDLPAYLANETFTVPVIVQSGK	0.07	0.024

Site 4

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	Average	SD
1462.599	2	2924.206	1707.775	1216.431	M5Gn2	IDADF N ATFYSMANK	0.14	0.029
1543.627	2	3086.262	1707.775	1378.487	M6Gn2	IDADF N ATFYSMANK	0.20	0.012
1624.655	2	3248.318	1707.775	1540.543	M7Gn2	IDADF N ATFYSMANK	0.32	0.032
1705.683	2	3410.373	1707.775	1702.598	M8Gn2	IDADF N ATFYSMANK	0.29	0.018
1786.707	2	3572.422	1707.775	1864.647	M9Gn2	IDADF N ATFYSMANK	0.04	0.011

ab domain

Average	SD
0.29	0.055
0.20	0.012
0.25	0.023
0.12	0.014
0.14	0.019

Average	SD
0.59	0.061
0.16	0.026
0.13	0.027
0.09	0.014
0.03	0.006

Average	SD
0.44	0.084
0.20	0.024
0.14	0.018
0.17	0.034
0.05	0.031

Average	SD
0.14	0.024
0.19	0.016
0.25	0.019
0.36	0.013
0.06	0.011

bb' domain

Average	SD
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.

Average	SD
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.

Average	SD
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.

Average	SD
0.28	0.035
0.33	0.028
0.28	0.021
0.10	0.013
0.01	0.004

b domain

Average	SD
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.

Average	SD
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.

Average	SD
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.
n.d.	n.d.

Average	SD
0.32	0.033
0.30	0.023
0.31	0.018
0.06	0.015
0.00	0.002

Table S5: Quantification results for glycopeptides from Pdip treated with kifunensine first and mannosidase by mass spectrometry

Site 1 (1st)										Site 1 (2nd)					Site 1 (3rd)										
ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	T0	T1	T3	T10	T30	T60	T0	T1	T3	T10	T30	T60	T0	T1	T3	T10	T30	T60	
1492.325	3	4474.981	3258.520	1216.461	M5Gn2	NITLAQIDCTENQDLCMEHNIPGFPSLK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1546.338	3	4637.023	3258.520	1378.503	M6Gn2	NITLAQIDCTENQDLCMEHNIPGFPSLK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1600.357	3	4799.080	3258.520	1540.560	M7Gn2	NITLAQIDCTENQDLCMEHNIPGFPSLK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1654.376	3	4961.135	3258.520	1702.615	M8Gn2	NITLAQIDCTENQDLCMEHNIPGFPSLK	0.01	0.15	0.56	0.93	0.93	0.95	0.00	0.21	0.51	0.91	0.91	0.91	0.90	0.00	0.22	0.55	0.94	0.98	0.95
1708.394	3	5123.190	3258.520	1864.670	M9Gn2	NITLAQIDCTENQDLCMEHNIPGFPSLK	0.99	0.85	0.44	0.07	0.07	0.04	1.00	0.79	0.49	0.09	0.09	0.09	0.09	1.00	0.78	0.45	0.06	0.00	0.00

Site 2 (1st)										Site 2 (2nd)					Site 2 (3rd)										
ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	T0	T1	T3	T10	T30	T60	T0	T1	T3	T10	T30	T60	T0	T1	T3	T10	T30	T60	
1398.581	2	2796.170	1579.704	1216.466	M5Gn2	NSDVNSIDYEGPR	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.01
1479.609	2	2958.226	1579.704	1378.522	M6Gn2	NSDVNSIDYEGPR	0.01	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.01
1560.637	2	3120.282	1579.704	1540.578	M7Gn2	NSDVNSIDYEGPR	0.01	0.01	0.01	0.01	0.01	0.02	0.00	0.01	0.02	0.02	0.02	0.03	0.03	0.00	0.01	0.01	0.02	0.03	0.03
1641.666	2	3282.339	1579.704	1702.635	M8Gn2	NSDVNSIDYEGPR	0.03	0.38	0.83	0.83	0.83	0.96	0.03	0.42	0.79	0.96	0.96	0.96	0.95	0.03	0.49	0.89	0.96	0.96	0.95
1722.693	2	3444.393	1579.704	1864.689	M9Gn2	NSDVNSIDYEGPR	0.96	0.61	0.14	0.14	0.00	0.00	0.96	0.57	0.19	0.00	0.00	0.00	0.00	0.96	0.50	0.10	0.00	0.00	0.00

Site 3 (1st)										Site 3 (2nd)					Site 3 (3rd)										
ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	T0	T1	T3	T10	T30	T60	T0	T1	T3	T10	T30	T60	T0	T1	T3	T10	T30	T60	
1477.063	3	4429.196	3212.710	1216.486	M5Gn2	GSQPAAVVADLPAYLANETFPVIVQSGK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1531.079	3	4591.244	3212.710	1378.534	M6Gn2	GSQPAAVVADLPAYLANETFPVIVQSGK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02
1585.098	3	4753.303	3212.710	1540.593	M7Gn2	GSQPAAVVADLPAYLANETFPVIVQSGK	0.00	0.00	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.02
1639.115	3	4915.353	3212.710	1702.643	M8Gn2	GSQPAAVVADLPAYLANETFPVIVQSGK	0.02	0.28	0.78	0.96	0.96	0.98	0.03	0.35	0.74	0.98	0.98	0.97	0.03	0.39	0.86	0.97	0.98	0.95	0.95
1693.131	3	5077.399	3212.710	1864.689	M9Gn2	GSQPAAVVADLPAYLANETFPVIVQSGK	0.98	0.72	0.21	0.03	0.00	0.00	0.96	0.64	0.24	0.00	0.00	0.00	0.96	0.60	0.12	0.02	0.00	0.00	0.00

Site 4 (1st)										Site 4 (2nd)					Site 4 (3rd)										
ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	T0	T1	T3	T10	T30	T60	T0	T1	T3	T10	T30	T60	T0	T1	T3	T10	T30	T60	
1462.599	2	2924.206	1707.775	1216.431	M5Gn2	IDADFNAIFYSMANK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1543.627	3	3086.262	1707.775	1378.487	M6Gn2	IDADFNAIFYSMANK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1624.655	2	3248.318	1707.775	1540.543	M7Gn2	IDADFNAIFYSMANK	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.00	0.01	0.01	0.02	0.02
1705.683	2	3410.373	1707.775	1702.598	M8Gn2	IDADFNAIFYSMANK	0.01	0.09	0.35	0.66	0.91	0.96	0.01	0.13	0.33	0.66	0.89	0.95	0.01	0.13	0.42	0.71	0.90	0.95	
1786.707	2	3572.422	1707.775	1864.647	M9Gn2	IDADFNAIFYSMANK	0.98	0.90	0.64	0.33	0.07	0.01	0.97	0.86	0.66	0.31	0.09	0.02	0.98	0.86	0.56	0.27	0.07	0.02	

Site 5 (1st)										Site 5 (2nd)					Site 5 (3rd)										
ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	T0	T1	T3	T10	T30	T60	T0	T1	T3	T10	T30	T60	T0	T1	T3	T10	T30	T60	
1228.900	3	3684.701	2468.250	1216.451	M5Gn2	LAPTYQELADTYANATSDVLIAK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1282.919	3	3846.758	2468.250	1378.508	M6Gn2	LAPTYQELADTYANATSDVLIAK	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01
1336.940	3	4008.819	2468.250	1540.569	M7Gn2	LAPTYQELADTYANATSDVLIAK	0.00	0.01	0.01	0.03	0.03	0.04	0.01	0.01	0.02	0.02	0.03	0.03	0.03	0.01	0.01	0.02	0.02	0.03	0.03
1390.956	3	4170.868	2468.250	1702.618	M8Gn2	LAPTYQELADTYANATSDVLIAK	0.02	0.22	0.62	0.94	0.95	0.93	0.03	0.24	0.59	0.93	0.95	0.95	0.03	0.26	0.65	0.94	0.96	0.96	
1444.973	3	4332.918	2468.250	1864.668	M9Gn2	LAPTYQELADTYANATSDVLIAK	0.97	0.77	0.36	0.03	0.01	0.01	0.96	0.75	0.38	0.04	0.01	0.01	0.96	0.73	0.32	0.03	0.01	0.01	

Table S6: Quantification results for glycopeptides from sPdp B by mass spectrometry

Site 1

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan stucture	peptide sequence	Average	SD
1384.28	3	4150.84	3258.52	892.32	(HexNAc2Hex3)	NI ¹⁵ TLAQIDCTENQDLCMEHNIPGFPSLK	0.15	0.012
1432.97	3	4296.91	3258.52	1038.39	(HexNAc2Hex3)Fuc	NI ¹⁵ TLAQIDCTENQDLCMEHNIPGFPSLK	0.08	0.005
1451.98	3	4353.94	3258.52	1095.42	(HexNAc2Hex3)HexNAc	NI ¹⁵ TLAQIDCTENQDLCMEHNIPGFPSLK	0.29	0.021
1481.65	3	4442.95	3258.52	1184.43	(HexNAc2Hex3)Fuc2	NI ¹⁵ TLAQIDCTENQDLCMEHNIPGFPSLK	0.02	0.004
1492.32	3	4474.98	3258.52	1216.46	(HexNAc2Hex3)Hex2	NI ¹⁵ TLAQIDCTENQDLCMEHNIPGFPSLK	0.20	0.028
1500.67	3	4500.01	3258.52	1241.49	(HexNAc2Hex3)FucHexNAc	NI ¹⁵ TLAQIDCTENQDLCMEHNIPGFPSLK	0.13	0.014
1546.34	3	4637.02	3258.52	1378.50	(HexNAc2Hex3)Hex3	NI ¹⁵ TLAQIDCTENQDLCMEHNIPGFPSLK	0.07	0.008
1568.36	3	4703.08	3258.52	1444.56	(HexNAc2Hex3)FucHexNAc2	NI ¹⁵ TLAQIDCTENQDLCMEHNIPGFPSLK	0.02	0.001
1600.36	3	4799.08	3258.52	1540.56	(HexNAc2Hex3)Hex4	NI ¹⁵ TLAQIDCTENQDLCMEHNIPGFPSLK	0.05	0.009
1654.38	3	4961.14	3258.52	1702.62	(HexNAc2Hex3)Hex5	NI ¹⁵ TLAQIDCTENQDLCMEHNIPGFPSLK	0.01	0.004

Site 2

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan stucture	peptide sequence	Average	SD
1228.52	2	2456.03	1579.70	876.33	HexNAc2Hex2Fuc	NSDV ¹⁵ NNSIDYEGPR	0.03	0.004
1236.51	2	2472.03	1579.70	892.32	(HexNAc2Hex3)	NSDV ¹⁵ NNSIDYEGPR	0.07	0.005
1301.55	2	2602.09	1579.70	1022.39	HexNAc2Hex2Fuc2	NSDV ¹⁵ NNSIDYEGPR	0.01	0.003
1309.55	2	2618.09	1579.70	1038.39	(HexNAc2Hex3)Fuc	NSDV ¹⁵ NNSIDYEGPR	0.11	0.007
1317.54	2	2634.09	1579.70	1054.38	(HexNAc2Hex3)Hex	NSDV ¹⁵ NNSIDYEGPR	0.01	0.001
1338.05	2	2675.10	1579.70	1095.40	(HexNAc2Hex3)HexNAc	NSDV ¹⁵ NNSIDYEGPR	0.10	0.005
1382.57	2	2764.15	1579.70	1184.44	(HexNAc2Hex3)Fuc2	NSDV ¹⁵ NNSIDYEGPR	0.07	0.006
1390.56	2	2780.13	1579.70	1200.42	(HexNAc2Hex3)HexFuc	NSDV ¹⁵ NNSIDYEGPR	0.00	0.002
1398.57	2	2796.15	1579.70	1216.45	(HexNAc2Hex3)Hex2	NSDV ¹⁵ NNSIDYEGPR	0.11	0.014
1411.09	2	2821.17	1579.70	1241.47	(HexNAc2Hex3)FucHexNAc	NSDV ¹⁵ NNSIDYEGPR	0.17	0.005
1439.59	2	2878.18	1579.70	1298.48	(HexNAc2Hex3)HexNAc	NSDV ¹⁵ NNSIDYEGPR	0.03	0.009
1479.59	2	2958.19	1579.70	1378.48	(HexNAc2Hex3)Hex3	NSDV ¹⁵ NNSIDYEGPR	0.04	0.009
1484.11	2	2967.23	1579.70	1387.52	(HexNAc2Hex3)Fuc2HexNAc	NSDV ¹⁵ NNSIDYEGPR	0.08	0.005
1492.11	2	2983.22	1579.70	1403.51	(HexNAc2Hex3)HexHexNAcFuc	NSDV ¹⁵ NNSIDYEGPR	0.02	0.000
1512.63	2	3024.25	1579.70	1444.55	(HexNAc2Hex3)FucHexNAc2	NSDV ¹⁵ NNSIDYEGPR	0.06	0.011
1560.62	2	3120.24	1579.70	1540.54	(HexNAc2Hex3)Hex4	NSDV ¹⁵ NNSIDYEGPR	0.03	0.009
1585.65	2	3170.30	1579.70	1590.60	(HexNAc2Hex3)Fuc2HexNAc2	NSDV ¹⁵ NNSIDYEGPR	0.04	0.007
1641.65	2	3282.30	1579.70	1702.59	(HexNAc2Hex3)Hex5	NSDV ¹⁵ NNSIDYEGPR	0.01	0.002
1722.68	2	3444.37	1579.70	1864.66	(HexNAc2Hex3)Hex6	NSDV ¹⁵ NNSIDYEGPR	0.00	0.001

Site 3

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan stucture	peptide sequence	Average	SD
1363.69	3	4089.07	3212.71	876.36	HexNAc2Hex2Fuc	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.03	0.006
1369.02	3	4105.07	3212.71	892.36	HexNAc2Hex3	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.07	0.009
1412.38	3	4235.13	3212.71	1022.42	HexNAc2Hex2Fuc2	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.01	0.002
1417.71	3	4251.13	3212.71	1038.42	(HexNAc2Hex3)Fuc	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.19	0.029
1466.40	3	4397.19	3212.71	1184.48	(HexNAc2Hex3)Fuc2	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.14	0.015
1477.06	3	4429.18	3212.71	1216.47	(HexNAc2Hex3)Hex2	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.10	0.018
1485.40	3	4454.20	3212.71	1241.49	(HexNAc2Hex3)FucHexNAc	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.21	0.018
1531.07	3	4591.22	3212.71	1378.51	(HexNAc2Hex3)Hex3	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.06	0.004
1534.09	3	4600.26	3212.71	1387.55	(HexNAc2Hex3)Fuc2HexNAc	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.09	0.014
1585.09	3	4753.28	3212.71	1540.57	(HexNAc2Hex3)Hex4	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.06	0.012
1601.78	3	4803.34	3212.71	1590.63	(HexNAc2Hex3)Fuc2HexNAc2	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.02	0.003
1639.11	3	4915.34	3212.71	1702.63	(HexNAc2Hex3)Hex5	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.03	0.007
1693.13	3	5077.39	3212.71	1864.68	(HexNAc2Hex3)Hex6	QSQPAVAVVADLPAYLAN ¹⁵ ETVTPVIVQSGK	0.01	0.003

Site 4

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan stucture	peptide sequence	Average	SD
1300.55	2	2600.10	1707.78	892.33	(HexNAc2Hex3)	IDADF ¹⁵ NATFYSMANK	0.06	0.017
1373.58	2	2746.15	1707.78	1038.38	(HexNAc2Hex3)Fuc	IDADF ¹⁵ NATFYSMANK	0.00	0.000
1402.09	2	2803.18	1707.78	1095.41	(HexNAc2Hex3)HexNAc	IDADF ¹⁵ NATFYSMANK	0.14	0.033
1462.60	2	2924.20	1707.78	1216.43	(HexNAc2Hex3)Hex2	IDADF ¹⁵ NATFYSMANK	0.04	0.007
1475.12	2	2949.23	1707.78	1241.46	(HexNAc2Hex3)HexNAcFuc	IDADF ¹⁵ NATFYSMANK	0.00	0.000
1483.12	2	2965.23	1707.78	1257.46	(HexNAc2Hex3)HexNAc1Hex1	IDADF ¹⁵ NATFYSMANK	0.17	0.016
1543.62	2	3086.24	1707.78	1378.47	(HexNAc2Hex3)Hex3	IDADF ¹⁵ NATFYSMANK	0.08	0.013
1564.14	2	3127.28	1707.78	1419.51	(HexNAc2Hex3)HexNAcHex2	IDADF ¹⁵ NATFYSMANK	0.06	0.014
1624.65	2	3248.31	1707.78	1540.53	(HexNAc2Hex3)Hex4	IDADF ¹⁵ NATFYSMANK	0.30	0.032
1645.17	2	3289.33	1707.78	1581.56	(HexNAc2Hex3)HexNAcHex3	IDADF ¹⁵ NATFYSMANK	0.02	0.006
1705.68	2	3410.36	1707.78	1702.58	(HexNAc2Hex3)Hex5	IDADF ¹⁵ NATFYSMANK	0.10	0.011
1786.71	2	3572.41	1707.78	1864.64	(HexNAc2Hex3)Hex6	IDADF ¹⁵ NATFYSMANK	0.02	0.006

Site 5

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan stucture	peptide sequence	Average	SD
1169.55	3	3506.66	2468.25	1038.41	(HexNAc2Hex3)Fuc	LAPTYQELADTYA ¹⁵ NATSDVLIK	0.05	0.020
1188.56	3	3563.68	2468.25	1095.43	(HexNAc2Hex3)HexNAc	LAPTYQELADTYA ¹⁵ NATSDVLIK	0.27	0.052
1218.24	3	3652.71	2468.25	1184.46	(HexNAc2Hex3)Fuc2	LAPTYQELADTYA ¹⁵ NATSDVLIK	0.02	0.004
1228.90	3	3684.70	2468.25	1216.45	(HexNAc2Hex3)Hex2	LAPTYQELADTYA ¹⁵ NATSDVLIK	0.15	0.064
1237.25	3	3709.74	2468.25	1241.49	(HexNAc2Hex3)FucHexNAc	LAPTYQELADTYA ¹⁵ NATSDVLIK	0.30	0.027
1256.25	3	3766.75	2468.25	1298.50	(HexNAc2Hex3)HexNAc2	LAPTYQELADTYA ¹⁵ NATSDVLIK	0.07	0.023
1285.93	3	3855.79	2468.25	1387.54	(HexNAc2Hex3)Fuc2HexNAc	LAPTYQELADTYA ¹⁵ NATSDVLIK	0.04	0.010
1304.94	3	3912.81	2468.25	1444.56	(HexNAc2Hex3)FucHexNAc2	LAPTYQELADTYA ¹⁵ NATSDVLIK	0.07	0.009
1339.95	3	4017.84	2468.25	1549.59	(HexNAc2Hex3)Fuc2HexNAcHex	LAPTYQELADTYA ¹⁵ NATSDVLIK	0.00	0.001
1353.62	3	4058.87	2468.25	1590.62	(HexNAc2Hex3)Fuc2HexNAc2	LAPTYQELADTYA ¹⁵ NATSDVLIK	0.02	0.002

Table S7: Quantification results for glycopeptides from sDip B by mass spectrometry

Site 1

ions used for quantification	z	[M+H] ⁺	peptide [M+H] ⁺	glycan	glycan structure	peptide sequence	EV		Gal	
							Average	SD	Average	SD
1384.28	3	4150.84	3258.52	892.32	(HexNac2Hex3)	NTLAQIDCTENQDLCEHNIPGFPSLK	0.1598	0.0033	0.1794	0.0043
1432.97	3	4296.91	3258.52	1038.39	(HexNac2Hex3)Fuc	NTLAQIDCTENQDLCEHNIPGFPSLK	0.1180	0.0017	0.1138	0.0024
1451.98	3	4353.94	3258.52	1095.42	(HexNac2Hex3)HexNac	NTLAQIDCTENQDLCEHNIPGFPSLK	0.2880	0.0107	0.3236	0.0057
1481.65	3	4442.95	3258.52	1184.43	(HexNac2Hex3)Fuc2	NTLAQIDCTENQDLCEHNIPGFPSLK	0.0380	0.0029	0.0108	0.0007
1492.32	3	4474.98	3258.52	1216.46	(HexNac2Hex3)Hex2	NTLAQIDCTENQDLCEHNIPGFPSLK	0.1230	0.0031	0.1276	0.0064
1500.67	3	4500.01	3258.52	1241.49	(HexNac2Hex3)FucHexNac	NTLAQIDCTENQDLCEHNIPGFPSLK	0.1595	0.0024	0.1177	0.0008
1506.00	3	4515.99	3258.52	1257.47	(HexNac2Hex3)HexHexNac	NTLAQIDCTENQDLCEHNIPGFPSLK	0.0214	0.0020	0.0260	0.0014
1546.34	3	4637.02	3258.52	1378.50	(HexNac2Hex3)Hex3	NTLAQIDCTENQDLCEHNIPGFPSLK	0.0376	0.0014	0.0386	0.0017
1560.02	3	4678.05	3258.52	1419.53	(HexNac2Hex3)Hex2HexNac	NTLAQIDCTENQDLCEHNIPGFPSLK	0.0030	0.0013	0.0064	0.0005
1568.36	3	4703.08	3258.52	1444.56	(HexNac2Hex3)FucHexNac2	NTLAQIDCTENQDLCEHNIPGFPSLK	0.0243	0.0052	0.0202	0.0040
1600.36	3	4799.08	3258.52	1540.56	(HexNac2Hex3)Hex4	NTLAQIDCTENQDLCEHNIPGFPSLK	0.0241	0.0032	0.0327	0.0018
1654.38	3	4961.14	3258.52	1702.62	(HexNac2Hex3)Hex5	NTLAQIDCTENQDLCEHNIPGFPSLK	0.0033	0.0009	0.0032	0.0019

Site 2

1228.52	2	2456.03	1579.704	876.33	HexNac2Hex2Fuc	NSDVENNSIDYEGPR	0.0422	0.0004	0.0270	0.0005
1236.51	2	2472.03	1579.704	892.32	(HexNac2Hex3)	NSDVENNSIDYEGPR	0.0504	0.0008	0.0643	0.0014
1301.55	2	2602.09	1579.704	1022.39	HexNac2Hex2Fuc2	NSDVENNSIDYEGPR	0.0316	0.0012	0.0122	0.0004
1309.55	2	2618.09	1579.704	1038.39	(HexNac2Hex3)Fuc	NSDVENNSIDYEGPR	0.1281	0.0037	0.1199	0.0021
1317.54	2	2634.09	1579.704	1054.38	(HexNac2Hex3)Hex	NSDVENNSIDYEGPR	0.0020	0.0002	0.0030	0.0001
1338.05	2	2675.10	1579.704	1095.40	(HexNac2Hex3)HexNac	NSDVENNSIDYEGPR	0.0893	0.0010	0.1251	0.0013
1382.57	2	2764.15	1579.704	1184.44	(HexNac2Hex3)Fuc2	NSDVENNSIDYEGPR	0.1291	0.0029	0.0676	0.0003
1390.56	2	2780.13	1579.704	1200.42	(HexNac2Hex3)HexFuc	NSDVENNSIDYEGPR	0.0000	0.0000	0.0387	0.0010
1398.57	2	2796.15	1579.704	1216.45	(HexNac2Hex3)Hex2	NSDVENNSIDYEGPR	0.0369	0.0010	0.0608	0.0013
1411.09	2	2821.17	1579.704	1241.47	(HexNac2Hex3)FucHexNac	NSDVENNSIDYEGPR	0.2090	0.0039	0.1414	0.0054
1419.08	2	2837.16	1579.704	1257.46	(HexNac2Hex3)HexHexNac	NSDVENNSIDYEGPR	0.0090	0.0004	0.0180	0.0005
1439.59	2	2878.18	1579.704	1298.48	(HexNac2Hex3)HexNac2	NSDVENNSIDYEGPR	0.0106	0.0007	0.0196	0.0007
1463.61	2	2926.22	1579.704	1346.52	(HexNac2Hex3)Fuc2Hex	NSDVENNSIDYEGPR	0.0000	0.0000	0.0595	0.0012
1479.59	2	2958.19	1579.704	1378.48	(HexNac2Hex3)Hex3	NSDVENNSIDYEGPR	0.0087	0.0016	0.0148	0.0003
1484.11	2	2967.23	1579.704	1387.52	(HexNac2Hex3)Fuc2HexNac	NSDVENNSIDYEGPR	0.1456	0.0012	0.0746	0.0004
1492.11	2	2983.22	1579.704	1403.51	(HexNac2Hex3)HexHexNacFuc	NSDVENNSIDYEGPR	0.0107	0.0008	0.0783	0.0019
1512.63	2	3024.25	1579.704	1444.55	(HexNac2Hex3)FucHexNac2	NSDVENNSIDYEGPR	0.0487	0.0023	0.0407	0.0010
1560.62	2	3120.24	1579.704	1540.54	(HexNac2Hex3)Hex4	NSDVENNSIDYEGPR	0.0061	0.0009	0.0097	0.0008
1585.65	2	3170.30	1579.704	1590.60	(HexNac2Hex3)Fuc2HexNac2	NSDVENNSIDYEGPR	0.0400	0.0015	0.0212	0.0007
1641.65	2	3282.30	1579.704	1702.59	(HexNac2Hex3)Hex5	NSDVENNSIDYEGPR	0.0016	0.0003	0.0029	0.0003
1722.68	2	3444.37	1579.704	1864.66	(HexNac2Hex3)Hex6	NSDVENNSIDYEGPR	0.0005	0.0002	0.0008	0.0002

Site 3

1363.69	3	4089.07	3212.71	876.36	HexNac2Hex2Fuc	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0338	0.0010	0.0224	0.0022
1369.02	3	4105.07	3212.71	892.36	HexNac2Hex3	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0462	0.0010	0.0651	0.0024
1412.38	3	4235.13	3212.71	1022.42	HexNac2Hex2Fuc2	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0225	0.0027	0.0103	0.0049
1417.71	3	4251.13	3212.71	1038.42	(HexNac2Hex3)Fuc	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.1847	0.0034	0.1513	0.0126
1466.40	3	4397.19	3212.71	1184.48	(HexNac2Hex3)Fuc2	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.2000	0.0103	0.1157	0.0040
1471.73	3	4413.19	3212.71	1200.48	(HexNac2Hex3)FucHex	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0000	0.0000	0.0904	0.0033
1477.06	3	4429.18	3212.71	1216.47	(HexNac2Hex3)Hex2	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0618	0.0021	0.0871	0.0044
1520.42	3	4559.25	3212.71	1346.54	(HexNac2Hex3)Fuc2Hex	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.2481	0.0084	0.1565	0.0077
1485.40	3	4454.20	3212.71	1241.49	(HexNac2Hex3)FucHexNac	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0242	0.0026	0.0384	0.0025
1531.07	3	4591.22	3212.71	1378.51	(HexNac2Hex3)Hex3	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0000	0.0000	0.1503	0.0095
1534.09	3	4600.26	3212.71	1387.55	(HexNac2Hex3)Fuc2HexNac	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.1248	0.0064	0.0663	0.0020
1585.09	3	4753.28	3212.71	1540.57	(HexNac2Hex3)Hex4	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0186	0.0013	0.0295	0.0021
1601.78	3	4803.34	3212.71	1590.63	(HexNac2Hex3)Fuc2HexNac2	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0263	0.0019	0.0075	0.0017
1639.11	3	4915.34	3212.71	1702.63	(HexNac2Hex3)Hex5	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0078	0.0018	0.0066	0.0013
1693.13	3	5077.39	3212.71	1864.68	(HexNac2Hex3)Hex6	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	0.0012	0.0014	0.0026	0.0007

Site 4

1300.55	2	2600.10	1707.775	892.33	(HexNac2Hex3)	IDADF ^N ATFYSMANK	0.1156	0.0043	0.1070	0.0040
1373.58	2	2746.15	1707.775	1038.38	(HexNac2Hex3)Fuc	IDADF ^N ATFYSMANK	0.0011	0.0001	0.0015	0.0024
1402.09	2	2803.18	1707.775	1095.41	(HexNac2Hex3)HexNac	IDADF ^N ATFYSMANK	0.3332	0.0074	0.3016	0.0061
1462.60	2	2924.20	1707.775	1216.43	(HexNac2Hex3)Hex2	IDADF ^N ATFYSMANK	0.0290	0.0012	0.0327	0.0010
1475.12	2	2949.23	1707.775	1241.46	(HexNac2Hex3)HexNacFuc	IDADF ^N ATFYSMANK	0.0000	0.0000	0.0023	0.0016
1483.12	2	2965.23	1707.775	1257.46	(HexNac2Hex3)HexNac1Hex1	IDADF ^N ATFYSMANK	0.2710	0.0044	0.2962	0.0027
1543.62	2	3086.24	1707.775	1378.47	(HexNac2Hex3)Hex3	IDADF ^N ATFYSMANK	0.0533	0.0040	0.0542	0.0024
1564.14	2	3127.28	1707.775	1419.51	(HexNac2Hex3)HexNacHex2	IDADF ^N ATFYSMANK	0.0539	0.0017	0.0588	0.0017
1624.65	2	3248.31	1707.775	1540.53	(HexNac2Hex3)Hex4	IDADF ^N ATFYSMANK	0.1003	0.0020	0.1007	0.0009
1645.17	2	3289.33	1707.775	1581.56	(HexNac2Hex3)HexNacHex3	IDADF ^N ATFYSMANK	0.0118	0.0001	0.0140	0.0011
1705.68	2	3410.36	1707.775	1702.58	(HexNac2Hex3)Hex5	IDADF ^N ATFYSMANK	0.0268	0.0005	0.0262	0.0006
1786.71	2	3572.41	1707.775	1864.64	(HexNac2Hex3)Hex6	IDADF ^N ATFYSMANK	0.0040	0.0002	0.0048	0.0002

Site 5

1164.22	3	3490.66	2468.25	1022.41	HexNac2Hex2Fuc2	LAPTYQELADTYANATSDVLIK	0.0075	0.0004	0.0028	0.0005
1169.55	3	3506.66	2468.25	1038.41	(HexNac2Hex3)Fuc	LAPTYQELADTYANATSDVLIK	0.0837	0.0036	0.0704	0.0015
1188.56	3	3563.68	2468.25	1095.43	(HexNac2Hex3)HexNac	LAPTYQELADTYANATSDVLIK	0.1405	0.0127	0.2366	0.0117
1218.24	3	3652.71	2468.25	1184.46	(HexNac2Hex3)Fuc2	LAPTYQELADTYANATSDVLIK	0.1054	0.0052	0.0604	0.0021
1223.57	3	3668.72	2468.25	1200.47	(HexNac2Hex3)FucHex	LAPTYQELADTYANATSDVLIK	0.0000	0.0000	0.0278	0.0025
1228.90	3	3684.70	2468.25	1216.45	(HexNac2Hex3)Hex2	LAPTYQELADTYANATSDVLIK	0.0506	0.0023	0.0874	0.0021
1237.25	3	3709.74	2468.25	1241.49	(HexNac2Hex3)FucHexNac	LAPTYQELADTYANATSDVLIK	0.2801	0.0133	0.1995	0.0039
1256.25	3	3766.75	2468.25	1298.50	(HexNac2Hex3)HexNac2	LAPTYQELADTYANATSDVLIK	0.1281	0.0105	0.0975	0.0050
1272.26	3	3814.77	2468.25	1346.52	(HexNac2Hex3)Fuc2Hex	LAPTYQELADTYANATSDVLIK	0.0000	0.0000	0.0294	0.0021
1285.93	3	3855.79	2468.25	1387.54	(HexNac2Hex3)Fuc2HexNac	LAPTYQELADTYANATSDVLIK	0.1087	0.0042	0.0633	0.0021
1304.94	3	3912.81	2468.25	1444.56	(HexNac2Hex3)FucHexNac2	LAPTYQELADTYANATSDVLIK	0.0371	0.0052	0.0547	0.0061
1339.95	3	4017.84	2468.25	1549.59	(HexNac2Hex3)Fuc2HexNacHex	LAPTYQELADTYANATSDVLIK	0.0085	0.0003	0.0452	0.0211
1353.62	3	4058.87	2468.25	1590.62	(HexNac2Hex3)Fuc2HexNac2	LAPTYQELADTYANATSDVLIK	0.0496	0.0012	0.0249	0.0083

Chapter 5

Concluding remarks and future perspectives

Concluding remarks and future perspectives

The goal of this thesis was to evaluate the potential of a recombinant glycoprotein-based vaccine for sheep against the blood-sucking parasite *Haemonchus contortus*. The vaccine consisted of heterologously expressed *H. contortus* gut proteases in glycoengineered insect cells. This system offered the ability to produce antigens carrying *N*-glycans very similar to the ones found in the pathogen. Unfortunately none of the tested candidates induced protection in sheep but during the scope of the study a high rate of polymorphism was detected in one of the most promising vaccine candidate proteins. A special focus was therefore set on the analysis of this newly discovered heterogeneity on both protein and *N*-glycans.

In Chapter 2, the insect cell expression system was chosen due to the potential of the targeted manipulation of the *N*-glycosylation machinery to engineer *N*-glycans by using nematode-specific glycosyltransferases. We glycoengineered High five insect cells by overexpressing the *C. elegans* galactosyltransferase GALT-1. The suitability of these manipulated cells as a recombinant expression system for the production of carbohydrate-based vaccines against *H. contortus* was tested. Opposed to some previous studies, the purified recombinant proteins were decorated with *N*-linked glycans and they were highly similar to the ones found on native proteins of *H. contortus* [1], including the galactosylated core α 1,6 fucose and the immunogenic core α 1,3 fucose. Furthermore, we produced various digestive proteases of *H. contortus* in our insect cell expression system to test combinations of these putative antigens as vaccines. We expected to raise protection levels in sheep against *H. contortus* by targeting multiple worm proteins, involved in the proteolytic digestion of hemoglobin (the major food source of the parasite), decorated with native nematode-like *N*-glycans. Despite a significant anti-protein and anti-glycan response stimulated in sheep vaccinated with the glycoengineered proteins, no protection was observed, suggesting that the glycans might contribute to the antigenicity but do not induce protective immunity.

Although we were not able to discover a protective antigen among the targets expressed and purified from insect cells, this system represents an ideal production system for future vaccine production. High quantities of glycoprotein with extensive Golgi-modified *N*-glycans could be generated with this fast and flexible expression

system. The cells were highly susceptible to *N*-glycan manipulations which indicates also future modifications could be possible, e.g. by overexpression of the *C. elegans* fucosyltransferase FUT-6 [2] to produce trifucosylated core *N*-glycans (after the removal of the α 1,6 mannose). Moreover, a recent study reappraised the High five *N*-glycome by using HPLC and MALDI-TOF MS/MS in combination with selected digestions of *N*-glycans and revealed that other *H. contortus* *N*-glycan modifications like phosphorylcholine and the LDNF motif are evident *N*-glycan modifications in High five insect cells [3]. It might be necessary to overexpress the glycosyltransferases responsible for the generation of the LDNF motif to increase the abundance of this epitope on the *N*-glycans.

Previous studies reported that reduced protective capacity of the recombinant vaccines might be due to (I) suboptimal or incorrect folding, (II) lack of various post-translational modifications, such as *N*-glycosylation or (III) the absence of other proteins critical in conferring protection in the native vaccine. We postulated that it was mainly due to the latter reason that the recombinant vaccines were not protective so far. Since the *H. contortus* genome encodes five different H11 isoforms and, in combination with the high level of genetic diversity in *H. contortus* [4], this might explain the failure of recombinant H11 protein to elicit a protective immune response. Based on this hypothesis, we examined in Chapter 3 the distribution and conservation of *H11* antigen genes within and between three different intercontinental parasite populations, as this might have fundamental implications on the development of a globally active recombinant H11 vaccine. H11 isoform sequences were cloned from cDNA and displayed an extreme genetic diversity in all three worm populations analyzed (Switzerland, Scotland and South Africa). Interestingly, the resulting amino acid sequence variation of the H11 proteins was found almost exclusively on the surface of the modeled protein structure (>88% on average), leaving the internal structure of the protein unaffected. Furthermore, we showed that the amino acid variation was affecting the *N*-linked glycosylation of the isoforms, leading to the loss or gain of half of the *N*-glycosylation sites among the isoforms. In addition, a mass spectrometry-based analysis revealed site-specific *N*-glycan processing of the proteins with differing *N*-glycan structures between isoforms and between the variants of one isoform. This site-specific *N*-glycan processing was detected in both, the recombinantly expressed proteins from insect cells, but also in the native H11 isoforms purified from adult *H. contortus* worms.

This natural structural complexity in epitopes and varying glycosylation of the native H11 proteins may be the key requirement for stimulating a very broad antibody response necessary for protection and might explain why antibodies generated by single recombinant vaccines induce only low levels of protection. Moreover, there was a substantial genetic diversity within and among *H. contortus* worm population. The data of Chapter 3, explain the lack of protection of recombinantly produced H11 in the sheep immunization trial in Chapter 2. This *H11* gene was synthesized by a company based on the published sequence in GenBank X94187.1, but shows 5 % genetic variation as compared to the H11 sequences analyzed in the Swiss laboratory strain used for the *H. contortus* infection in the vaccination trial performed in Zürich. It remains a substantial challenge to produce recombinant versions inducing the same levels of protection as with native vaccines, as demonstrated by over 100 published vaccine trials with recombinant antigens over the last decades that were unsuccessful against multiple parasitic helminths.

In Chapter 3, we hypothesized that the H11 variation is driven by the need to evade the host immune response. Using maximum-likelihood models, we showed that positive diversifying selection is acting on the H11 genes, offering support for this hypothesis. We compared the high surface variation and rearrangement of *N*-glycans in H11 with the very well-studied example of the highly variable human immunodeficiency virus 1 (HIV-1) that avoids recognition by the host immune response with its variability. It has been shown that the combination of the high surface variation of HIV-1 with the *N*-glycan shielding allows the virus to conceal itself from neutralizing antibodies, as only 3 % of the surface was still immunoglobulin-domain accessible [5].

An important next step would be to prove the hypothesis that native H11 vaccinated sheep sera are protected due to inhibition of the gut protease (in particular H11) enzyme activities. In a first approach, one could establish *in vitro* enzyme activity assays with purified recombinantly expressed H11 antigen candidates and the purified immunoglobulins from protective sheep sera binding to these antigen. If this can be shown, an antibody epitope mapping could give a greater insight into the most important surface regions of H11 to be targeted for future vaccine generation. In a parallel attempt a focus could be led on the observed *N*-glycan variability in the different H11 variants. Earlier studies showed that human immunodeficiency viruses could acquire a neutralization escape phenotype by introducing the glycosylation pattern of

the corresponding escape variant, this involved a combination of glycan additions, removals and shifts in *N*-glycosylation sites ('evolving glycan shield') [6]. Due to this fact, a reasonable question is if introducing new *N*-glycosites by site-directed mutagenesis in H11 leads to inaccessibility of the protein structure by generated host immunoglobulins. Here the insect cell system again offers a great tool to introduce this kind of modifications and later on use the recombinantly expressed H11 versions (with additional *N*-glycans) in protein-antibody binding studies, e.g. H11 enzyme activity assays.

We identified distinct geographical surface patches on H11 isoforms when the amino acid variation was marked on the modelled protein by the different geographical isolation of the worm population. Not only the genetic background of the parasite populations is different among the different countries, but also the selection pressures acting on them might differ from region to region. One might expect that different mutations are important in different locations. In the case of the benzimidazole anthelmintic resistance induced by isotype-1 β -tubulin variation, this is known to be true. For example, the resistant β -tubulin haplotypes with the F167Y and F200Y mutations are frequently found in the UK but not in India, where the E198A mutation is widespread [7]. Studying more H11 sequences of different populations will allow to study geographical selection pressures and in addition will let to the identification of amino acid replacements that are selected in all populations. The H11 characterization by single worm sample sequencing would increase the quantitative readout in future studies. Specific regional studies will be needed to test which amino acid sites are of general global importance to lead to the development of a globally effective, non-geographical limited vaccine.

A less well-studied fact is that *N*-glycan structures add an additional level of information and variability to the protein surface. In Chapter 4 [8], we examined in more detail how the differential processing of *N*-glycans is generated by the non-template driven *N*-glycan processing machinery in insect cells and how the processing is influenced by the underlying protein sequence. The yeast model protein PDI1 was therefore expressed in insect cells and the novel mass spectrometry-based analytics revealed site-specific *N*-glycan processing of the protein. Molecular dynamics simulation, mutational analysis and kinetic studies of *in vitro* processing events showed that the PDI1 protein conformation had a defining role in the *N*-glycan processing, it determined

the final glycan structure by interacting with the covalently linked *N*-glycan. Further studies are needed, to identify which exact amino acid residues make contact with the glycans. This information will allow to selectively influence the final glycosylation on the protein by manipulating the primary amino acid sequence of the protein. This would lead to a new and useful tool to further reach native state of glycosylation on vaccine candidates.

In summary, this work presented a different view on the requirements for vaccine-induced protection. Many parasite vaccine candidates have failed to be protective. It was always speculated that a lack of native posttranslational modifications resulted in non-protection. Here we show that high genetic variations in *H. contortus* populations might be the main reason for failures in the recombinant vaccine generation. It is absolutely crucial that future research unravels the requirements for a correct epitope presentation, the underlying mechanisms of vaccine protection and the complex immune response necessary for conferring resistance.

References

1. Paschinger, K. and I.B. Wilson, *Two types of galactosylated fucose motifs are present on N-glycans of Haemonchus contortus*. Glycobiology, 2015. **25**(6): p. 585-90.
2. Yan, S., et al., *Array-assisted characterization of a fucosyltransferase required for the biosynthesis of complex core modifications of nematode N-glycans*. J Biol Chem, 2013. **288**(29): p. 21015-28.
3. Stanton, R., et al., *The underestimated N-glycomes of lepidopteran species*. Biochimica et Biophysica Acta (BBA) - General Subjects, 2017. **1861**(4): p. 699-714.
4. Troell, K., et al., *Global patterns reveal strong population structure in Haemonchus contortus, a nematode parasite of domesticated ruminants*. Int J Parasitol, 2006. **36**(12): p. 1305-16.
5. Pancera, M., et al., *Structure and immune recognition of trimeric pre-fusion HIV-1 Env*. Nature, 2014. **514**(7523): p. 455-61.
6. Mascola, J.R. and D.C. Montefiori, *HIV-1: nature's master of disguise*. Nat Med, 2003. **9**(4): p. 393-4.
7. Gilleard, J.S. and E. Redman, *Genetic Diversity and Population Structure of Haemonchus contortus*. Adv Parasitol, 2016. **93**: p. 31-68.
8. Hang, I., et al., *Analysis of site-specific N-glycan remodeling in the endoplasmic reticulum and the Golgi*. Glycobiology, 2015. **25**(12): p. 1335-49.

Curriculum vitae

Personal Data

Name	Susanna Maria Fleurkens
Date of birth	22.10.1983
Place of birth	Visp, Switzerland
Nationality	Dutch and Spanish
Address	Würzwies 10 8064 Zürich

Education

2014 - 2017	PhD thesis in Life Sciences Eidgenössische Technische Hochschule, ETH Zürich, Switzerland Group of Prof. Markus Aebi
2007 - 2009	Master of Science in Microbiology Eidgenössische Technische Hochschule, ETH Zürich, Switzerland
2003 - 2008	Bachelor of Science in Biology Eidgenössische Technische Hochschule, ETH Zürich, Switzerland
1998 - 2003	High school Gymnasium Kollegium Spiritus Sanctus, Brig, Switzerland