Structural Proteomics of the Fungal Cell Wall

Dissertation

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer. nat.)

des Fachbereichs Chemie der Philipps-Universität Marburg

vorgelegt von

Viktoria Reithofer
aus Mönichkirchen am Wechsel, Österreich

Marburg an der Lahn, Dezember 2020

Die vorliegende Dissertation wurde von September 2016 bis Dezember 2020 am Fachbereich Chemie der Philipps-Universität Marburg unter Leitung von Prof. Dr. Lars-Oliver Essen angefertigt.

Vom Fachbereich Chemie der Philipps-Universität Marburg
(Hochschulkennziffer 1180) als Dissertation angenommen am _____

Erstgutachter(in): Prof. Dr. Lars-Oliver Essen

Zweitgutachter(in): Prof. Dr. Hans-Ulrich Mösch

Tag der Disputation: _____

“If knowledge can create problems, it is not through ignorance that we can solve them.”

- Isaac Asimov

Eidesstattliche Erklärung

Ich erkläre, dass eine Promotion noch an keiner anderen Hochschule als der Philipps-Universität Marburg, Fachbereich Chemie, versucht wurde.

Hiermit versichere ich, dass ich die vorliegende Dissertation

Structural Proteomics of the Fungal Cell Wall

selbstständig, ohne unerlaubte Hilfe Dritter angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Dritte waren an der inhaltlich-materiellen Erstellung der Dissertation nicht beteiligt; insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden. Mit dem Einsatz von Software zur Erkennung von Plagiaten bin ich einverstanden.

_____ _____

Ort/Datum Unterschrift (Viktoria Nina Reithofer)

Verlauf des wissenschaftlichen Werdegangs

In der elektronisch publizierten Version nicht enthalten

Zusammenfassung

Pilze sind von einer dicken Schicht aus Kohlenhydraten und Proteinen umgeben, die für die Lebensfähigkeit der Zelle essentiell ist – der pilzlichen Zellwand. Proteine sind auf unterschiedliche Arten in dieses Organell integriert: einige sind kovalent an den Kohlenhydratanteil der Zellwand gebunden, entweder über Glycosylphosphatidylinositol (GPI)-Anker oder alkaliempfindliche Bindungen, andere indirekt über Disulfidbindungen. Zellwandproteine sind an unterschiedlichen zellulären Funktionen beteiligt, wie der Zellwandbiosynthese, der Adhäsion an Oberflächen oder der Sensorik.

Im ersten Teil dieser Arbeit wurden die GPI-verankerten Proteine des thermophilen Modellorganismus *Chaetomium thermophilum* identifiziert. Zunächst wurde eine Vorhersage der an Zellwand und Plasmamembran befindlichen GPI-Proteine durchgeführt. Die Vorhersage wurde durch den massenspektrometrischen Nachweis der GPI-verankerten Zellwandproteine in isolierten Zellwänden ergänzt. Die detektierten Proteine wurden hinsichtlich ihrer Funktionen und mutmaßlichen Rollen analysiert. Interessante Targets für pharmazeutische Anwendungen und Grundlagenforschung konnten ermittelt werden, u. a. Gel1/2, Kre9/Knh1 und Ecm33. Zusätzlich wurde die Ultrastruktur der Zellwand von *C. thermophilum* mittels Transmissionselektronenmikroskopie analysiert, wobei kurze Mikrofibrillen in der äußeren Zellwandschicht und Ähnlichkeit zu der Zellwand von *S. cerevisiae* festgestellt werden konnten.

Die Arbeit behandelt im zweiten Teil die Analyse der A-Domänen der *Candida glabrata* Adhäsine Awp1 und Awp3, die Mitglieder des Adhäsinclusters VI sind. Obwohl diesem humanpathogenen Pilz bestimmte Virulenzfaktoren - z. B. zur Hyphenbildung - fehlen, werden *C. glabrata* Infektionen häufig beobachtet, wobei sein großes Repertoire an Adhäsinen einer der wesentlichen Gründe sein sollte. Awp1A und Awp3A bestehen beide aus einer β-Helix-Domäne und einer α-Kristallin-Domäne. Sie ähneln strukturell kohlenhydratbindenden Proteinen, z. B. Polysaccharid-Lyasen. Allerdings konnte keine Bindung von Kohlenhydraten an Awp1-Typ Adhäsinen nachgewiesen werden. Ein Sequenzähnlichkeitsnetzwerk leitet eine hohe Ähnlichkeit zu den Adhäsinen Awp2 und Awp4 des Adhäsinclusters V ab und untermauert damit frühere Klassifizierungen. Die Strukturen von Awp1 und Awp3 geben erste Einblicke in neue Typen von Adhäsinen in *C. glabrata*, zu denen Adhäsine der Cluster V und VI gehören.

Weiterhin wurde der G-Protein-gekoppelte Rezeptor Pth11 aus *C. thermophilum* analysiert. Er enthält eine N-terminale CFEM-Domäne - diese Domäne kommt ausschließlich in Pilzzellwand- und Plasmamembranproteinen vor -, die als Ligandenbindungsstelle vorhergesagt wurde. Die CFEM-Domäne von *Ct*Pth11 besteht aus fünf α-Helices und weist zwei potenzielle Bindungsstellen auf, die durch F48 geteilt werden. Bestimmte Orientierungen des Aminosäurerestes F48 ermöglichen die Bildung eines Tunnels durch die Domäne. Ein Modell der CtPth11-CFEM-Domäne und der Transmembranregion - basierend auf der Vorhersage benachbarter Reste mittels Sequenzkovarianzanalyse - zeigt, dass beide potenziellen Bindungsstellen zugänglich sind. In einem Fragment-Screen wurden vier Fragmente an der gleichen Bindestelle gebunden; drei davon konnten in die jeweiligen Elektronendichten modelliert werden. Diese hydrophoben Fragmente sind in der hydrophoben Bindestelle platziert und weisen nur wenige zusätzliche Interaktionen auf, was zu der Hypothese passt, dass Pth11 hydrophobe Charakteristika auf der Pflanzenoberfläche wahrnimmt.

Summary

Fungi are surrounded by a thick layer of carbohydrates and proteins, which is essential for the cell's viability – the fungal cell wall. Proteins are incorporated into this organelle in different ways: some are covalently linked to the carbohydrate moiety of the cell wall via Glycosylphosphatidylinositol (GPI)-anchors or alkali-sensitive linkages, others are indirectly attached to the cell wall via disulfide bonds. Cell wall proteins are involved in various cellular functions, such as cell wall biosynthesis, adhesion to external surfaces, or sensing.

The GPI-anchored cell wall proteome of the thermophilic model organism *Chaetomium thermophilum* was identified in the first part of this thesis. First, a prediction of GPI-proteins, anchored to the cell wall and the plasma membrane was done. The prediction was then complemented by mass-spectrometric identification of GPI-anchored cell wall proteins in isolated cell walls. The detected proteins were then analyzed concerning their functions and putative roles and interesting targets for pharmaceutical applications and fundamental research were established, including Gel1/2, Kre9/Knh1, and Ecm33. In addition, the ultrastructure of the *C. thermophilum* cell wall was analyzed via transmission electron microscopy, revealing short microfibrils in its outer layer and its similarity to the cell wall of *S. cerevisiae*.

The thesis then advances to the analysis of the A-domains of the *Candida glabrata* adhesins Awp1 and Awp3, which are members of adhesin cluster VI. Although the fungal pathogen lacks certain virulence factors – such as hyphae formation – *C. glabrata* infections are commonly observed; its large repertoire of adhesins is believed to be the reason therefore. Awp1A and Awp3A both consist of a β-helix domain and an α-crystallin domain. They are structurally similar to carbohydrate binding proteins, e. g. polysaccharide lyases, but carbohydrate binding could not be observed. A sequence similarity network (SSN) elucidates their high similarity to cluster V adhesins Awp2 and Awp4 and thereby reinforces previous classifications. The structures of Awp1 and Awp3 provide first insights into new types of adhesins in *C. glabrata* that include the adhesin clusters V and VI.

Furthermore, the G-protein coupled receptor Pth11 from *C. thermophilum* was analyzed. It contains an N-terminal CFEM domain – a domain exclusively found in fungal cell wall and plasma membrane proteins – that is predicted to be the ligand binding site. The *Ct*Pth11 CFEM domain consists of five α-helices and reveals two potential binding sites, divided by F48. Distinct conformers of F48 allow formation of a tunnel through the domain. A model of the *Ct*Pth11 CFEM domain and transmembrane region – based on prediction of neighboring residues via sequence covariation analysis – shows that both potential binding sites are accessible. In a fragment screen, four fragments were bound in the same cavity; three of them could be fitted into their respective electron densities. These hydrophobic fragments are placed in the hydrophobic cavity, with only few additional interactions, which is in accordance with the proposal that Pth11 senses hydrophobic cues on the plant surface.

Table of Contents

1. Introduction

1. 1. The Fungal Cell Wall

Fungi are covered by a 110 – 200 nm thick carbohydrate layer, the fungal cell wall. The wall provides high stability to the cell, but is also subject to constant remodelling[1]. It constitutes 15 – 30% of the total dry mass of the cell in vegetative *Saccharomyces cerevisiae*[2]. Its importance is additionally underlined by the fact that approximately one-fifth of the yeast genome is dedicated to cell wall biosynthesis and remodeling[1]. The fungal cell wall fulfills various functions that are crucial for the cell's viability: Maintenance of osmotic homeostasis, protection from mechanical damage, determination of the cell shape along the whole cell cycle and providing a scaffold for extracellular proteins. Proteins within the wall vary in their function, amongst other things they are involved in cell wall synthesis and remodeling, sensing, adhesion, or nutrient acquisition[1,3].

Since the cell wall is an essential compartment of the fungal cell and is at the same time distinct from the cell walls or membranes of mammals, plants or bacteria, it is generally considered a promising target for the development of antifungal drugs[1,4].

1. 1. 1. Structure of the Fungal Cell Wall

The unique structure of the fungal cell wall enables it to fulfill its diverse functions. A schematic representation of the cell wall is depicted in Figure 1. The cell wall is often divided into an inner layer, which is rich in carbohydrates, and an outer layer, which is rich in protein[5]. The two layers can be differentiated in transmission electron microscopy (TEM) images. The inner wall consists of chitin, β-1,3-glucan, and β-1,6-glucan. A thin layer of chitin surrounds the plasma membrane and provides rigidity to the cell wall. Chitin is essential for cell wall integrity[3]; cell wall defects are often compensated by the fungus through increased levels of chitin in the cell wall[2]. In *S. cerevisiae* only 1.5 – 6% of the cell wall mass consist of chitin[3], whereas in filamentous fungi – like *Aspergillus fumigatus* – it can constitute up to 15% of the whole cell wall mass[5]. The major component of the cell walls of fungi characterized so far is β-1,3-glucan, which forms a three-dimensional network. Other components, such as certain proteins, are embedded in this network[3]. Highly branched β-1,6-glucan was identified in several fungi, including yeast-like *Saccharomycetales*. It plays a role in crosslinking the different constituents of the fungal cell wall[5]. Cell wall proteins are usually highly glycosylated by the addition of branched mannose chains and are therefore called mannoproteins. They are enriched in the outer layer of the cell wall[5].

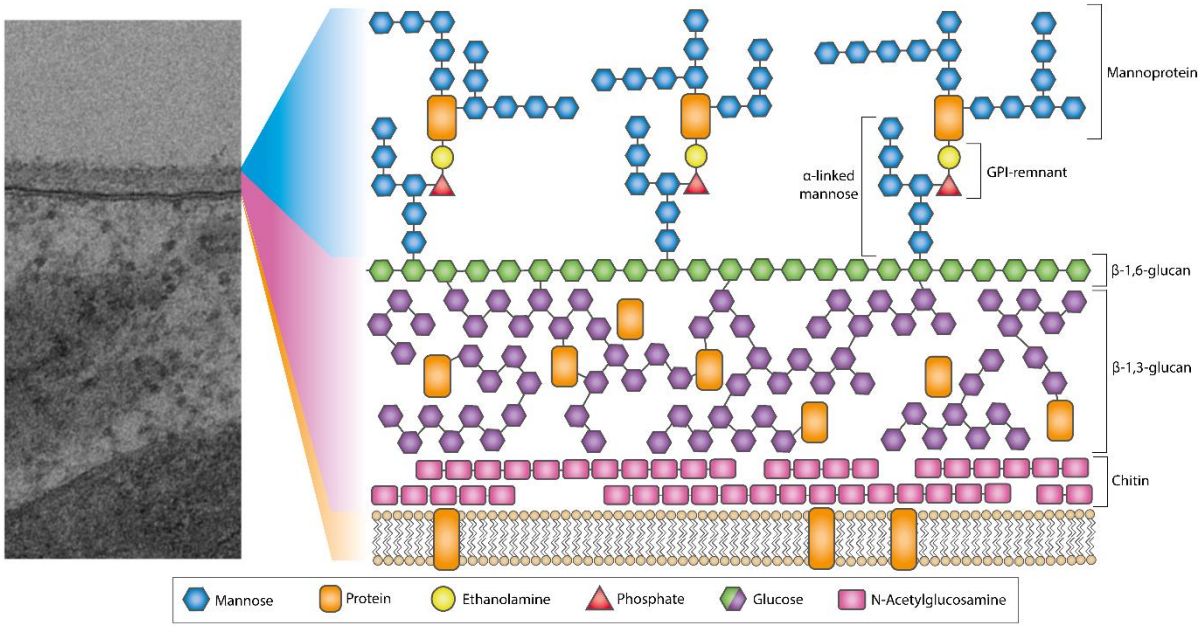


Figure 1: Schematic representation of the fungal cell wall, adapted from Cassone (2013)[6]

In TEM images the two layers of the fungal cell wall can be distinguished – in this example an image of the *C. thermophilum* cell wall is shown. The protein-rich outer wall is more electron dense - i.e. appears darker on the image - than the inner, carbohydrate-rich wall. The plasma membrane is visible as a very electron dense bilayer. The right panel shows a schematic representation of the cell wall. The inner layer of the cell wall consists of chitin, β-1,3-glucan, and β-1,6-glucan, as well as proteins. In the outer layer of the cell wall, mannoproteins can be found. Many of those are GPI-anchored proteins, connected to the β-1,6-glucan moiety of the cell wall via a few mannose units and a remnant of the GPI-anchor.

The composition of the cell wall varies considerably in different fungi. Additional components were identified in some fungi. The most striking example might be melanin, which is responsible for the black color of certain fungi[5]. Other fungi were found to lack particular cell wall components. For example, no β-1,6-glucan could be detected in the cell wall of *A. fumigatus*[7].

1. 1. 2. Incorporation of proteins into the fungal cell wall

Proteins are incorporated into the cell wall in different ways: A few proteins are ester-linked to the β-1,3-glucan moiety of the cell wall and can be released by treatment with mild alkali, they are therefore often referred to as **ASL (alkali sensitive linkage) cell wall proteins** (CWPs). Also the term **PIR (proteins with internal repeats) CWPs** is commonly used for those proteins, because they contain multiple repeats of the sequence DGQ[hydrophobic amino acid]Q[2,3]; a linkage between the central glutamine residue (Q) and β-1,3-glucan attaches them to the cell wall[8]. PIR-CWPs can form several linkages, thus they are able to interconnect glucans. Single PIR repeats can also be found in certain glycosylphosphatidylinositol (GPI)-anchored CWPs. Fungal cell walls also contain **disulfide-linked CWPs**, which are thought to be connected to the

cell wall either directly or indirectly by being linked to other proteins. They can be released from the cell wall using sulfhydryl reagents[2].

The majority of proteins in fungal cell walls are **GPI-CWPs**. In eukaryotic genomes approximately 1% of encoded proteins are post-translationally modified by addition of a GPI-anchor. The anchor's core structure is conserved in mammals, protozoa, and yeast; modifications can be species- or even tissue-specific. GPI-anchored proteins have to undergo a maturation process before they reach the cell surface: The GPI-anchor is pre-assembled in the endoplasmic reticulum (ER) to which the protein is directed by a signal peptide. In the lumen of the ER, a specific signal sequence is recognized at the protein's C-terminus. The C-terminal end of the protein, up to the so-called "ω site", is removed and replaced by the GPI-anchor. GPI-proteins then go through the secretory pathway, during which glycans and lipids of the GPI-anchor are subject to several modifications. At the cell surface, a lipid portion of the GPI-anchor embeds it into a single leaflet of the membrane[9]. In fungi, proteins can then be linked to the β-1,6-glucan moiety of the cell wall via a remnant of the GPI-anchor. This is achieved by a transglycosylation reaction, catalyzed by a member of the glycoside hydrolase (GH) 76 family[10]. In this context, it should be noted that possibly the majority, but not all GPI proteins are relocated to the carbohydrate moiety of the cell wall; some remain at the plasma membrane, others are found in both locations[9].

As mentioned above, GPI-anchored proteins have certain features that can be used for their identification, specifically an N-terminal signal peptide and a C-terminal GPI anchor attachment sequence. The GPI anchor attachment sequence itself also possesses particular characteristics: the GPI-attachment site (ω-site) is typically a G, A, S, N, D, or C. N-terminal from the ω site lies the ω- region that consists of around 10 polar amino acids (ω-10 to ω-1), which serve as a flexible linker. ω+2 is restricted to G, A, S, or V, it is followed by a spacer region of 4 – 19 amino acids and a stretch of hydrophobic amino acids that varies in length, but is able to span the membrane. Upon GPI-anchor attachment, the peptide bond between ω and ω+1 is cleaved[9,11,12].

Consensus sequences for the GPI anchor attachment sequence have been described in several publications[11–13]. In this study, detection of GPI-anchored proteins has been done using the Big-PI Fungal Predictor[12]. In addition, the following sequence was used for detection of GPI-anchored proteins via a pattern search[11]:

[NSGDAC]–[GASVIETKDLF]–[GASV]–X(4,19)–[FILMVAGPSTCYWN](10)>

The final location of GPI-anchored proteins in fungi – i. e. the plasma membrane or the cell wall – is proposed to be influenced by residues in the ω- region of the GPI attachment signal sequence. Proteins that are located at the plasma membrane usually contain basic amino acids in positions ω-1 and ω-2[11], typically in form of a dibasic motif[13]. The ω- region of GPI-anchored proteins that are sorted to the cell wall is considerably different: typically, V, I or L are located at positions ω-4 and ω-5 and Y or N at ω-2[11].

1. 2. *Chaetomium thermophilum* – a thermophilic model organism for biochemical studies

Proteins derived from thermophilic organisms are generally considered more stable than their corresponding mesophilic orthologues[14]. The most prominent example for this phenomenon might be the DNA-polymerase of *Thermus aquaticus*[15]. The production of more heat tolerant proteins is not only of high interest for industrial applications[16], but also biochemical and structural studies profit from the usage of thermally stable proteins, as these also tend to be highly stable at lower temperatures[14]. For this reason, proteins derived from thermophilic organisms are enthusiastically used for *in vitro* studies, rather than their orthologues originating from mesophilic organisms[17].

In this context, the thermophilic fungus *C. thermophilum* provides a well suited model organism for *in vitro* studies on eukaryotic proteins. The filamentous fungus belongs to the *Ascomycetes* and grows in rotten organics at temperatures of up to 60 °C, with an optimal growth temperature of 50 – 55 °C[18]. The genome of *C. thermophilum* has first been published in 2011[19]. It is available at https://c-thermophilum.bork.embl.de, with annotations updated and curated in 2014. Additionally, its proteome has been analyzed via mass spectrometry, resulting in the identification of 4266 proteins from 7227 predicted protein coding sequences[18]. Increased solubility of heterologously expressed proteins originating from *C. thermophilum* compared to their orthologues from other fungi has been described on several occasions[10,18,20]. Seemingly the fungus also enjoys a certain popularity among structural biologists, as suggested by the 314 PDB entries of proteins derived from *C. thermophilum* (as of November 25th, 2020). Although the fungus is a popular model organism, it has not yet been widely used for the study of cell wall proteins. Structurally characterized *C. thermophilum* cell wall proteins include the glycoside hydrolases (GH) Dfg5 (PDB: 6RY0 and related entries)[10] and Lam55 (PDB: 5M5Z and 5M60)[21].

1. 3. Adhesins in *C. glabrata* – important contributors to the virulence of a yeast-like fungus

The yeast *C. glabrata* is a mammalian commensal that can cause mucosal, blood stream and medical-device related infections. Especially immunocompromised patients are severely affected by *Candida* infections[22,23]. The opportunistic pathogen *C. glabrata* is the second most cause of these infections in human after *Candida albicans*, with increased numbers over the years. In addition, the prerequisite that *C. glabrata* is naturally resistant against azole class antifungal drugs complicates treatment of infections. Interestingly, the organism is phylogenetically more closely related to *S. cerevisiae* than to other *Candida* species and it lacks certain virulence factors, such as hyphae formation[23]. However, *C. glabrata* possesses a remarkably large number of putative adhesins, which are thought to compensate for the lack of other virulence factors[24,25]. These are proteins on the cell's surface that enable the fungus

to adhere to a variety of biologic and abiotic substances. Adhesion to host tissue is considered a critical first step in the establishment of fungal infections and also adhesion to medical devices, followed by biofilm formation, has been described[22].

Adhesins are GPI-anchored proteins, most of which share a particular domain architecture: Being GPI-anchored, they apparently possess an N-terminal signal peptide. The signal peptide is followed by the so called "A-domain" or "effector domain", which harbors the adhesive function. A central serine/threonine-rich region of low complexity and of various lengths – also referred to as "B-domain" – acts as a proteoglycan-like stalk to present the A-domain on the surface of the fungal cell. Lastly, the C-terminal domain contains the GPI anchor attachment signal sequence and is required for the integration of the protein into the cell wall via a GPI-anchor[22,26].

Obviously, the exact number of adhesins in a fungus cannot be specified, but one can compare the numbers of already identified adhesion-like encoding genes in different fungi. This reveals that *C. glabrata* contains an exceptionally large number of adhesins, specifically 67 putative adhesins, which can be identified by domain architecture in the genome of the *C. glabrata* strain ATCC2001/CBS138[22]. In comparison, 25 adhesins were described in *C. albicans* by de Groot *et al.* in 2013[22]. In this context, the plasticity of the *C. glabrata* genome is worth mentioning, i. e. the genome of the organism is highly dynamic. This feature is also found in other pathogens and enables adaptation to environmental changes. In addition, many adhesins are encoded in subtelomeric regions of the genome. Those are regions with a high amount of sequence repeats and therefore particularly susceptible to rearrangements. The presence of sequence repeats also increases the complexity of correct sequencing[27].

Applying the specific domain architecture of adhesins as a criterion for the identification of adhesins, De Groot *et al.* bioinformatically identified novel putative adhesins within the second assembly of the *C. glabrata* genome (2004). Four of those were confirmed via mass spectrometric analysis of the cell walls of different *C. glabrata* strains (ATCC 90876 and ATCC 2001) under varying growth conditions in 2008. Those novel putative adhesins were named **Adhesin-like wall protein (Awp)** 1-4 and represent the first identified members of the Awp family of *C. glabrata* adhesins[24]. Two more novel adhesins were identified in *C. glabrata* stationary phase cells and in biofilms by Kraneveld *et al.* in 2011 and named Awp5/6[25]; Awp7-13 were identified in hyperadhesive clinical isolates of *C. glabrata* in 2015[28]. The putative adhesins Awp1-14 are members of different clusters of *C. glabrata* adhesins, the classification being based on a phylogenetic tree, which was generated using the N-terminal regions of the sequences. The current classification of *C. glabrata* adhesins was published with the newly assembled genome of the organism by Xu *et al.*[27] and generally corresponds to the classification presented by de Groot *et al.* in 2008[24,27].

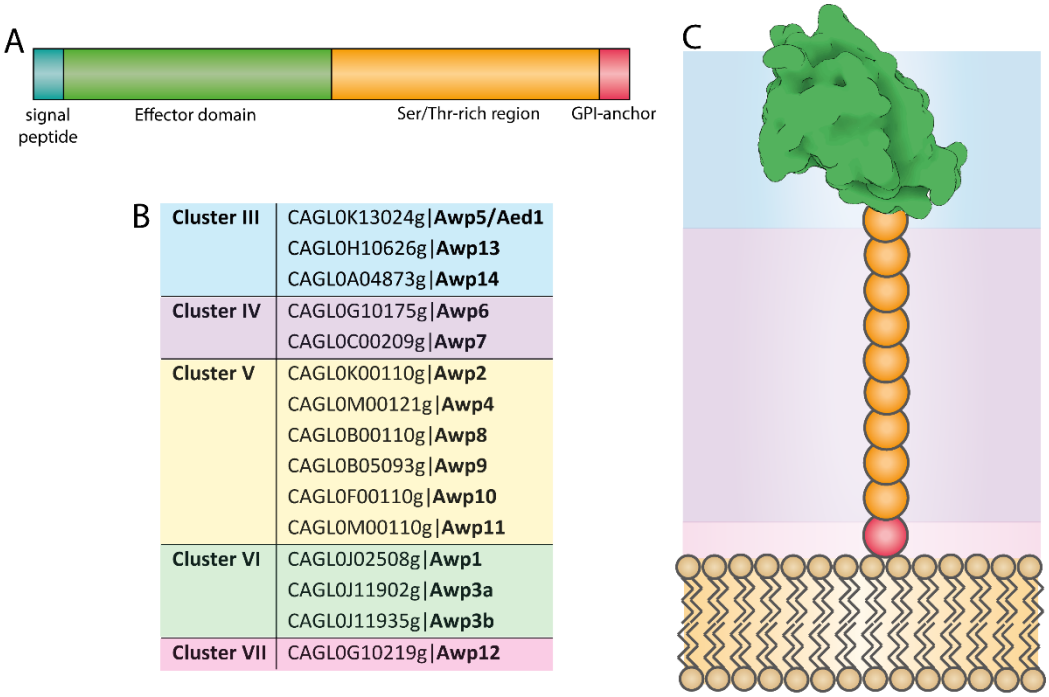


Figure 2: Domain architecture and model of a typical adhesin, classification of Awp's in different clsuters

A) The distinct domain architecture of *C. glabrata* adhesins: The proteins carry an N-terminal signal peptide that targets them to the cell wall. The N-terminal effector domain (or A-domain) has the adhesive function. It is followed by a Ser/Thr-rich region, also referred to as B-domain, which displays the A-domain at the cell's surface. Finally, adhesins are connected to the cell wall or the plasma membrane via a GPI-anchor, they therefore have a GPI anchor attachment sequence. B) Awp's are members of different clusters, as shown in the table. Colors of the clusters were chosen according to Xu *et al.*[27]. The classification is based on Xu *et al.*[27], de Groot *et al.*[24] and Gómez-Molero *et al.*[28]. C) Model of an adhesin: the protein is anchored to the cell wall via a GPI-anchor. A proteoglycan-like stalk (represented by orange spheres) presents the effector domain on the cell surface. The effector domain (shown as a green surface representation of Epa1A) has the adhesive function.

C. glabrata contains 7 different clusters of adhesins, summarized in Appendix I; classification of Awp's is shown in Figure 2. The **Epithelial adhesion (Epa)** family forms **cluster I**. The Epa family consists of 20 members[27], structural information is available on Epa1, Epa6, and Epa9; all containing an anthrax protective antigen 14 (PA14) domain. These proteins bind various carbohydrates, which can be found on the surface of epithelial cells[26,29,30]. Also some other *Candida* species, which are closely related to *C. glabrata*, contain Epa genes. 12 and 9 Epa orthologs were identified in *C. bracarensis* and *C. nivariensis*, respectively, both pathogenic fungi. In contrast, only one Epa gene was found in the non-pathogenic fungus *Nakaseomyces delphensis*, underlining the important role of Epa family members in virulence[31]. **Cluster II** is formed by the **PA14 domain containing Wall Protein (Pwp)** family of adhesins, which has seven members. However, information on this family is rather limited[22]; Pwp7 was shown to be involved in adhesion to human endothelial cells[32]. **Cluster III** contains 14 members, including Awp5/Aed1[27] – which is proposed to be required for adhesion to human endothelial cells[32] –, as well as Awp13 and Awp14. Awp6, which was shown to be upregulated in biofilms[25], and Awp7 constitute **cluster IV**[27]. **Cluster V** contains several Awp's, namely Awp2, Awp4, and

Awp8-11. Proteome mass-spectrometry analysis of hyperadhesive *C. glabrata* strains revealed that the number of peptides from cluster V corresponded to the number of identifiable peptides from Epa family members, suggesting that this cluster also plays an important role in cell adhesion[28]. **Cluster VI** contains Awp1, Awp3a, and Awp3b, amongst other members. The Awp3 gene was misassembled in the 2004 reference genome, the current assembly led to the identification of two paralogs, named Awp3a and Awp3b[27]. Awp12 is a member of **cluster VII** and also its first member to be identified in cell walls via proteome analysis. This is the first indication for biological relevance of cluster VII adhesins[28].

Interestingly, homology of Awp1 and Awp2 – which are members of cluster VI and V, respectively – to Awa1, Hpf1, and Hpf1' from yeast has been described[24]. Awa1 – "awa" is Japanese for foam – is a GPI-anchored cell wall protein unique to sake yeast, which is essential for foam-formation and surface hydrophobicity[33]. Haze protective factors (Hpf) have first been described by Waters *et al.* in 1994[34]. They are cell wall proteins of several *S. cerevisiae* strains and are contained in isolates of wine, where they are proposed to compete with wine proteins for the components that form visible protein aggregates – i. e. haze[35].

The Awp family represents the second largest family of adhesins in *C. glabrata*, but most members are still uncharacterized; structural and biochemical information is lacking. Nevertheless, the identification of these proteins in cell wall isolates of different *C. glabrata* strains, especially in clinical isolates of hyperadhesive strains, indicates that they play significant roles in cell adhesion[28].

1. 4. Proteins with a CFEM domain

The **CFEM (common in several fungal extracellular membrane proteins)** domain is exclusively found in fungal membrane or cell wall proteins. It has a size of around 60 amino acids, with the following consensus sequence:

$$PxC[A/G]x_2Cx_{8-12}Cx_{1-3}[x/T]Dx_{2-5}CxCx_{9-14}Cx_{3-4}Cx_{15-16}$$

The formation of 4 disulfide bonds by the eight cysteines of the domain was first predicted by Kulkarni *et al.* [36] and could be confirmed in structural studies on the CFEM domain containing protein Csa2[37]. The domain can occur in one or more copies in a protein, it is usually located at the N-terminus. N-terminal signal sequences, transmembrane spans or GPI-anchor sequences are often identified in CFEM domain containing proteins[36]. Proteins with a CFEM domain fulfil a variety of functions[37–40]. A classification of CFEM-proteins was done by Dr. Vitali Kalugin via a Sequence Similarity Network (SSN). It is shown in Figure 3 and reveals various families of CFEM domain containing proteins. These also differ in function and domain architecture[41].

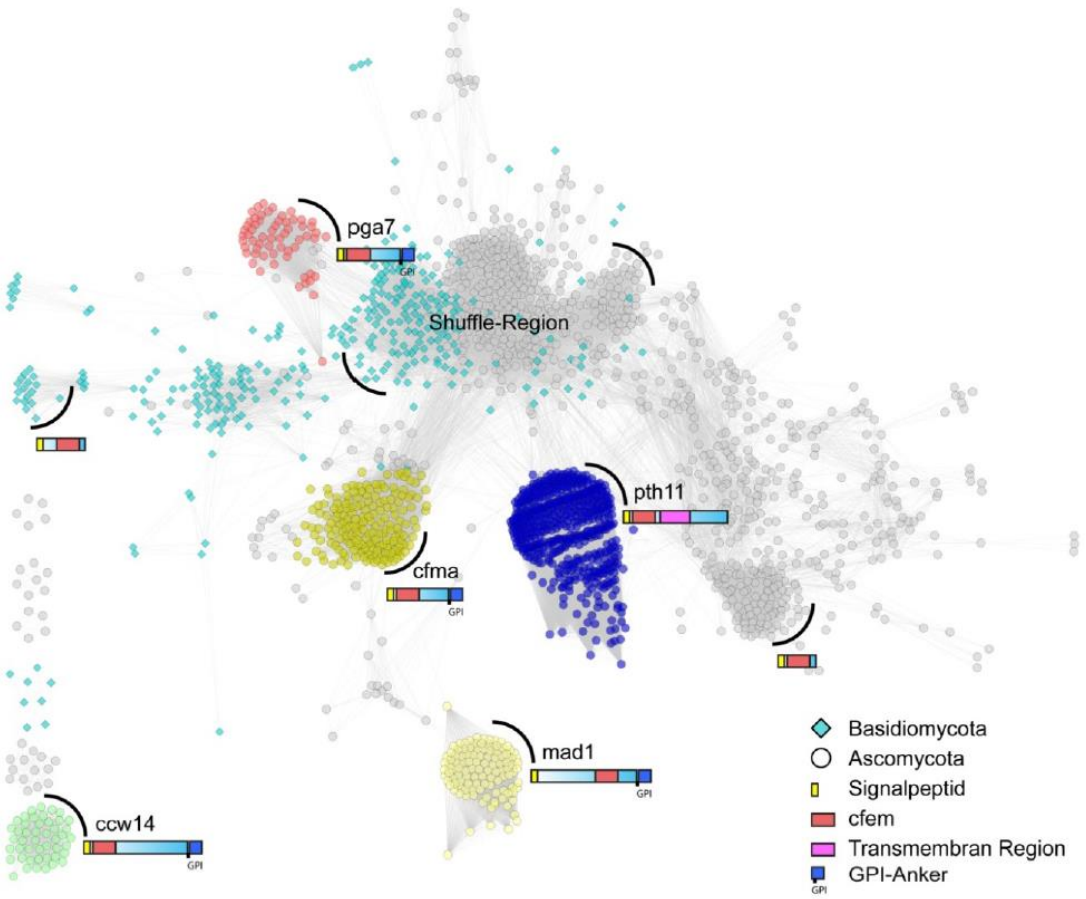


Figure 3: SSN of proteins with a CFEM-domain, created by Dr. Vitali Kalugin[41]

The SSN (created by Dr. Vitali Kalugin) shows different families of proteins with a CFEM domain. Additionally, the domain architecture of the families is given. Protein sequences are represented by so called "nodes" in the network, their relationship to each other is indicated by "edges", which are depicted as lines. The distance between two protein sequences is defined by the BLAST E-value: If a certain E-value cutoff is exceeded, no edge is displayed. The lower the E-value (and thus the more similar two sequences are to each other), the closer together two nodes are. The families Ccw14 (light green) and Mad1 (light yellow) are further away from the main body of the proteins that form the network. Following other families could be identified in the network: Pth11 (dark blue), Pga7 (orange) and Cfma (bright yellow)[41].

The first CFEM domain containing proteins that were identified are Ccw14 (Covalently-linked cell wall protein 14, formerly known as Icwp) from *S. cerevisiae* and Aci1 (Mac1 interacting protein 1) from the rice blast fungus *Magnaporthe oryzae*[39,42]. Ccw14 is a small (238 AA) GPI-CWPs, which is important for maintenance of cell wall integrity. It consists of a signal peptide, the CFEM domain and a GPI anchor attachment sequence[39]. Aci1 shows the same domain architecture; it interacts with the adenylate cyclase Mac1, an essential player in appressorium formation in the *M. oryzae* and is therefore important for the organism's pathogenicity[42,43]. Members of the Pga7 family are proposed to be involved in heme-iron acquisition from hemoglobin in the cell walls of fungal pathogens. Also Csa2, which is structurally characterized in complex with heme, is a member of this family[37].

In this work, the focus will reside on the non-canonical GPCR Pth11, which was shown to be important for appressorium formation in *M. oryzae*[40,44]. The fungus is one of the most relevant plant pathogens worldwide[45,46] and its control poses a major challenge[46]. *M. oryzae* exhibits a remarkable disease cycle that begins with the landing of a conidium on the plant leaf. It forms a germ tube, which quickly develops into an infection structure, the so-called appressorium. The mature appressorium then develops a penetration peg, which enables the fungus to penetrate the plant cell wall. *M. oryzae* can thereby intrude into the cells of the host plant, where hyphae spread, recognizable by lesions on the plant surface. Within 7 days a new disease cycle is induced, as these lesions present numerous freshly developed conidia[44,45].

Pth11 has an N-terminal signal peptide to target the protein to the cell membrane, followed by the CFEM domain, 7 transmembrane helices, and an unknown cytoplasmic domain[40,44]. The N-terminal CFEM domain was shown to be vital for the protein's function via several approaches: first, a deletion of the CFEM domain leads to disruption of appressorium formation and therefore also of plant cell infection, as does a disruption of disulfide bonds within the domain. Complementation with the CFEM domain of *C. albicans* Csa1 cannot compensate for loss of the Pth11 CFEM domain. This observation underlines the functional diversity of different CFEM domains[40]. Pth11 is thought to respond to certain surface cues[44], but its ligand remains unknown[40]. Recently, this GPCR type has also been shown to play a role in the virulence of *Fusarium graminearum*, a plant pathogen that infects cereals and causes the disease Fusarium head blight[47]. Pth11 is regarded as a promising target for the development of novel antifungal agents in agriculture[40].

1. 5. Objectives of the thesis

The first part of the thesis will be focused on the identification of cell wall proteins in *C. thermophilum*. The fungus has been shown to be a promising model organism for biochemical and structural studies of eukaryotic proteins on several occasions[10,19,20]. Also genetic manipulation of *C. thermophilum* is feasible, so that it can be used as a source for purification of thermally stable native macromolecular assemblies[48]. However, the cell wall of the fungus has not yet been characterized, which brings up the first goal of this work: The cell wall proteome of *C. thermophilum* is investigated to reveal attractive candidates for the biochemical and structural characterization of CWPs. These can be proteins involved in cell wall assembly, remodeling, or integrity, which are expected to be of interest of further research to understand these processes in fungi. In addition, the characterization of the GPI-anchored cell wall proteome could also prove to be a useful tool for the identification of new targets for antifungal drugs. To characterize the *C. thermophilum* GPI- and cell wall proteome, a prediction of GPI-anchored proteins will be done using bioinformatics methods. Furthermore, *C. thermophilum* cell walls are isolated and analyzed via mass spectrometry, enabling the identification of GPI-CWPs.

The focus of the thesis will then advance to structural and biochemical studies on certain cell wall proteins: first, adhesins of the Awp family from *C. glabrata* will be analyzed, then the ligand-binding CFEM domain of the GPCR Pth11 from *C. thermophilum* will be characterized.

Mass spectrometric detection of various Awp proteins in the cell walls of various *C. glabrata* strains and clinical isolates suggests that they play a significant role in the infection process[22]. Awp1 and Awp3, which are members of adhesin cluster VI, will be the focus of this thesis. The sequence of Awp3 was misassembled in the older version of the *C. glabrata* genome, which was used for the initial identification of these proteins[25,27]. A *de novo* assembly of the *C. glabrata* genome in 2020 revealed two paralogs of Awp3, named Awp3a and Awp3b. Nevertheless, the sequence of the Awp3 A-domain used in this work remained the same and corresponds to the paralog Awp3b. The sequence of Awp1 remained unchanged[27]. The effector domains of Awp1 and Awp3b will be produced in *E. coli*, purified and structurally characterized. The structures of cluster VI adhesins are expected to provide insights into a novel class of adhesins in *C. glabrata*, as they lack any similarity to the PA14 domain containing Epa family of adhesins. A SSN will be used to elucidate their relationship to other adhesins and to reinforce classification of certain adhesin clusters. In addition, carbohydrate binding studies will be conducted on the Awp1 and Awp3b A-domains.

Characterization of even another adhesin cluster will be pursued by heterologous expression, purification and structural characterization of the A-domain of the cluster III adhesin Awp14. Other members of this cluster, specifically Awp5/Aed1, were shown to adhere to human epithelial cells[32], indicating a function in virulence.

Concerning cell wall proteins with a CFEM domain, the CFEM domain of the GPCR Pth11 from *C. thermophilum* will be characterized. The *C. thermophilum* orthologue was chosen because heterologous expression of the CFEM domain from *M. oryzae* Pth11 in *E. coli* did not result in production of soluble protein. *Ct*Pth11 was identified using the SSN presented above[41]. The *Ct*Pth11 CFEM domain will be produced in *E. coli*, purified and structurally analyzed. Using a fragment screening approach, new information on putative natural ligands of the protein will be obtained.

2. Materials

2. 1. Chemicals

1,5-Pentanediol	*Sigma*
2'-Deoxycytidine 5'-triphosphate disodium salt (dCTP-Na_2)	*Thermo Fisher*
2'-Deoxyguanosine 5'-triphosphate trisodium salt (dGTP-Na_3)	*Thermo Fisher*
2-Bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol (Bis-Tris)	*Sigma*
3-(N-Morpholino)propanesulfonic acid (MOPS)	*Roth*
3-Fucosyllactose	
3-O-(β-D-Galactopyranosyl)-D-galactopyranose	*Carbosynth*
Acetic acid	*VWR*
Agar-agar	*Roth*
Agarose	*Invitrogen*
Ammonium persulfate (APS)	*Merck*
Beta glucan (Barley)	*Megazyme*
Boric acid (H_3BO_3)	*Grüssing GmbH*
Bromphenolblue	*Roth*
Calcium chloride ($CaCl_2$)	*Fluka*
CM-curdlan	*Megazyme*
cOmplete Protease Inhibitor Cocktail	*Roche*
Coomassie brilliant blue R-250	*Serva*
Dextrin (potato)	*Sigma*
Dipotassium phosphate (K_2HPO_4)	*Merck*
Dithiothreitol (DTT)	*Merck*
Erbium(III) chloride ($ErCl_3$)	
Ethanol	*VWR*
Ethylenediaminetetraacetic acid (EDTA)	*Merck*
Gadolinium (III) acetate ($Gd(OAc)_3$)	*Alfa Aesar*
Galα1-3Gal	*Dextra*
Galα1-3Galβ1-4Gal	*Dextra*
Galβ1-3GalNAc	*Dextra*
Galβ1-3GalNAcβ1-4Galβ1-4Glc	*Dextra*
Galβ1-3GlcNAc	*Dextra*
Galβ1-4GlcNAc	*Dextra*
Glucosamine	*Roth*
Glucose	*Roth*
Glycerol	*Roth*
Glycine	*Sigma*

Hydrochloric acid (HCl)	*VWR*
Imidazole	*Merck*
Iron(III) sulfate hydrate ($Fe^2(SO_4)_3$)	*Merck*
Isopropanol	*VWR*
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	*Gerbu*
Kanamycin sulfate	*VWR*
lacto-N-neotetraose	
lacto-N-tetraose	
Laminarin	
Lewis[a] trisaccharide	*Dextra*
Magnesium chloride ($MgCl_2$)	*Merck*
Magnesium sulfate ($MgSO_4$)	*VWR*
Manganese(II) chloride ($MnCl_2$)	*Sigma*
Mannopentaose	*Dextra*
Mannose	*Merck*
Mannotetraose	*Dextra*
Midori Green	*Biozym*
N,N'-diacetylchitobiose	*Dextra*
Peptone	*Difco*
Polyethylene glycol 8000 (PEG 8000)	*Sigma*
Potassium acetate (CH_3COOK)	*Merck*
Rotiphorese Gel 30 (37,5:1)	*Roth*
Rubidium chloride (RbCl)	*Sigma*
Saccharose	*VWR*
Sodium chloride (NaCl)	*VWR*
Sodium dihydrogen phosphate (NaH_2PO_4)	*Merck*
Sodium dodecyl sulfate (SDS)	*AppliChem*
Sodium hydroxide (NaOH)	*AppliChem*
Sorbitol	*Sigma*
Terbium(III) chloride ($TbCl_3$)	
Tetramethylethylenediamine (TEMED)	*Roth*
Tris(hydroxymethyl)aminomethane (Tris)	*Roth*
Tryptone	*Th. Geyer*
Virkon	*VWR*
Yeast extract	*Th. Geyer*
Ytterbium(III) chloride ($YbCl_3$)	
α1,2-mannobiose	*Dextra*
α1,3-mannobiose	*Dextra*
α1,4-mannobiose	*Dextra*
α1,6-mannobiose	*Dextra*
β-Hydroxy-4-morpholinepropanesulfonic acid (MOPSO)	
β-mercaptoethanol	*Roth*

2. 2. Equipment

Device	**Model (*Manufacturer*)**
Autoclave	T-Line (*Fedegari*)
Balance	PC2200 (*Mettler*)
	LabStyle 54 (*Mettler Toledo*)
Bead mill	FastPrep-24 (*MP Biomedicals*)
Centrifuge bottles	1L Superspeed CB with sealing (*Nalgene*)
	JA-20 (*Beckman*)
Centrifuge rotors	F6S 6x1000Y (*Thermo Fisher*)
	JA-20 Fixed Angle Rotor (*Beckman*)
Centrifuges	Centrifuge 5810 R (*Eppendorf*)
	Heraeus Fresco 21 (*Thermo Fisher*)
	J2-HS (*Beckman*)
	Lynx 6000 (*Sorvall*)
Chromatography columns	HiLoad 26/600 Superdex 200 pg (*GE Healthcare*)
	HiLoad 16/600 Superdex 200 pg (*GE Healthcare*)
	HiLoad 26/600 Superdex 75 pg (*GE Healthcare*)
	HiLoad 16/600 Superdex 75 pg (*GE Healthcare*)
	Protino Ni-NTA Column 5 mL (*Macherey-Nagel*)
Chromatography system	NGC Chromatography System (*Bio-Rad*)
Crystallization plate documentation	Rock Imager (*Formulatrix*)
Crystallization robot	Honeybee 963 (*Digilab*)
Electrophoresis chambers	(*Feinmechanische Werkstatt, Chemistry department, PUM*)
	Mini-PROTEAN Tetra Vertical Electrophoresis Cell (*Bio-Rad*)
Gel documentation	Computer E.A.S.Y. (*UVP*)
	Thermal printer UP-D 895 (*Sony*)
	UV-transilluminator (*Herolab*)
Heating block	BT3 (*Grant Instruments*)
Incubators	Certomat IS (*Sartorius*)
	FED-53 (*Binder*)
	Innova S44i (*Eppendorf*)
	Multitron (*InforsHT*)
Microfluidizer	Emulsifier C5 (*Avestin*)
Microscopes	B601 (*Olympus*)
	MZ 8 (*Leica*)
Microwave	(*LG*)
MilliQ water dispenser	Seralpur Pro90CN (*Seralpur*)
Peristaltic pump	Pump drive 5201 (*Heidoph*)
pH meter	HI2020 edge (*Hanna Instruments*)
Pipets	Research variable 100 – 1000 µL (*Eppendorf*)

	Research variable 20 – 200 µL (*Eppendorf*)
	Research variable 10 – 100 µL (*Eppendorf*)
	Research variable 1 – 10 µL (*Eppendorf*)
	Research plus variable 0.1 – 2.5 µL (*Eppendorf*)
Power Boxes	EPS 301 (*Amersham Biosciences*)
Spectrometers	NanoDrop 800 Spectrophotometer (*Thermo Fisher*)
	OD 600 (*Implen*)
Spin concentrators	Amicon Ultra-15 (3 – 30 kDa MWCO) (*Millipore*)
Thermocycler	GeneAmp PCR System 2400 (*Perkin Elmer*)
	Rotor-Gene Q (*Qiagen*)
Thermomixer	Comfort (*Eppendorf*)
Waterbath	NK22 (*Haake*)
X-ray sources/beamlines	Beamlines ID23-1/2, ID29 (*ESRF, Grenoble*)
	Beamlines X06SA (PXI), X06DA (PXIII) (*SLS, Villigen*)

2. 3. Commercial kits, enzymes, and consumables

Crystallization and Fishing Equipment	EasyXtal 15-Well Tools (*Qiagen*)
	MRC 2 Well UVP (*Swissci*)
	VIEWseal (*Greiner BIOone*)
Crystallization Screens	NeXtal Tubes JCSG Core I Suite (*Qiagen*)
	NeXtal Tubes JCSG Core II Suite (*Qiagen*)
	NeXtal Tubes JCSG Core III Suite (*Qiagen*)
	NeXtal Tubes JCSG Core IV Suite (*Qiagen*)
	NeXtal Tubes AmSO4 Suite (*Qiagen*)
	NeXtal Tubes Classics Suite (*Qiagen*)
	Morpheus (*Molecular Dimensions*)
	Morpheus II (*Molecular Dimensions*)
Cuvettes (single use)	67.724 (*Sarstedt*)
DNA Ladder	1 kb DNA Ladder (*NEB*)
DNA-Ligase	T4 DNA Ligase (*NEB*)
DNA-Polymerase	Phusion Polymerase (2U/µL) (*NEB*)
	Phusion HF-Buffer (5x) (*NEB*)
Gel extraction kit	QIAquick Gel Extraction Kit (*Qiagen*)
Miniprep kit	QIAprep Spin Miniprep Kit (*Qiagen*)
PCR purification kit	QIAquick PCR Purification Kit (*Qiagen*)
Pipet tips	(*Sarstedt*)
Protein Ladder	Pierce Unstained Protein MW Marker (*Fermentas*)
Reaction tubes	(*Sarstedt*)
Restriction Enzymes	*Bam*HI (*NEB*)
	*Eco*RI-HF (*NEB*)

	*Hind*III-HF (*NEB*)
	*Nhe*I-HF (*NEB*)
	*Ssp*I-HF (*NEB*)
	CutSmart (10x) (*NEB*)
Sterile filters	Bottle-top filters (*Millipore*)
	Filtropur S 0.2 (*Sarstedt*)
	Filtropur S 0.45 (*Sarstedt*)
	Ultrafree-MC (*Millipore*)
Sypro Orange	SYPRO Orange Protein Gel Stain (*Thermo Fisher*)

2. 4. Oligonucleotides, vectors, and DNA

2. 4. 1. List of oligonucleotides used for gene amplification

Table 1: List of primers used for amplification (restriction sites underlined, overlaps used for LIC bold)

Name	Sequence (5' – 3')	Target
*Sc*Ecm33 21 - 360 fwd	CATGGCTAGCAACTCAACTACTTCTATTCCAT	pET-28a(+)
*Sc*Ecm33 21 - 360 rev	AGTAAGCTTTTACTTAACGGAGGTAGATGTGGCA	pET-28a(+)
*Sc*Pst1 20 - 357 fwd	AGCTGCTAGCGCTACTTCCTCTTCTTCCAGCAT	pET-28a(+)
*Sc*Pst1 20 - 357 rev	AGTGGATCCTTAGGATGATGCACCATTTTTGCA	pET-28a(+)
*Sc*Ecm33 35 - 148 fwd	ATAAGCTAGCACTTCTGCCACTGCTACTGCTCA	pET-28a(+)
*Sc*Ecm33 35 - 148 rev	AGTAAGCTTTTAGTCAGAAACAATAATGTTGTT	pET-28a(+)
*Ca*Pst1 25 - 351 fwd	ATAAGCTAGCAACAAATGTTCATTCTCTAAAACTT	pET-28a(+)
*Ca*Pst1 25 - 351 rev	AGTAAGCTTTTAATGAGTACAAACATAATTGTGACCT	pET-28a(+)
*Cg*Ecm33 21 - 357 fwd	ATAAGGATCC*ACATCTGACGATGTTCCATCTGGG*	pET-28a(+)
CgEcm33 21 - 357 rev	ATTAAGCTTTTAAGTAGCACCGTTCTTGCAGACGAA	pET-28a(+)
*Kp*Ecm33 35 - 360 fwd	TGCAGCTAGCATTTCAATTGCATCTGGATGTAGT	pET-28a(+)
*Kp*Ecm33 35 - 360 rev	AATGGATCCTTAAGCAGCAGAGCACTGATACTCA	pET-28a(+)
*Ca*Ecm33 32-360 fwd	ATGCGCTAGCAAATCTGAATGTTCATTCAAAGATTTC	pET-28a(+)
*Ca*Ecm33 32-360 rev	ATGCAAGCTTTTAGGTTTGTCTGTCTTCACATTGGAATT	pET-28a(+)
*Ca*Pst1 24-354 fwd	ATACGCTAGCTCAAACAAATGTTCATTCTCTAAA	pET-28a(+)
*Ca*Pst1 24-354 rev	AGTAAGCTTTTAATTAGCTGGATGAGTACAAACA	pET-28a(+)
*Sc*Ecm33 21-160 rev	AGTAAGCTTTTACAAAGTGGAGAAACCTTCGACACTT	pET-28a(+)
*Ct*Ecm33 fwd	CAGAGGATCCAGCTGCAAGGCGACGACGACGACT	pET-28a(+)
*Ct*Ecm33 rev	CAGTAAGCTTTTAGGCAGCAGCGTTGTCGCTCGTGCAG	pET-28a(+)
G0RYL2 fwd	ATGCGCTAGCACCGACTTCCCGCCCAACA	pET-28a(+)
G0RYL2 rev	AGCTGGATCCTTACGCAAGAATGCCACCGCAAAAGC	pET-28a(+)
G0S002 fwd	ATGCGCTAGCGAGGCTTCTTCTAGTGTCAG	pET-28a(+)
G0S002 rev	AGCTGGATCCTTAAGCCCACTTGCCGCAGATGCCCTG	pET-28a(+)
G0S3S8 fwd	ACGAGCTAGCGACGCCCAGCCCACTCTTCCT	pET-28a(+)
G0S3S8 rev	AGCTGGATCCTTAAGCAGTCGGCAGATCGCTCACTT	pET-28a(+)
G0S9T6 fwd	ACGAGCTAGCCAGTCTATTGACACCCTTGACCCCT	pET-28a(+)
G0S9T6 rev	AGCTGGATCCTTAAGCAGGGGAGGGAGCCGCAGTGA	pET-28a(+)

G0SBA5 fwd	ACGAGCTAGCAGCACCACTGCCACGGCTACCTC	pET-28a(+)
G0SBA5 rev	AGCTGGATCCTTAGGCCGGTGTGACGGCAACGCAAT	pET-28a(+)
G0SBE2 fwd	ACGAGCTAGCGTCGATGCCCCCGGATCGCTGTTGT	pET-28a(+)
G0SBE2 rev	AGCTGGATCCTTAGCTCTTTGGCGTGACACCGCACAT	pET-28a(+)
G0SDR6 fwd	ACGAGCTAGCGACCCAATTCCCTCTGCCGCGGT	pET-28a(+)
G0SDR6 rev	AGCTGGATCCTTAGTTGAGAACACAGTCGCAGACCTT	pET-28a(+)
Awp6 fwd	ATGCGGATCCATCGAACCAACAACCACGCTA	pET-28a(+)
Awp6 rev	ATCGGAATCCCTACCAGGCAGTAACAATACCTG	pET-28a(+)
*Sc*Ecm-LIC fwd	**TACTTCCAATCCAATGCA**AACTCAACTACTTCTATTCCAT	pET-LIC
*Sc*Ecm-LIC rev	**TTATCCACTTCCAATGTTATTA**CTTAACGGAGGTAGATGTGGCA	pET-LIC
*Ct*Ecm-LIC fwd	**TACTTCCAATCCAATGCA**AGCTGCAAGGCGACGACGACGA	pET-LIC
*Ct*Ecm-LIC rev	**TTATCCACTTCCAATGTTATTA**GGCAGCAGCGTTGTCGCTCGTGC	pET-LIC
Awp1 I165M fwd	AATACAGGCACAATGAATTACGAAAGT	SDM
Awp1 I165M rev	ACTTTCGTAATTCATTGTGCCTGTATT	SDM
Awp1 I285M fwd	ACACAGACAGGTATGCTTACTGTTACC	SDM
Awp1 I285M rev	GGTAACAGTAAGCATACCTGTCTGTGT	SDM

Genomic DNA (gDNA) from *S. cerevisiae* (*Sc*), *C. albicans* (*Ca*), *C. glabrata* (*Cg*) and *Komagataella phaffii* (*Kp*) were used as templates for amplification of the desired genes. gDNA is the complete chromosomal DNA of an organism, containing introns and exons. Primers were therefore designed with care to avoid introduction of noncoding sequences into the final expression construct.

As the thermophilic fungus *C. thermophilum* (*Ct*) contains a high number of introns, usage of gDNA as a template for gene amplification is not applicable. Therefore, complementary DNA (cDNA) of *C. thermophilum* was used in this work. The preparation of cDNA is achieved by isolation of the organism's complete RNA, which is subsequently amplified via Reverse Transcriptase (RT)-Polymerase chain reaction (PCR) using poly-A primers. In this step only the polyadenylated messenger RNA (mRNA) is amplified, thus cDNA only contains sequences of proteins that are transcribed. *C. thermophilum* cDNA used in this work was received from two sources: as a generous gift from Dr. Patrick Pausch and by isolation of cDNA, executed by Christin Schulz.

2. 4. 2. pET-28a(+)

The pET vectors are used for the recombinant overproduction of target proteins in *E. coli*. They were originally developed by Studier and Moffat[49] and can currently be acquired from *Novagen*. pET-28a(+) is a translation vector, accordingly no ribosome binding site needs to be inserted, but the vector contains the ribosome binding site from the phage T7 major capsid protein. Thus, combination with a suitable *E. coli* strain (a T7 expression host) is essential. The protein expression is also controlled by the lac operator, which facilitates induction of protein

expression by addition of lactose or its structural analogue IPTG to the cell's growth medium. pET-28a(+) also contains a kanamycin resistance cassette, allowing application of selective pressure by addition of kanamycin to the growth medium. The origin of replication (*ori*) ensures that the vector can be copied by the cell. With pET-28a(+) being a low copy plasmid, around 15 – 20 copies per cell are produced.

Target sequences are inserted into the multiple cloning site of the vector, which contains a variety of restriction enzyme target sites. This ensures that appropriate restriction enzymes can be chosen for cloning. An N-terminal His_6-Tag, followed by a thrombin cleavage site, and a C-terminal His_6-Tag are encoded next to the multiple cloning site and can be added to the target protein as desired.

In this work, an N-terminal His_6-Tag was added to target proteins that were cloned into pET28-28a(+). The plasmid map of pET28a_Awp1A is shown below as an example.

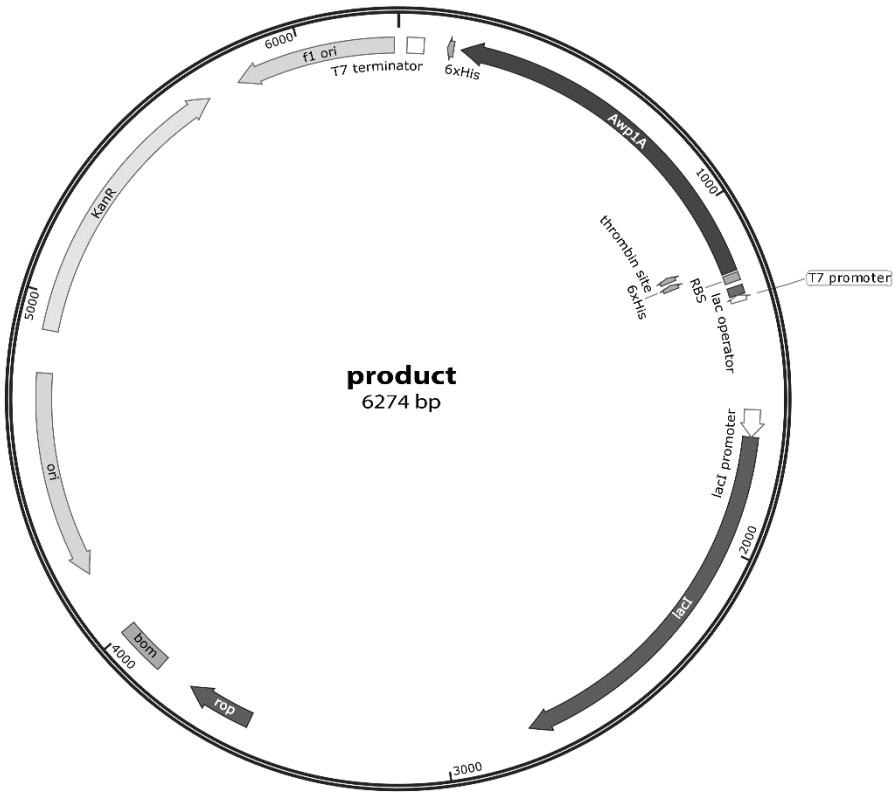


Figure 4: Visualization of pET28a_Awp1 as an example of a plasmid map created in this work

2. 4. 3. pET-vectors designed for Ligation Independent Cloning (LIC)

Vectors containing an N-terminally His_6 tagged solubility tag, followed by a TEV cleavage site and a LIC cloning site were acquired via *Addgene* from the Scott Gradia laboratory. Target proteins cloned into those vectors therefore have an N-terminal His_6-Tag enabling purification via IMAC, as well as a solubility tag, which can both be removed via cleavage with TEV protease. The LIC cloning site itself is the same in all three vectors, making the inserts compatible with each of them. Following *Addgene* vectors were used: pET His6 GST TEV LIC cloning vector (1G) (Plasmid #29655), pET His6 MBP TEV LIC cloning vector (1M) (Plasmid #29656), pET His6 Mocr TEV LIC cloning vector (1O) (Plasmid #29658).

Glutathione S-transferase (GST) and maltose binding protein (MBP) are commonly used solubility tags, which also facilitate binding to certain columns and can therefore be used for affinity purification. With a size of 13.8 kDa, monomeric Ocr (Mocr) is the smallest of those three tags and does not confer binding to a specific column matrix[50]. Thus, the N-terminal His_6-Tag encoded on the LIC vector is indispensable for affinity purification in this construct.

2. 4. 5. Plasmids used in this work

Table 2: List of plasmids that were used in this work

Name	Comments
pET28a_*Sc*Ecm33	
pET28a_CgEcm33	
pET28a_*Kp*Ecm33	
pET28a_*Ca*Ecm33	
pET28a_*Ca*Pst1	
pET28a_*Ct*Ecm33	
pET28a_G0SBA5	
pET28a_G0S9T6	Pga7
pET28a_G0SBE2	Pth11
pET28a_*Ct*Pth11 36-101	received by Dr. Vitali Kalugin
pET28a_*Ct*Mad1 391-453	received by Dr. Vitali Kalugin
pET28a_Awp1	
pRSETa_Awp1	received by Dr. Piet de Groot
pET28b_Awp2	received by Dr. Piet de Groot
pET28a_Awp3	
pRSETa_Awp3	received by Dr. Piet de Groot
pET28b_Awp4	received by Dr. Piet de Groot
pRSETa_Awp2	received by Dr. Piet de Groot
pRSETa_Awp4	received by Dr. Piet de Groot
pRSETa_Awp5	received by Dr. Piet de Groot
pET28a_Awp5	
pRSETa_Awp6	received by Dr. Piet de Groot

pRSETa_Awp7	received by Dr. Piet de Groot
pET28a_Awp7	
pET28a_Awp8	received by Dr. Piet de Groot
pET28a_Awp9	received by Dr. Piet de Groot
pET28a_Awp10	received by Dr. Piet de Groot
pRSETa_Awp12	received by Dr. Piet de Groot
pET28a_Awp12	
pET28a_Awp13	received by Dr. Piet de Groot
pET28b_Awp14	received by Dr. Piet de Groot
pET28a_Awp6	
pET28a_Awp1 I165M I285M 'SeMet'	
Mocr-pET_*Sc*Ecm33	
Mocr-pET_Awp2	
Mocr-pET_Awp4	
Mocr-pET_Awp9	
MBP-pET_*Sc*Ecm33	
MBP-pET_*Ct*Ecm33	
MBP-pET_Awp2	
MBP-pET_Awp4	
MBP-pET_Awp9	
MBP-pET_Awp10	
GST-pET_*Sc*Ecm33	
GST-pET_*Ct*Ecm33	
GST-pET_Awp2	
GST-pET_Awp4	
GST-pET_Awp8	
GST-pET_Awp9	
GST-pET_Awp10	
pBC542_empty	received by Dr. Piet de Groot
pEH070_Awp3	received by Dr. Piet de Groot

Numerous plasmids were created for the overproduction of fungal proteins in *E. coli*. Plasmids that resulted in successful production and purification of the protein are written in bold, yeast plasmids in italics. The sequences of all plasmids used in this work were verified via sequencing.

2. 5. Organisms

2. 5. 1. Escherichia coli DH5α

Genotype: F- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_{k-}, m_{k+}) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ-

E. coli DH5α (*Invitrogen*) have a high plasmid replication rate. Accordingly, the strain is well suited for the production of plasmids. Accordingly, chemically competent *E. coli* DH5α were used for this purpose.

2. 5. 2. Escherichia coli BL21 (DE3) Gold

Genotype: *F^- omp*T *gal dcm lon hsd*S_B(r_B-m_B-) λ*(DE3 [lacI lac UV5-T7p07 ind1 sam7 nin5]) [malB$^+$]K-12(λ^S)*

E. coli BL21(DE3) Gold (*Invitrogen*) is one of the standard strains used for heterologous production of proteins using the T7 expression system. Chemically competent cells from this strain were used for production of proteins that do not contain any disulfide bonds.

2. 5. 3. Escherichia coli SHuffle T7 Express

Genotype: *fhuA2 lacZ::T7 gene1* [lon] *ompT ahpC gal λatt::*pNEB3-r1-*cDsbC* (Spec^R, *lacI^q*) Δ*trxB sulA11 R(mcr-73::miniTn10*--Tet^S)2 [dcm] *R(zgb-210::Tn10* --Tet^S) *endA1 Δgor Δ(mcrC-mrr)114::IS10*
E. coli SHuffle T7 Express (*Invitrogen*) is a strain designed for heterologous production of proteins containing disulfide bonds using the T7 expression system. Disulfide bond formation is enabled by the deletion of *gor* and *trxB* and introduction of the disulfide isomerase DsbC.

2. 5. 4. Chaetomium thermophilum DMSZ No.: 1495

The strain *C. thermophilum* var. *thermophilum* La Touche 1950 (DMS No.: 1495), originally isolated from wheat straw compost in the UK, was used in this work. Fungal spores (dried on a filter paper) and cultivation protocols were kindly provided by the group of Prof. Dr. Ed Hurt (Heidelberg University Biochemistry Center).

2. 6. Software and Algorithms

Software or algorithm	**Version (if applicable)**
CCP4i and *CCP4i2* software suite[51]	7.0.067
PHENIX suite[52]	1.14-3260
WinCoot[53]	0.8.9
XDS[54]	
ARP/wARP Webservice[55]	8.0
Cytoscape[56]	3.7.1
PyMOL	4.5.0
BLAST[57]	
Clustal Omega[58]	
ProtParam[59]	

3. Methods

3. 1. Bioinformatics methods

Bioinformatics has become an essential tool for practice in biological sciences. It is generally understood to be the application of information techniques for organization of biological information and understanding it. The field of bioinformatics includes the storage and retrieval of information from databases, as well as providing effective ways to computationally analyze this information or to carry out predictions[60].

Bioinformatics applications are continuously updated and enhanced, so it is hardly possible to keep track of all the latest advancements. Nevertheless, a basic understanding of the algorithms commonly used in these applications is beneficial for understanding the results and limitations of an application.

3. 1. 1. Prediction of GPI-anchored proteins in C. thermophilum

The prediction of the GPI-anchored proteins in *C. thermophilum* depends on three specific characteristics: firstly, GPI-anchored proteins contain an N-terminal signal peptide, which targets them to the ER, where the GPI-anchor is attached to the protein. Secondly, they do not contain any transmembrane helices. Lastly, the GPI anchor attachment sequence has characteristic features and can therefore be recognized[9]. The workflow used here was done together with Dr. Piet de Groot and has already been described in 2003[11].

The sequences of all proteins included in version 3.0 of the *C. thermophilum* genome – which was the newest version of the genome available at the time of the analysis – were retrieved from the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) database.

The presence of the N-terminal signal peptide was analyzed using SignalP 5.0 with Eukarya set as an organism group. SignalP 5.0 uses a machine learning approach to recognize signal peptides, applying a deep artificial neural network of the recurrent type[61]. Artificial neural networks are widely used for many different applications. They consist of several layers of nodes or “neurons”, where each neuron in a layer is connected to each neuron of the next layer. The connections between the neurons propagate information from one layer to the next one via a propagation function, which assigns a certain weight to a connection that is descriptive for the relative importance. A learning process is used to define the weights of the connections. There is a wide variety of neural network architectures (see https://www.asimovinstitute.org/neural-network-zoo/). In recurrent networks, such as the one used in SignalP 5.0, certain layers do not only obtain information from the previous layer, they also feed on previous information from themselves. Additionally, the implementation of

long/short term memory enables memorizing features from the beginning of a sequence, while already classifying positions further downstream[62].

At this point it has to be noted that the presence or absence of a signal peptide does not equal secretion of a protein or no secretion. Few proteins are secreted without signal peptides and a few have a signal peptide, but are not secreted[61].

Protein sequences, in which a signal peptide was detected by SignalP 5.0 were further analyzed for absence of transmembrane helices using TMHMM v. 2.0. As indicated by the name, a hidden Markov model (HMM) – an algorithm well suited for pattern detection – is used for identification of potential transmembrane helices[63]. As the GPI anchor attachment sequence is usually recognized as a transmembrane helix, C-termini of the proteins were ignored in the prediction.

Protein sequences were then further analyzed for presence of a GPI anchor attachment sequence using the Big-PI Fungal Predictor (http://mendel.imp.ac.at/gpi/fungi_server.html)[12]. In addition, a pattern search was applied for identification of GPI-anchored proteins, using the following pattern: [NSGDAC]–[GASVIETKDLF]–[GASV]–X(4,19)–[FILMVAGPSTCYWN](10)>[11].

3. 2. Cell wall extraction and analysis

3. 2. 1. Cultivation of C. thermophilum

C. thermophilum (DMSZ No.: 1495) spores were received as a kind gift by the group of Dr. Ed Hurt (Heidelberg University Biochemistry Center). To reactivate spores on a filter paper 50 µL CCM medium (composition described below) were pipetted onto the paper, followed by incubation for 10 min. The filter paper was then laid onto a CCM agar plate (spore side down), which was put into a plastic bag together with a wet towel, sealed tightly and incubated at 54 °C for 2 days. Subsequently, half a plate was used to inoculate 150 mL CCM medium. Therefore, mycelium was cut into small pieces and as much agar as possible was removed. Liquid cultures were incubated at 54 °C, 100 rpm, for 1 day. Mycelium was then either harvested or used for production of new spores. For harvesting, cells were strained through a gauze, then washed with deionized water. The mycelium was then dried by pressing it between some sheets of paper towel, frozen in liquid nitrogen and stored at -80 °C.

Spores were grown on rice agar, which was produced by cooking 75 g of brown rice for 2 h in 1 L water. 15 g agar were added, the rice broth was filtered through a sieve to remove the rice seeds and the volume was refilled to 1 L. Rice agar was filled into beakers (50 mL each) and autoclaved. 50 mL rice agar were then inoculated with 2 mL mycelium grown in a liquid culture, closed tightly and incubated at 37 °C until black spores could be seen on the surface of the agar (at least 7 days). Spores were harvested in 1 M sterile sorbitol by scratching the agar surface with a sterile spatula. The presence of spores in the solution was verified by microscopy. Spore aliquots were then frozen in liquid nitrogen and stored at – 80 °C.

For the proteomic analyses of the *C. thermophilum* cell wall, 250 µL spore solution were used to directly inoculate 150 mL liquid CCM medium. Cultures were incubated at 54 °C, 100 rpm, for 2 days, then harvested and either directly used for cell wall isolation or stored as described above.

CCM medium	
Sucrose	3 g/L
NaCl	0,5 g/L
$K_2HPO_4 \cdot 3\ H_2O$	0,65 g/L
$MgSO_4 \cdot 7\ H_2O$	0,5 g/L
Fe(III)sulfate-hydrate	0,01 g/L
Tryptone	5 g/L
Peptone	1 g/L
Yeast extract	1 g/L
Dextrine (potatoe) (dissolved in ¼ of the final volume, heated, then added to the medium)	15 g/L
Agar added for plates	20 g/L

3. 2. 2. Cell wall isolation

Different approaches can be used for the isolation of certain components of the fungal cell wall, depending on the intended purpose of the experiments. For example, the exposed surface proteins of a cell can be identified by digestion of living cells using proteases, followed by identification of the released peptides via mass spectrometry (MS)[64]. Obviously, cell surface "shaving" does not yield in a complete picture of the cell wall proteome, as some proteins are not sufficiently exposed to the surface or not digested by the protease for other reasons e. g. heavy glycosylation[65]. In this work, the cell wall material was isolated from broken cells to achieve determination of the cell wall proteome of *C. thermophilum*. The workflow used was also described by de Groot *et al.*[66].

C. thermophilum mycelium was resuspended in 10 mM Tris-HCl, pH 7.5 and divided into 2 mL screw-cap cups. Glass beads and 10 µL protease inhibitor (cOmplete™ Protease Inhibitor Cocktail, *Roche*) were added. Cells were then lyzed in a FastPrep Homogenizer (*MPBio*) for 60 sec, at a speed of 6.5 m/s. Cell lysis was repeated until full breakage of the cells could be observed under the microscope; the samples were kept on ice for 5 min after each run. The lysate was then extensively washed with 1 M NaCl to remove intracellular contaminants. Additionally, the glass beads were removed in this step. Subsequently, 0.5 mL SDS extraction buffer (50 mM Tris-HCl, 100 mM EDTA, 150 mM NaCl, 2% SDS, pH 7.8) per 100 mg wet weight cell walls were added, as well as 8 µL β-mercaptoethanol per mL of extraction buffer. The extraction was done by incubation in a boiling water bath for 10 min; then the cell wall material was pelleted 5 min at 1800 g, the supernatant was removed and the extraction step

repeated. The treatment of the cell wall material with denaturing and reducing agents is intended to remove proteins that are not covalently incorporated into the cell wall[65]. The isolated cell walls were then washed with ddH_2O by centrifugation at 1800 g for 5 min, until SDS was fully removed. Complete removal of SDS was assessed by the absence of foam formation. The cell walls were freeze dried and stored at -20 °C.

3. 2. 3. Mass-spectrometric analysis of isolated cell walls

The proteomic analysis of isolated cell walls was done in the *MarMass* facility for MS. The analysis protocol was outlined with Dr. Uwe Linne.

The isolated cell walls were resuspended in Urea and proteins were digested by addition of Sequencing Grad Modified Trypsin (*Serva*) and incubated at 37 °C overnight. Peptides were desalted and concentrated using Chromabond C18WP spin columns (*Macherey-Nagel*, Part No. 730522). Finally, Peptides were dissolved in 25 µL of water with 5% acetonitrile and 0.1% formic acid.

The mass spectrometric analysis of the samples was performed using an Orbitrap Velos Pro mass spectrometer (*Thermo Scientific*). An Ultimate nanoRSLC-HPLC system (*Dionex*), equipped with a custom end-fritted 50cm x 75µm C18 RP column filled with 2.4 µm beads (*Dr. Maisch*) was connected online to the mass spectrometer through a Proxeon nanospray source. 1-15 µL (depending on peptide concentration and sample complexity) of the tryptic digest were injected onto a 300µm ID x 1cm C18 PepMap pre-concentration column (*Thermo Scientific*). Automated trapping and desalting of the sample was performed at a flowrate of 6 µL/min using water/0.05% formic acid as solvent.

Separation of the tryptic peptides was achieved with the following gradient of water/0.05% formic acid (solvent A) and 80% acetonitrile/0.045% formic acid (solvent B) at a flow rate of 300 nL/min: holding 4% B for five minutes, followed by a linear gradient to 45%B within 30 minutes and linear increase to 95% solvent B in additional 5 minutes. The column was connected to a stainless steel nanoemitter (*Proxeon*, Denmark) and the eluent was sprayed directly towards the heated capillary of the mass spectrometer using a potential of 2300 V. A survey scan with a resolution of 60000 within the Orbitrap mass analyzer was combined with at least three data-dependent MS/MS scans with dynamic exclusion for 30 s either using CID with the linear ion-trap or using HCD combined with Orbitrap detection at a resolution of 7500.

Data analysis was performed with Proteome Discoverer 2.4 (*Thermo Scientific*) with *SEQUEST* as search engine. The search libraries used were the proteome translated from the *C. thermophilum* genome v 3.0 (downloaded from NCBI) and a list of common contaminants found in proteome analysis (provided by the *MarMass* facility). Sequence coverage, number of identified peptides, number of unique peptides and Sequest HT score were used to assess

the quality of the results. Particularly the Sequest HT score was used for the evaluation, and identified proteins with a score below 40 were not included in the further analysis. Finally, the identified proteins were sorted manually: Contaminants from other cellular components (e.g. cytosol or plasma membrane) were removed from the list of GPI-CWPs and the function of each identified protein was assigned by database analysis and other sequence analysis methods.

3. 2. 4. Imaging of C. thermophilum cell walls via Transmission Electron Microscopy (TEM)

Transmission electron microscopy (TEM) imaging was used to reveal the cell wall structure of *C. thermophilum*.

A TEM consists of an electron optical column, a vacuum pump and a sample chamber. The electron optical column is kept under vacuum; it contains the electron source ("electron gun"), a lens system and a detector. An electron beam is generated by applying heat or a strong electric field to a cathode in the electron gun (a tungsten filament or LaB_6 cathode). The gun also contains an anode, which is a disc with an axial hole. The electrons emerging from the cathode are accelerated towards the anode and pass through the central hole at constant energy. The energy of the electrons can be controlled by the voltage (often 80 kV - 200 kV) applied on the cathode. The electron beam then passes a lens system with magnetic lenses inside the electron optical column. The energy and speed of the electrons remain unchanged as they pass through the column; only the path is adjusted to focus the beam on the sample, which is usually an ultrathin section (less than 100 nm thick) of the specimen. An image can be obtained, because electrons are scattered when they hit an atomic nucleus (elastic scattering). On leaving the sample, diffracted electrons are shielded by the contrast aperture and cannot reach the detector. Visualization is often realized by a fluorescent screen placed at the base of the column; charge-coupled device (CCD) cameras are widely used to capture images[67].

Well-grown mycelium was used for recording TEM images of *C. thermophilum*. Fixation, embedding, microtomy, and imaging were done by Dr. Thomas Heimerl from the Synmikro Electron Microscopy Facility.

3. 3. Molecular Biology Methods

3. 3. 1. Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is one of the standard methods in molecular biology. It constitutes an *in vitro* method for amplification of specific nucleic acid sequences that has first been described by Mullis *et al.* in 1983.

It consists of three steps that are repeated in cycles: denaturation, annealing, and elongation. In the first step, the template double strand is split into two single strands by heat. Then, primers anneal to the flanking regions of the DNA sequence to be amplified. In the elongation step, a heat stable polymerase synthesizes the missing complementary strand. The annealed primers serve as the starting points for elongation. The execution of these steps in cycles leads to an exponential amplification of the desired DNA product, as long as the polymerase is still intact and required components are sufficiently available. Usually, 25 to 30 cycles are performed[68].

Experimental parameters for PCRs done in this work are shown below.

	Volume or weight
Template DNA	20 – 50 ng
Forward primer (5 µM)	2.5 µL
Reverse primer (5 µM)	2.5 µL
Phusion HF-Buffer (5x)	10.0 µL
Phusion Polymerase (2 U/µL)	0.5 µL
dNTPs (10 mM each)	1.0
ddH_2O	Ad 50 µL

	Temperature	**Duration**	
Initial Denaturation	98 °C	5 min	
Denaturation	98 °C	15 sec	35 x
Annealing	55 – 58 °C	20 sec	35 x
Elongation	72 °C	15 – 30 s/kbp	35 x
Final Elongation	72 °C	5 min	
Cool down	4 °C	∞	

3. 3. 2. Agarose gel electrophoresis

Agarose gel electrophoresis is used for separation of nucleic acids based on the size of the molecules, using an electric field. The agarose gel provides a matrix, through which smaller molecules migrate faster than larger ones. It is covered by a conductive buffer and an electric field is applied, causing the negatively charged nucleic acid samples to migrate from the cathode to the anode.

1% agarose gel were used for analysis of DNA fragments, such as PCR products. The 1% gel consists of 0.65 g agarose, which were dissolved in 70 mL TBE buffer (0.1 M Tris, 0.1 M boric acid, 2 mM EDTA) by boiling in a microwave. The gel was allowed to cool to approximately 55 °C, then 2.5 µL Midori green were added and the gel was poured. When the gel was completely solidified, 5 µL sample for an analytical run or 45 µL sample for a preparative run were applied. The gel was then run at 120 V for 1 h and finally examined on an imager under UV illumination.

3. 3. 3. PCR purification and gel extraction

To ensure that further work with the amplified DNA fragments is not disturbed by contaminations – like nucleotides, primers, enzymes or salts – the PCR products were purified. Kits designed for this purpose are sold by many manufacturers, in this work the QIAquick PCR Purification Kit (*Qiagen*) was used. A buffer containing isopropanol and guanidine hydrochloride (PB buffer in the kit) is added to the PCR product. The mixture is then applied on a silica matrix, which binds the DNA in the presence of chaotropic agents. An ethanol containing buffer (PB) is then used to remove nucleotides, primers, enzymes, and salts. Finally, the PCR products can be eluted using a low salt buffer (EB) or water[69].

When specific DNA fragments needed to be extracted from a preparative agarose gel, the QIAquick Gel Extraction Kit (*Qiagen*) was used. First, the desired DNA fragment was carefully excised, removing as much agarose gel around the band as possible, while keeping the UV exposure time short. The gel piece is dissolved in a guanidine thiocyanate-containing buffer (QG) at 50 °C, followed by addition of isopropanol. The sample was then applied to the column, washing and elution are done in the same way as for PCR purification.

Detailed protocols for performing PCR purifications or gel extractions can be found in the manufacturer's manual.

3. 3. 4. DNA-modification: digestion and ligation

In a standard cloning procedure, as performed in this thesis, the insertion of a target gene into the desired vector requires certain manipulations of the PCR product and the vector: First, both are cut with restriction enzymes that produce *stick ends*. Afterwards, insert and vector can be combined using a DNA ligase.

Specific enzymes serve as tools for these modifications. Endonucleases are able to cleave the phosphodiester bonds of the DNA, either non-specifically or at specific sites called restriction sites. Such restriction sites typically consist of a palindromic sequence of 4 to 8 bp. Depending on the restriction enzyme used, either *sticky ends* (where one DNA strand has a short overhang compared to the other one, e. g. after restriction with *Eco*RI) or *blunt ends* (without an overhang, e. g. after restriction with *Ssp*I) are obtained. Usually, both the insert and the

vector are cut with the same restriction enzyme that produces a *sticky end*. The overhangs then have a complementary sequence, giving a specific direction for the introduction of the insert. The DNA molecules are then combined using a DNA ligase, which is capable of forming a phosphodiester bond between the 5'-phosphate and the 3'-OH group of adjacent nucleotides.

A typical restriction digest is done at 37 °C (depending on the enzymes used) for around 2 h, followed by heat inactivation at either 65 °C or 80 °C, depending on the restriction enzyme used. The restriction digest has following composition:

	Volume or weight
CutSmart buffer (10x)	2 µL
Restriction enzyme 1	1 µL
Restriction enzyme 2	1 µL
DNA	1 µg
ddH_2O	ad 20 µL

Ligation was usually done overnight at 16 °C, using T4 DNA ligase. Since the success of a ligation depends partially on the insert/vector ratio, a molar ratio of 3:1 was aimed for. The composition of the ligation mix is as follows:

	Volume or weight
Vector	50 ng
Insert	x
T4 DNA Ligase	1 µL
Ligase buffer (10x)	2 µL
ddH_2O	ad 20 µL

3. 3. 5. Preparation of competent cells and plasmid transformation

Competence is defined as the ability of a cell to take up DNA from its surrounding. Some bacteria (e. g. *Bacillus subtilis*) are naturally competent, others are made competent by enhancing their membrane permeability, either physically (by electroporation) or chemically (by salt treatment, followed by a heat shock)[70,71]. *E. coli* is not a naturally competent organism; chemically competent *E. coli* cells were used in this work.

For the preparation of competent *E. coli*, cells from a glycerol stock (as supplied by the manufacturer) were plated onto an LB-agar plate, which was incubated overnight at 37 °C. A single colony was used to inoculate a 5 mL preculture; the preculture was incubated overnight at 37 °C, shaking. 1 mL from the preculture was transferred to 50 mL LB-medium; the cells were grown at 37 °C, 225 rpm, until an OD_{600} or 0.5 – 0.6 was reached. The cells were then harvested by centrifugation at 3200 g, 4 °C, for 15 min. The supernatant was carefully removed; the pellet was resuspended in 15 mL sterile TBF-I buffer on ice. The cells were then pelleted again and the cell pellet was resuspended in 2 mL TFB-II buffer. 50 µL aliquots were

prepared and rapidly frozen in liquid nitrogen. The competent cells were then stored at – 80 °C for further use.

TBF-I	
Rubidium chloride (RbCl)	100 mM
Manganese(II) chloride ($MnCl_2$)	50 mM
Potassium acetate (CH_3COOK)	30 mM
Calcium chloride dihydrate ($CaCl_2 \cdot 2\ H_2O$)	10 mM
Glycerol	15% (v/v)

TBF-II	
Rubidium chloride (RbCl)	10 mM
Calcium chloride dihydrate ($CaCl_2 \cdot 2\ H_2O$)	10 mM
MOPS	10 mM
Glycerol	15% (v/v)

For the transformation of plasmids into competent *E. coli*, a 50 µL aliquot was thawed on ice and approximately 50 ng DNA were added. Transformation of a ligation was done with 10 µL ligation mix. Cells were then incubated on ice for 30 min, subjected to a 45 sec heat shock at 42 °C and cooled on ice for approximately 2 min. 1 mL LB medium was added and competent *E. coli* were allowed to recover at 37 °C for around 1 h. Recovered cells were pelleted at 3500 g for 2 min, 900 µL supernatant were removed and the pellet was resuspended in the remaining LB medium. The cells were then plated onto an LB agar plate containing the appropriate antibiotics and incubated overnight at 37 °C.

LB medium	
Tryptone	10 g/L
Yeast extract	5 g/L
NaCl	10 g/L
NaOH (10 M)	400 µL/L

LB agar	
Tryptone	10 g/L
Yeast extract	5 g/L
NaCl	10 g/L
NaOH (10 M)	400 µL/L
Agar	15 g/L

3. 3. 6. Plasmid preparation

Plasmids are – usually circular – DNA molecules within a cell that do not belong to the chromosome of the organism – i. e. they are extrachromosomal. After transformation they remain in the cytosol of *E. coli*, as long as a selective pressure is present. In this work, plasmids providing resistance against certain antibiotics, in most cases kanamycin, were used. Additionally, they are replicated in *E. coli*, which makes the plasmid preparation a convenient tool for multiplication of desired plasmids[72].

5 mL LB medium containing the appropriate antibiotics were inoculated with a single colony from a transformation plate and incubated overnight at 37 °C, 225 rpm. The plasmid preparation was then performed according to the manual provided with the QIAprep Spin Miniprep Kit (*Qiagen*), which is based on the alkaline extraction procedure, described by Birnboim and Doly[73]. In brief, the cells were pelleted and then thoroughly resuspended in a buffer containing EDTA and RNaseA (P1). Alkaline lysis was then achieved by addition of buffer P2, which consists of NaOH and SDS and also serves the denaturation of proteins and high molecular weight DNA. Addition of a third buffer (N3) then leads to neutralization of the solution by potassium acetate and facilitation of DNA binding to the silica matrix by guanidine hydrochloride. The mixture is centrifuged at 17000 g for 10 min, leaving the plasmid DNA in the supernatant. The supernatant is applied to a silica matrix, washed with an ethanol containing buffer and finally eluted in elution buffer (EB) or water.

A detailed protocol for performing the plasmid preparation can be found in the manufacturer's manual.

3. 3. 7. Site-directed mutagenesis (SDM)

Site-directed mutagenesis (SDM) is a method for introduction of specific changes in the nucleotide sequence of a plasmid. The mutations are introduced during a PCR, in which forward and reverse primers completely overlap and the desired nucleotide change is located in their center. The entire plasmid is copied in the elongation step of the PCR, with the primers used as the starting point. Copies of the plasmid therefore contain the mutation and serve as templates for the following rounds of PCR (in addition to the original template). To eliminate plasmids that do not contain the desired mutation, the PCR mix is digested with *Dpn*I, a restriction endonuclease that degrades methylated DNA[74].

A double mutant of Awp1 (I165M, I285M) was created for attempts to determine the structure. The mutations were introduced in succession via SDM, using a protocol based on the one described by Bachman[74]. Composition of PCR mixture and the thermocycler program are shown below.

	Volume or weight
Template Plasmid	10 ng
Forward primer (5 µM)	1 µL
Reverse primer (5 µM)	1 µL
Phusion HF-Buffer (5x)	10.0 µL
Phusion Polymerase (2 U/µL)	0.5 µL
dNTPs (10 mM each)	1.0
ddH_2O	Ad 50 µL

	Temperature	Duration	
Initial Denaturation	98 °C	5 min	
Denaturation	98 °C	30 sec	18 x
Annealing	55 °C	30 sec	
Elongation	72 °C	210 sec	
Final Elongation	72 °C	5 min	
Cool down	4 °C	∞	

After completion of the PCR, 1 µL *Dpn*I was added to the mixture and digestion was performed for 1 h at 37 °C, followed by heat inactivation for 20 min at 60 °C. The plasmids were purified using the QIAquick PCR Purification Kit (*Qiagen*), analyzed via agarose gel electrophoresis and transformed into *E. coli* DH5α. Success of the SDM was assessed by sequencing.

3. 3. 8. Ligation-Independent Cloning (LIC)

In Ligation/Ligase-Independent Cloning (LIC), the 3' – 5' exonuclease activity of T4 DNA Polymerase utilized to generate an overlap of around 15 base pairs between vector and insert. The overlap is created by addition of dCTP/dGTP to the insert/linearized vector. The addition of dCTP leads to single strand digestion of the blunt ends, until a C is reached; upon addition of dGTP, the digestion is stopped at a G. The resulting single stranded overlaps then enable the integration of the insert into the vector without the help of ligase. The remaining nicks are repaired in *E. coli* after transformation. LIC takes less time than the classical cloning procedure and a variety of vectors containing the same overlap sequence are available, making the inserts compatible with different vectors. However, only vectors that have been designed for LIC can be used[75].

The inserts were amplified from already existing plasmids, using a standard PCR as described in chapter 3. 3. 1. The PCR products were then purified using the QIAquick PCR Purification Kit (*Qiagen*). Vectors were linearized by digestion with the blunt end creating restriction enzyme *Ssp*I for 3 h at 37 °C (reaction mixture shown below), followed by heat inactivation at 65 °C for 20 min.

	Volume or weight
CutSmart buffer (10x)	2 µL
*Ssp*I-HF	1 µL
vector	1 µg
ddH_2O	ad 20 µL

The linearized vectors were then purified via preparative agarose gel electrophoresis (see 3. 3. 2.). A LIC reaction was done for both, inserts and vectors, at 22 °C for 40 min, using following reaction mixtures:

	Volume or weight
Linearized vector	x
CutSmart buffer (10x)	2 µL
dGTP (25 mM)	2 µL
DTT (100 mM)	1 µL
T4 DNA Polymerase(3 000 U/mL)	0.2 µL
ddH_2O	ad 20 µL

	Volume or weight
Insert	y
CutSmart buffer (10x)	2 µL
dCTP (25 mM)	2 µL
DTT (100 mM)	1 µL
T4 DNA Polymerase(3 000 U/mL)	0.2 µL
ddH_2O	ad 20 µL

The LIC reaction was stopped by heat inactivation at 75 °C for 20 min. For annealing, equivalent amounts of insert and vector were mixed (total 8 µL), and the reaction volume was filled up to 20 µL with water. After a 30-minute incubation at room temperature, the annealing mixture was transformed into *E. coli* DH5α.

3. 4. Protein biochemistry

3. 4. 1. Analytical overproduction of proteins and cell lysis

To evaluate whether a protein can be produced in *E. coli* – in the best case soluble and in large amounts – overexpression was performed on an analytical scale. In this way, different factors that influence protein overproduction were evaluated, such as the strain of *E. coli* used, expression temperatures and durations, or type and concentration of the inducing agent.

For each overexpression condition that was tested, plasmids were transformed into various *E. coli* strains via heat-shock transformation as described in chapter 3. 3. 5. Cells were then plated onto LB agar plates containing the appropriate antibiotic; plates were incubated overnight at 37 °C. One colony was used to inoculate a 5 mL overnight culture with antibiotics, which was incubated at 37 °C, 225 rpm, overnight. Then, 50 mL LB containing the appropriate antibiotics were inoculated 1:50 and the cells were grown at 37 °C, 225 rpm, until an OD_{600} of approximately 0.6 was reached. Expression was induced by addition of either IPTG or lactose and the cultures were further incubated at different expression temperatures and durations. Parameters that were tested in small scale expression are summarized below.

***E. coli* strains**	**Inducing agents**	**Temperature/Duration**	
BL21 (DE3) Gold	IPTG	37 °C	3 h
SHuffle T7 Express	Lactose	30 °C	Overnight
BL21 Star (DE3)		18 °C	48 h
Rosetta		12 °C	72 h
Origami			

When the analytical overexpression was finished, the cells were pelleted by centrifugation at 3200 g, 4 °C, for 20 min. The cell pellets were resuspended in Ni-NTA buffer 1, transferred into screw-cap cups, 1 µL lysozyme (50 mM) and glass beads were added. Cell lysis was done in a FastPrep Homogenizer (*MPBio*), run twice for 60 sec at 6.5 m/s; between the runs the cells were cooled on ice for 5 min. To divide the soluble and the insoluble fraction, the lysed cells were centrifuged for 10 min at 17000 g, 4 °C. The supernatant was removed and the pellet was resuspended in 1 mL Ni-NTA buffer 1. Both were analyzed for presence of the desired protein via SDS-PAGE.

Expression conditions that were proven to produce soluble protein in analytical scale overexpression were upscaled. The description of those conditions can be found in chapter 3. 4. 2.

Ni-NTA buffer 1

NaH_2PO_4	50 mM
NaCl	300 mM
	pH 8.0

3. 4. 2. Preparative overexpression of proteins

For preparative overexpression of a desired protein, plasmids were transformed into E. coli and a colony from the transformation plate was used to inoculate 50 mL LB with 35 µg/mL kanamycin (Kan^{35}). The 50 mL starter culture was incubated at 37 °C, 225 rpm, overnight. Several 5 L baffled Erlenmeyer flasks containing 2 L LB+Kan^{35} were inoculated with starter culture (ratio 1:100) and cells were grown at 37 °C, 140 rpm, until an OD_{600} of around 0.6 was reached. Expression was induced by addition of 0.1 mM IPTG and the incubation temperature was lowered. The expression conditions for the constructs used in this work are as follows:

Construct	***E. coli* strain**	**Temperature/ Duration**
pET28a_Awp1	SHuffle T7 Express	12 °C/72 h
pET28a_Awp3	SHuffle T7 Express	12 °C/72 h
pET28a_Awp14	BL21 (DE3) Gold	12 °C/72 h
pET28a_*Ct*Pth11	SHuffle T7 Express	18 °C/48 h

The cells were then harvested by centrifugation at 3200 g, at 4 °C, for 20 min. The cell pellets were resuspended in Ni-NTA buffer 1 and washed by centrifugation at 4000 rpm, 4 °C, 20 min. The supernatant was removed; the pellets were stored at -80 °C for further use.

3. 4. 3. Cell lysis

The cells were lysed mechanically, either by subjecting them to high pressure using an emulsifier, or by sonication with ultrasound. For both methods, the cell pellet was thawed in a water bath at room temperature and then resuspended in Ni-NTA buffer 1. Complete resuspension of the cells is critical, as clumps may remain unbroken during sonication or clog the emulsifier tubes. In addition, mechanical cell lysis is associated with the generation of heat, and many proteins are sensitive to heat. To avoid excessive thermal effects on the proteins, cooling of the lysate is essential.

Cell lysis by sonication was performed with the cells kept on ice. The resuspended cells were sonicated for a total of 9 minutes, divided into 3 cycles, applying pulses with 50% intensity. Between the cycles the cell lysate was mixed and cooled on ice for 5 minutes. When the emulsiflex C5 (*Avestin*) was used for cell lysis, the equipment was first pre-cooled with ice for around 30 min. Then the cell suspension was passed 3 times through the emulsifier, applying pressures between 50 000 and 100 000 kPa.

The lysate was then cleared by centrifugation at 18000 rpm, 4 °C, for 30 min (J2-HS, *Beckman*), and the supernatant was sterile-filtered using a 0.45 µm syringe filter.

3. 4. 4. Protein purification

In order to perform protein analysis techniques, a certain purity level of the desired protein must be achieved, with the necessary degree of purity depending on the technique. Crystallographic studies in particular require a highly pure sample (> 95%). In addition, the homogeneity and monodispersity of the desired protein should be ensured, which demands a combination of several purification steps. The standard routine in many protein crystallography laboratories is to perform affinity chromatography (most commonly using a His-Tag), followed by size exclusion chromatography (SEC) as a polishing step. However, for the purification of some proteins inclusion of further steps may be required.

3. 4. 4. 1. Immobilized metal affinity chromatography (IMAC)

Affinity chromatography is enabled by the addition of a tag to the target protein, which mediates binding to a specific column matrix, while untagged proteins pass directly through the column. All proteins purified in this work contain an N-terminal His_6-Tag, which is compatible with Immobilized metal affinity chromatography (IMAC), the most commonly used chromatographic technique. Various metal ions have an affinity to histidine, in this work Ni^{2+}, immobilized by the chelating agent nitrilotriacetic acid (NTA), was used. The method is therefore also referred to as Ni-NTA chromatography. Elution of the desired proteins is achieved by the addition of imidazole, which displaces the bound target protein[76].

A peristaltic pump was used to apply the cleared and sterile-filtered cell lysate on a 5 mL Ni-NTA column (*Macherey-Nagel*), equilibrated with at least 5 column volumes (CV) Ni-NTA buffer 1. To evaluate appropriate imidazole concentrations in the wash and elution buffer for each protein, a step-wise increase of imidazole concentrations was done in the first purification (4 CV per step). The fractions were then analyzed via SDS-PAGE for presence and purity of the desired protein; fractions containing the target protein were pooled and subjected to the next purification step. Further purifications only consisted of sample application, a wash step, and elution. Imidazole concentrations of wash and elution buffers are summarized below for each protein.

	Imidazole concentration [mM]	
Protein	**Wash**	**Elution**
Awp1A	30	250
Awp3A	20	500
Awp14A	15	250
*Ct*Pth11	20	500

3. 4. 4. 2. Size exclusion chromatography (SEC)

Size exclusion chromatography (SEC) was done after IMAC as a polishing step. Molecules are separated based on their size, but also the shape, or more exactly their hydrodynamic diameter, plays a role in the separation process. Operation of SEC benefits from two differently accessible volumes, the external volume and the internal volume. The internal volume is the liquid within the porous matrix of the SEC column, which is typically composed of beads. The external volume is the liquid between the beads and is also called void volume. Smaller molecules travel through both the external and the internal volume; they migrate more slowly through the column. Molecules larger than the beads of the column matrix only pass through the external volume and elute at the void volume. A variety of different resins are available for SEC, adjusted to the size and type of the molecule, as well as the choice of eluent and other parameters. In this work, Superdex resins were used, which consist of a dextran matrix bound to cross-linked agarose[77].

The pooled fractions from IMAC containing the desired protein were concentrated to a final volume of approximately 2 mL. The sample was then filtered to remove aggregates or physical contaminants (foreign particles) using centrifugal filter units (Ultrafree-MC, *Merck*). It was then applied on a SEC column that had been equilibrated with sterile-filtered and degassed SEC buffer. Choice of the SEC column was based on the size and the expected quantity of the desired protein. SEC was run on an NGC Chromatography System, eluting proteins were detected by absorption at 280 nm and collected in 1.5 mL fractions. After SEC, the purity of the desired protein was assessed via SDS-PAGE and the sample was either concentrated or flash-frozen in liquid nitrogen and stored at – 80 °C for further use.

Protein	**Columns**	**SEC buffer**
Awp1A	26/600 Superdex 200 pg 16/600 Superdex 200 pg	20 mM Tris-HCl, 300 mM NaCl, pH 8.0
Awp3A	26/600 Superdex 200 pg 16/600 Superdex 200 pg	20 mM Tris-HCl, 300 mM NaCl, pH 8.0
Awp14A	26/600 Superdex 200 pg 16/600 Superdex 200 pg	20 mM Tris-HCl, 300 mM NaCl, pH 8.0
*Ct*Pth11	26/600 Superdex 75 pg	50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0

3. 4. 5. Protein concentration

The pooled fractions from IMAC and from SEC were concentrated using Amicon Ultra concentrators (*Millipore*). The concentrators are available with different molecular weight cut-offs (MWCO), a MWCO of 30 kDa was used for the concentration of Awp1A, Awp3A, and Awp14A. For concentrating *Ct*Pth11, a MWCO of 3 kDa was chosen. Concentrators were first rinsed with dH_2O, then the membrane was equilibrated with the buffer, in which the protein was currently contained, by centrifugation at 3200 g, at 4 °C, for 5 min. The protein solution was then filled into the concentrator and centrifuged at 3200 g, at 4 °C, for 15 min. The concentration step was repeated until either the desired amount or concentration of the protein solution was reached. The protein solution was mixed between each concentration step.

3. 5. Protein analysis

3. 5. 1. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is used to separate proteins based on their size, and consecutively visualize them via staining of the gel. It was used to visualize the desired proteins and to roughly estimate their purity by analyzing each fraction after IMAC and SEC. Protein samples are mixed with a SDS-PAGE loading buffer, which contains SDS and β-mercaptoethanol. These components (often used in combination with heating of the sample to 95 °C for several min) ensure that proteins are linearized, as SDS is able to denature proteins and β-mercaptoethanol reduces disulfide bonds. Simultaneously, the negatively charged SDS attaches to the linearized proteins and hides their surface charges, resulting in a constant mass/charge ratio. Proteins are then separated in the gel by application of an electrical field. In this work, discontinuous SDS-PAGE was used, a technique that provides improved separation compared to continuous SDS-PAGE.

4 µL sample were mixed with 4 µL 2x SDS-PAGE loading buffer and pipetted into the pockets of a gel with a 4.5% (v/v) stacking gel and either a 12% (v/v) or 15% (v/v) separation gel. One pocket of the gel was loaded with 5 µL Pierce Unstained Protein MW Marker. SDS-PAGE was run in an SDS-PAGE chamber filled with SDS-PAGE running buffer with an EPS 301 power box set to 35 mA per gel, until the sample reached the end of the separation gel. Protein bands were then visualized by staining the gel with hot Coomassie for 5 min, followed by destaining in hot destain solution, until bands were clearly visible.

SDS-PAGE, 12 gels

	Stacking gel (4.5%)	Separation gel (12%)	Separation gel (15%)
dH_2O	29.8 mL	32 mL	18.3 mL
Stacking gel buffer	12.5 mL	-	-
Separation gel buffer	-	20 mL	20 mL
Acrylamide (30%)	6.67 mL	32 mL	40 mL
SDS (10% w/v)	500 µL	800 µL	800 µL
APS (10% w/v)	500 µL	800 µL	800 µL
TEMED	50 µL	80 µL	80 µL

Stacking gel buffer		Sepeartion gel buffer	
Tris/HCl, pH 6.8	625 mM	Tris/HCl, pH 8.8	1.125 M
		Saccharose	30% (w/v)

2x SDS-PAGE loading buffer		10x SDS-PAGE running buffer	
Tris/HCl, pH 6.8	62.5 mM	Tris	30.3 g
Glycerol	15%	Glycine	144.4 g
β-mercaptoethanol	4% (v/v)	SDS	10 g
SDS	4% (w/v)	ddH_2O	ad 1 L
Bromphenolblue	a pinch		

Coomassie		Destain	
Coomassie brilliant blue R250	3.2 g	Ethanol	400 mL
Ethanol	400 mL	Acetic acid	80 mL
Acetic acid	80 mL	dH_2O	400 mL
dH_2O	400 mL		

3. 5. 2. Determination of protein concentration

There are various analytical methods available for the determination of protein concentration. The type of method that can be used depends, among other things, on the composition of the protein solution (defined or undefined), the properties of the protein and the choice of buffer. Time considerations and reproducibility are also important selection criteria. In this thesis UV spectroscopic analysis was used to determine the protein concentrations. The method is based on UV absorbance, usually measured at 280 nm, which relies on the aromatic amino acids. However, it has to be kept in mind that these show strong differences in their absorption behavior. The absorption maximum of both tryptophan and tyrosine is 280 nm, while the maximum of phenylalanine is about 260 nm. In addition, the protein's structure can change the absorption behavior[78].

The extinction (*E*) of the sample was measured using a NanoDrop photometer. Knowing E, the Lamber-Beer law can be applied to determine the protein concentration:

$$E = \varepsilon \cdot c \cdot d$$

$$c_m = \frac{E \cdot MW}{\varepsilon \cdot d}$$

E: extinction; ε: molar absorptivity; c: concentration; d: length of the solution the light passes through; c_m: mass concentration; MW: molecular mass

In addition, the molecular mass of the protein and its extinction coefficient are required for determination of its concentration. These values were calculated from the amino acid sequence using the online tool *ProtParam*, which is available on the ExPASy Bioinformatics Resource Portal[59].

3. 5. 3. Thermal shift assay (TSA)

A wide variety of methods are available for the characterization of protein-ligand interactions. The thermal shift assay (TSA) – also referred to as differential scanning fluorimetry or thermofluor assay – offers a relatively high throughput, while it can be easily performed with standard lab equipment.

Ligand binding is associated with a change in protein stability, usually it leads to stabilization of the protein. In a TSA, the thermal stability of protein solutions containing ligands is measured with the aim of detecting changes in melting temperature – i. e. thermal shifts. The detection of protein unfolding is facilitated by addition of SYPRO Orange, a component that shows low fluorescence in polar environments and high fluorescence in non-polar environments. Upon denaturation of the protein, its non-polar core is exposed, leading to an increase of fluorescence signal. As a result, a melting curve is obtained, of which the melting temperature is the inflection point (maximum of the first derivative).

	Volume
Awp1A/Awp3A (50 µM)	4 µL
Glycan (50 mM*)	4 µL
SYPRO Orange (1:62.5)	4 µL
SEC buffer	ad 40 µL

* if not indicated otherwise

Binding of Awp1A and Awp3A to various disaccharides and oligosaccharides was analyzed in a TSA using 40 µL reaction volumes. The experiment was run in a RotorGene Q (*Qiagen*), the temperature was raised by 0.2 °C each 4 sec, from 25 °C to 90 °C. Gain optimization was done

manually. The TSA mixture is shown below; reference measurements were done by adding the solvents of the ligands to the protein solution and 8x SYPRO Orange. Following carbohydrates were used: Laminarin, beta glucan (barley, 0,1%), CM-curdlan (0,01%), 3-O-(β-D-galactopyranosyl)-D-galactopyranose, Galβ1-3GlcNAc, Galβ1-3GalNAc, N,N'-diacetylchitobiose, Galα1-3Gal, 3-Fucosyllactose, Lewis[a] trisaccharide, lacto-N-tetraose, lacto-N-neotetraose, Galα1-3Galβ1-4Gal, Galβ1-4GlcNAc, Manα1-6Man, Manα1-2Man, Manα1-3Man, Manα1-4Man, mannotetraose, mannopentaose and Galβ1-3GalNAcβ1-4Galβ1-4Glc.

3. 5. 4. High throughput glycan binding studies at the Consortium for Functional Glycomics

Another high-throughput method for screening specific interactions is suspension array technology, which uses glass slides printed with specific components (DNA, peptides, glycans). The Consortium of Functional Glycomics (CFG) offers the implementation of so-called glycan arrays, which allow screening for binding of a protein to several hundred immobilized glycans.

Purified protein samples were sent to the Consortium for Functional Glycomics (CFG), where binding of Awp1A and Awp3A was examined on the newest version of the Mammalian Glycan Array (version 5.2), as described by Heimburg-Molinaro *et al.*[79]. For both samples a protein concentration of 50 µg/mL in SEC buffer was used. Detection was carried out via an anti-His antibody, coupled to AlexaFluor 488 (*Qiagen*). Glycan array data was deposited at the CFG, under the identifier cfg_rRequest_3531.

3. 6. Determination of protein structures

Structural biology is concerned with the analysis of the 3D structures of biological macromolecules, especially proteins and nucleic acids. Proteins play an essential role in every aspect of life and the structure of a protein is uniquely suited to its function. Therefore, new insights can be generated into a protein's function via the determination of its structure.

Three major methods are commonly used for the determination of structures of biological macromolecules: X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and 3D electron microscopy (3D-EM). Obtained structures are deposited in the Protein Data Bank (PDB), an open access database for the 3D structures of large biological molecules. 88.8% of the structures in the PDB have been determined via X-ray crystallography, highlighting the importance of the method for biological sciences. NMR spectroscopy accounts for 7.9% of structures in the PDB and 3D-EM for 3.2%. Around 0.2% of structures have been determined using multiple methods or other methods (e. g. neutron diffraction or solution scattering).

In recent years, 3D-EM has gained popularity due to advances in the technology of detectors and in image processing, enhancing the resolution that can be achieved with these structures[80]. Nevertheless, also X-ray crystallography has seen major recent developments: various X-ray free electron lasers were put into operation. Among other applications, they provide the possibility to determine crystal structures in a time-resolved manner (time-resolved serial femtosecond crystallography), thereby providing direct insights into functional reactions of the sample[81].

3. 6. 1. Protein crystallization

The crystallization process is a bottleneck in the process of protein structure determination. First protein crystals were already described in 1840, but remained a laboratory curiosity for a few decades. From the 1880ies on, protein crystallization was done as a purification method, until – around 1930 – protein crystals acquired a new application, when X-ray crystallography was applied for the determination of the structures of biological macromolecules[82]. Although the crystallization process has been observed for 160 years, the exact requirements for crystallization are still unknown and the process remains unpredictable. However, a few requirements that lead to a higher probability of crystallization are known: protein samples have to be pure and monodisperse. These prerequisites are ensured by the purification process.

Crystallization itself is reached by a slow decrease of the solubility of a protein by addition of precipitants. In some cases, this leads to formation of a so called nucleus, around which crystals then grow. The process is often described via a phase diagram (see Figure 5) and depends on many different factors, such as protein concentration, precipitant concentration, pH, temperature, additives, ligands, inhibitors, coenzymes, and many others. Nowadays, the bottleneck of crystallization is tackled by trying out a large amount of different crystallization

conditions. Various methods are available, the most common ones are microbatch experiments, vapor diffusion, dialysis, and free interface diffusion. Many labs use sitting drop vapor diffusion setups, which can be pipetted by robots in a short time, usually in a 96 well format. Nevertheless, the overall success of crystallizing a protein in structural biology laboratories is estimated to be around 30 – 40%[83].

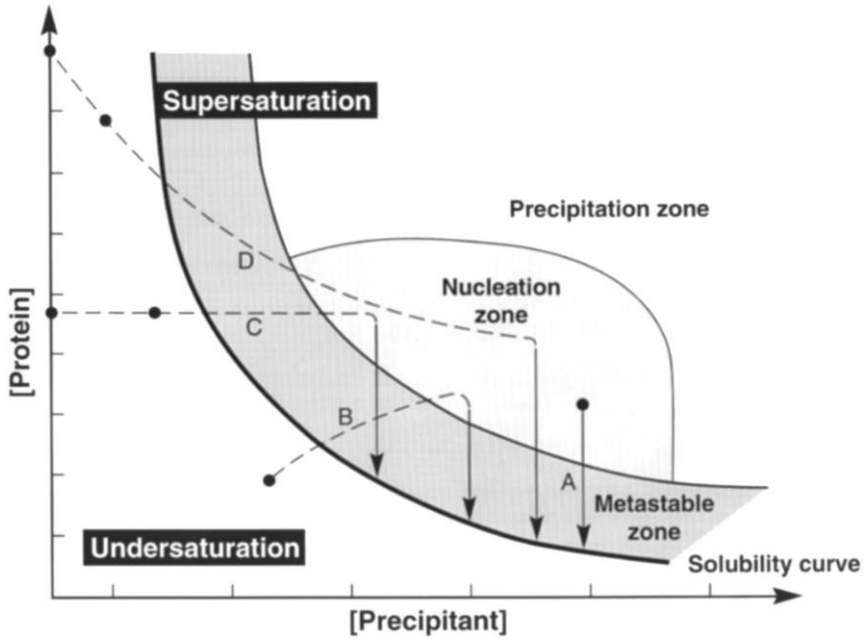


Figure 5: Protein crystallization phase diagram[84]

A phase diagram with commonly varied parameters of a crystallization experiment – protein concentration and precipitant concentration – is displayed. Additionally, the crystallization curves of the most common protein crystallization methods are indicated: A) Batch crystallization, B) Vapour diffusion, C) Dialysis, D) Free-interface diffusion (liquid/liquid diffusion). All methods aim to reach the nucleation zone, from which the system progresses through the metastable zone to finally arrive at the solubility curve.

Initial crystallization experiments were done in the MarXtal crystallization facility in a sitting drop vapor diffusion setup, using a variety of commercially available screens (see below). The screens were pipetted with a crystallization robot (Honeybee 963, *Digilab*). The reservoir of MRC 2 Well plates (*Swissci*) were filled with 80 µL mother liquor, 300 nL mother liquor and 300 nL protein solution were pipetted in each well. The plates were then sealed with sealing film and incubated at 18 °C in a Rock Imager (*Formulatrix*) crystallization imager, a system that is also documenting crystal growth.

Protein	Protein concentrations	Crystallization Screens
Awp1A	48 mg/mL 24 mg/mL	JCSG Core I (*Qiagen*), JCSG Core II (*Qiagen*), JCSG Core III (*Qiagen*), JCSG Core IV (*Qiagen*), Morpheus (*Molecular Dimensions*), Morpheus II (*Molecular Dimensions*), Classics (*Qiagen*)
Awp3A	24 mg/mL 12mg/mL	JCSG Core I (*Qiagen*), JCSG Core II (*Qiagen*), JCSG Core III (*Qiagen*), JCSG Core IV (*Qiagen*), Morpheus (*Molecular Dimensions*), Morpheus II (*Molecular Dimensions*), Classics (*Qiagen*)
Awp14A	22 mg/mL 11 mg/mL	JCSG Core I (*Qiagen*), JCSG Core II (*Qiagen*), JCSG Core III (*Qiagen*), JCSG Core IV (*Qiagen*), Morpheus (*Molecular Dimensions*), Morpheus II (*Molecular Dimensions*), Classics (*Qiagen*), Classics Lite (*Qiagen*)
***Ct*Pth11**	10.8 mg/mL 5.4 mg/mL	JCSG Core I (*Qiagen*), JCSG Core II (*Qiagen*), JCSG Core III (*Qiagen*), JCSG Core IV (*Qiagen*), Morpheus (*Molecular Dimensions*), Morpheus II (*Molecular Dimensions*), Classics (*Qiagen*), AmSO4 (*Qiagen*)

3. 6. 2. Optimization of crystallization conditions

If a crystallization condition is identified in initial crystallization experiments, it is often followed by an optimization of said condition with the purposes of crystal reproduction for additional experiments and growing crystals with a better diffraction quality. Usually, two factors influencing protein crystallization are altered around the original condition, e. g. pH and precipitant concentration.

3. 6. 2. 1. Optimization of Awp1A crystals

A 24-well hanging drop vapor diffusion optimization screen was pipetted, as depicted in the scheme in Figure 6. The original crystallization solution contained 0.1 M MOPSO/Bis-Tris pH 6.5, 10% (w/v) PEG 8000, 20% 1,5-pentanediol, 0.5 mM erbium(III) chloride hexahydrate, 0.5 mM terbium(III) chloride hexahydrate, 0.5 mM ytterbium(III) chloride hexahydrate, and 0.5 mM yttrium(III) chloride hexahydrate[85]. For the optimization screen the ratios of MOPSO and Bis-Tris were changed to alter the pH and the concentrations of both precipitants were varied in the same proportion to each other. Erbium(III) chloride, terbium(III) chloride and ytterbium(III) chloride were present at a concentration of 0.5 mM each. A drop size of 1.2 µL was chosen, composed of 0.6 µL reservoir and 0.6 µL protein solution. Two drops were set, one using a protein concentration of 48 mg/mL, the other one using 24 mg/mL. The crystallization plate was incubated at 20 °C.

precipitant	30 : 70 MOPSO : Bis-tris	40 : 60 MOPSO : Bis-tris	50 : 50 MOPSO : Bis-tris	60 : 40 MOPSO : Bis-tris	70 : 30 MOPSO : Bis-tris	80 : 20 MOPSO : Bis-tris
5 % PEG 10 % pentane-diol						
7.5 % PEG 15 % pentane-diol						
10 % PEG 20 % pentane-diol						
12.5 % PEG 25 % pentane-diol						

Figure 6: Optimization screen for Awp1A

The pipetting scheme of the optimization screen is shown. MOPSO and Bis-Tris were used at a final concentration of 0.1 M MOPSO/Bis-Tris. The concentrations of PEG 8000 are given in (w/v), the concentrations of 1,5-pentanediol are given in (v/v). Buffer mixing ratios were varied along the x-axis, the precipitant concentrations were changed along the y-axis. All reservoir solutions contained 0.5 mM erbium(III) chloride, 0.5 mM terbium(III) chloride, and 0.5 mM ytterbium(III) chloride.

3. 6. 2. 1. Optimization of Awp3A crystals

The original crystallization condition of Awp3A contained 0.2 M $MgCl_2$, 0.1 M Tris pH 7.0 and 2.5 M NaCl. The condition was optimized in a hanging drop vapor diffusion setup, as described for Awp1A. The pH was varied along the y-axis of the optimization, in a range from 7.0 to 8.5, using increments of 0.5. Different precipitant concentrations were used along the x-axis of the screen, ranging from 1 M NaCl to 3.5 M NaCl, in increments of 0.5 M. The salt concentration (0.2 M $MgCl_2$) remained unchanged. Additionally protein solution:reservoir ratios of 1:1, 1:2 and 2:1 were used. The crystallization plate was incubated at 20 °C.

3. 6. 3. Crystal harvesting and soaking

When crystals stop growing or when the next beam time at a synchrotron is approaching, protein crystals are harvested and stored in liquid nitrogen for transport to the synchrotron, where diffraction data is collected. For some crystals, soaking may be required or desired at that point. The soaking process can serve different purposes, e. g. protection from ice formation, introduction of heavy atoms for phasing, or introduction of ligands into the protein crystals.

Protein crystals typically have a high solvent content with usually observed values around 50% and a range from around 30% to 85%. In many cases, the solvent is an aqueous solution that

will form ice upon freezing. Ice formation lowers diffraction quality by disruption of the protein crystal structure and ice rings can be observed on the diffraction images. Thus, the process is often prevented by soaking protein crystals with cryoprotectants, such as glycerol, ethylene glycol or MPD. The necessity of introducing additional cryoprotectants depends on the crystallization condition.

Another purpose of protein crystal soaking is the introduction of heavy atoms for solving the phase problem. Some phasing methods, such as multiple wavelength anomalous dispersion (MAD) or single wavelength anomalous dispersion (SAD) require the presence of heavy atoms showing said anomalous dispersion. These can be naturally present in the protein, for example as cofactors or as part of a ligand. If that is not the case, the experimenter can choose between several methods for the introduction of heavy atoms: The substitution of the methionine residues in a protein by selenomethionine (SeMet) is a common method called SeMet labeling. It has the advantage that the number of heavy atom sites within the asymmetric unit of the protein crystal – a variable that may be decisive for the phasing process – is already known. However, protein production, purification and crystallization may have to be adapted when working with SeMet labeled proteins. Additionally, some proteins do not have a sufficient amount of Met residues and one or even a few mutations need to be introduced– as a rule of thumb, at least one SeMet per 100 AA is required for phasing[86,87]. Another approach for heavy atom derivatization is soaking already existing crystals in heavy atom containing solutions. Soaking is a lot swifter than SeMet labeling, because overproduction, purification and crystallization do not have to be repeated. However, it is not predictable whether the protein crystal will endure the soaking process and whether the protein will bind the metal. A variety of heavy metal compounds are available for phasing purposes and also iodine and bromide can be used[87].

In some cases, ligands or a variety of potential ligands are introduced into the protein crystal by soaking. The structural context of ligand binding in a protein using an already known ligand can be examined in this way, but also ligand screening experiments are often conducted by soaking. In the recent years, fragment-based lead discovery (FBLD) has become a conventional approach for drug discovery. Protein crystals are soaked with a variety of low-molecular-mass molecules, i. e. fragments. If binding is observed, the fragments can be combined or upsized into lead compounds[88].

In this work, crystals were harvested and usually directly flash-frozen in liquid nitrogen without any additional cryoprotectant. To enable phasing of Awp3A by single wavelength anomalous diffraction (SAD), crystals were transferred to a drop of mother liquid, containing 50 mM Gd(III) acetate. They were allowed to sit in this drop for 90 min and then flash-frozen in liquid nitrogen without any additional cryoprotectant.

Table 3: Fragment concentrations and soaking times used for fragment binding experiments on the *Ct*Pth11 CFEM domain

Fragment No.	Concentration [mM]	Soaking times	Fragment No.	Concentration [mM]	Soaking times
1	50	10 sec	46	100	25 min
2	50	23 h	47	100	30 min
3	50	23 h	48	100	1 min, 20 min
4	50	23 h	49	100	20 sec
5	50	23 h	50	50*	2 ½ h
6	100	2 min	51	100	1 h, 19 h
7	50	20 min	52	100	3 min
8	100	26 h	58	100	5 min
9	50	2 min, 4 min	59	100	3 h
10	50	26 h	60	100	7 min, 10 min
11	100*	30 sec	61	100	3 h, 19 h
12	100*	30 sec, 1 min	62	100	6 min
13	50	26 h	63	50	19 h, 24 h
14	100*	10 sec, 30 sec	64	50	19 h, 24 h
15	100*	15 sec	65	100	3 h
16	100	5 min	66	100	2 h
17	100	5 min	67	50	15 sec, 6 min
18	100	30 min, 50 min	68	50	3 h, 24 h
20	50	3 h, 26 h	69	100	30 sec
21	50	3 h, 26 h	70	50	3 h
22	100*	10 min, 20 min	71	100	1 h, 3 h
23	100	5 min, 3 h	72	100	3 h, 24 h
24	100	20 min, 1 h	73	100	3 h, 24 h
25	50	1 h	74	50	3 h, 24 h
26	100	1 min	75	50	3 h, 24 h
27	100	1 ½ h, 3 h	76	50	3 h, 24 h
28	100	1 ½ h, 3 h	77	50	3 h, 24 h
29	100	1 ½ h	78	50	3 h, 24 h
31	50	1 h, 3 h	79	100*	2 min
32	100	4 min, 10 min	80	50	3 h, 24 h
33	100	10 min, 24 h	81	50	3 h
34	50	3 h, 24 h	83	100	30 sec
35	100	30 sec	84	100	10 sec, 1 min
36	100	10 min	85	100	15 min, 20 min
37	100	15 min, 50 min	86	50	3 h
38	50	3 h, 4 h	87	100	20 min
39	100	15 min, 90 min	88	100	1 ½ h
40	100	1 h	89	100	12 min
41	50	1 h, 24 h	90	50*	10 sec
42	100	30 min	91	100	14 min, 15 min
43	100	30 min	92	50*	30 sec
44	100*	10 sec, 3 min	93	100	3 h
45	100*	15 sec, 1 min	94	100	1 min, 2 min

* Fragment powder remaining undissolved was centrifuged and the supernatant was used for soaking.

Fragment number refers to the Frag Xtal Screen from *Jena Bioscience*.

Crystals of the *Ct*Pth11 CFEM domain were soaked with fragments from the Frag Xtal Screen (*Jena Bioscience*), which were received as a kind gift from the group of Prof. Dr. Gerhard Klebe and Prof. Dr. Andreas Heine. 1 M fragment stock solutions (in DMSO) were mixed with mother liquor and glycerol (2.4 M ammonium sulfate, 0.8% MPD, 20% glycerol) to reach a final fragment concentration of either 100 mM or 50 mM, depending on the solubility of the fragment. Crystal soaking times of 3 h in the fragment containing solution were aimed for. If that time was not achievable, crystals were soaked as long as possible (up to 26 h), i.e. the crystals were harvested as soon as severe fractures were observed or when 26 h had passed. A summary of the soaking experiments conducted on *Ct*Pth11 can be found in Table 3. Appendix II contains a list of all datasets collected during the soaking experiments and the fragments used.

3. 6. 4. Principles of X-ray diffraction

The principles of X-ray diffraction are described in several excellent resources, both online and offline, in open-access resources and available for purchase. The topic is therefore only described briefly in this work.

The prerequisite for being able to collect meaningful high resolution X-ray diffraction data is the presence of a crystal. Crystals are characterized by the periodic arrangement of a certain motif within a three-dimensional lattice. The repeating motif is referred to as the *unit cell*, the whole crystal can be recreated by translation of the *unit cell* in the three lattice directions (a, b, c). The smallest fragment of the crystal is the *asymmetric unit*, from which the unit cell can be recreated by symmetry operations. The symmetry of the molecules within a crystal is described by the *crystallographic space groups*. A combination of the seven *crystal systems* (*triclinic*, *monoclinic*, *orthorhombic*, *tetragonal*, *trigonal*, *hexagonal*, *cubic*) and 14 *Bravais lattices* results in 230 *space groups*. But because proteins are chiral molecules, certain symmetry operations (such as inversion or reflection) cannot be performed. Thus, 65 *space groups* are viable in protein crystallography. Another common characteristic of crystals is *mosaicity*. In an impeccable crystal, all *unit cells* would be perfectly aligned. Naturally, most crystals are not perfect and show slight displacements of blocks of unit cells relative to each other, i.e. *mosaicity*.

When the crystal is placed in an X-ray beam, an X-ray diffraction pattern is the result. Diffraction occurs when X-rays with a wavelength that approximately corresponds the lattice parameters of the protein crystal are directed at the crystal and can be explained by the *Bragg model*. It describes in which circumstances constructive interference of scattered X-ray beams can occur, resulting in *Bragg reflexes*. In this context, the crystal is regarded as a set of equally spaced planes that are parallel to each other (*Bragg planes*). Each plane acts as mirrors for the incident X-ray beam – the angle of incidence (θ) equals the angle of scattering. Constructive interference of X-rays that are scattered from adjacent planes can only occur under certain circumstances, formulated as *Bragg's law* (see formula below and Figure 7).

$$2d \sin\theta = n\lambda$$

d = spacing between planes; θ = angle between plane and X-ray; n = integer; λ = X-ray wavelength

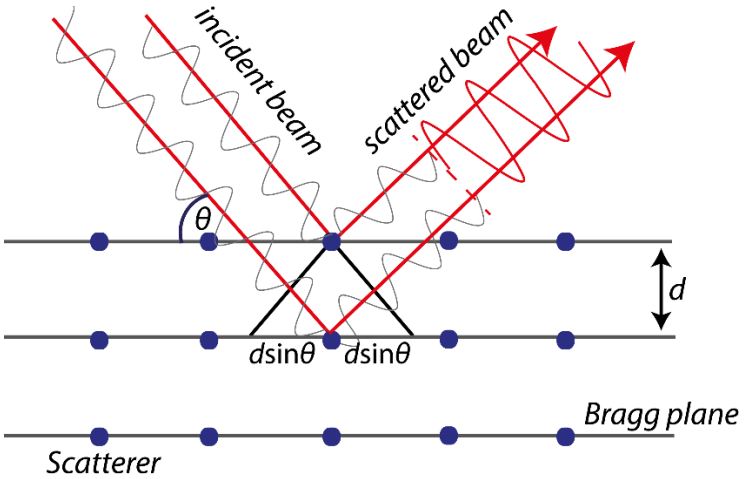


Figure 7: Visual representation of Bragg's law

X-ray beams with the wavelength λ meet scatterers at the imaginary *Bragg planes* – which are separated by the distance *d* – at an angle of incidence θ. When the total path length difference 2*d*sinθ is an integer number of λ, Bragg's law is fulfilled and constructive interference will occur.

3. 6. 5. Practical approach to data collection

All datasets from this thesis were measured at the ESRF (beamlines ID29, ID23-1, ID23-2) or at the SLS (beamlines PXI or PXIII).

When X-ray diffraction data is collected from a crystal at a synchrotron beamline, several criteria should be adapted to measure high-quality datasets in a reasonable amount of time[89]. Automatic sample changers at the synchrotrons mount the loop onto a goniometer, where the crystal is cooled in a cryostream at 100 K. Room temperature measurements are rather unusual nowadays, because the radiation damage is lower at colder temperatures, allowing the collection of a complete dataset from a single crystal. At least two test exposures at orthogonal orientations (e. g. 0° and 90°) are done for determining the space group and estimating optimal data collection parameters. The required calculations are automatically done by the data acquisition and analysis software. At the ESRF, the data collection software is *MxCuBE* and data collection strategies are calculated by the *EDNA* framework[90]; at the SLS both tasks are run by the automatic data analysis software *DA+*[91]. Nevertheless, results from the data analysis software should be examined by the experimenter and manual estimations of the parameters should be done if required. During the work for this PhD thesis, the estimation of a crystal's resolution by *DA+* was found to be flawed in many cases; usually the appropriate crystal-to-detector distance had to be estimated manually. Exposure times and beam intensity were often accepted as indicated by the data analysis software. Datasets for

SAD phasing were tendentially collected with lower beam intensity to prevent extra radiation damage, to which the presence of heavy atoms can contribute significantly. The wavelength for measurement of native datasets was set to approximately 1 Å (12.398 keV); when anomalous data was measured, it was changed accordingly (see Pike *et al.*[87]). Usually, rotation ranges of in total 180° to 360° were collected, even if not necessarily required for collecting a complete dataset. As a larger range of rotation yields multiple measurements of the symmetry-equivalent reflections, it theoretically results in higher-quality data. However, radiation damage has to be taken into account (amongst other factors)[92].

Datasets for S-SAD phasing of *Ct*Pth11 were measured according to a specialized data collection strategy described by Basu *et al.*[93], together with Dr. Vincent Olieric from the SLS, Villigen. The wavelength was set to 5.5 keV (2.25 Å) and a raster scan was used to determine the best diffracting location within the crystal. Then, the first 360° ω dataset was collected. The starting angle for data collection was altered by +5° in κ and ϕ orientations and another 360° ω dataset was measured. This protocol was repeated, until the data collection statistics revealed significant radiation damage. Using this method, four datasets with acceptably low radiation damage could be collected from a single crystal.

3. 6. 6. Data processing and data reduction

Data processing consists of several steps: First, the space group of the crystal is determined – a procedure called indexing. Then, the intensities of the measured reflexes are integrated and finally, they are scaled[54]. During data reduction, the data is scaled to produce internally consistent data. Several datasets can also be combined in this step, a process called merging[94]. In this work, mainly the program *XDS*[54] was used for data processing and *AIMLESS*[95] was used for data reduction.

In *XDS*, data processing is done in consecutive steps: XYCORR, INIT, COLSPOT, IDXREF, DEFPIX, INTEGRATE, and CORRECT. Each step produces a log file, named after the step, with the appendix ".LP" added. In the XYCORR step, geometrical corrections are applied if required. Correction files have to be specified in the *XDS* input file ("XDS.INP"). INIT then calculates the gain of detector, i. e. it differentiates between background and reflexes. In the COLSPOT step, strong reflections are identified, which are then used for indexing during IDXREF. Possible space groups are determined in this step. XDS chooses the space group with the highest symmetry and a good quality of fit for further processing steps. If the chosen space group is found to be incorrect, the space group can also be specified by the user to enforce correct cell constants. In the subsequent process – DEFPIX – certain pixels of the detector are labelled to be ignored during the integration step. INTEGRATE then calculates the intensities of the reflections in the dataset and CORRECT corrects the calculated intensities for decay, absorption, and variations of detector surface sensitivity. Processing statistics are provided in the "CORRECT.LP" file and the output file – "XDS_ASCII.HKL" – is generated[54].

After processing a dataset, the space group was reviewed and corrected if required. Additionally, the quality of the data was assessed and the resolution of the dataset was estimated using the statistics provided in the "CORRECT.LP" file. CC(1/2) and *I*/σ were treated as main indicators for the dataset's resolution. The dataset was then processed again, with space group and resolution already specified in XDS.INP.

Data processing is followed by data reduction, which was done using *AIMLESS*[95], run within *CCP4i2*[51]. In this procedure the space group is determined a second time, because the indexing in the integration program only detects the lattice symmetry, which may not reflect the true symmetry. Symmetry related observations of reflections are then scaled and merged and a free-R set is generated, by default with 5% of the data[94,95]. Statistics given by AIMLESS were then used to reassess data quality and the resolution.

3. 6. 7. Structure determination – solving the phase problem (SAD, S-SAD & MR)

To be able to determine the structure of a molecule from X-ray diffraction data, the *phase problem* must be solved. The *real space* (i.e. the electron density function) and the *reciprocal space* (i.e. the structure factors measured in the diffraction experiment) are related to each other via the *Fourier transform*. The *real space* can be used to calculate the *reciprocal space*, but not vice versa, because some information is lost during data acquisition. This missing information are the phases of the X-ray waves, therefore the dilemma of lacking information is referred to as the *phase problem*. The *phase problem* is described by following equation:

$$\rho(xyz) = \frac{1}{V}\sum_{\substack{hkl \\ -\infty}}^{+\infty} |F(hkl)| \cdot e^{-2\pi i[hx+kx+lz-\phi(hkl)]}$$

ρ(xyz) = function of electron density at position xyz, V = Volume of the unit cell, |F(hkl)| = structure factor amplitudes, Φ(hkl) = phase associated with F_{hkl}

Several methods are available for the determination of phases. The most commonly used method is Molecular Replacement (MR), which requires a model of a similar structure. Other methods do not rely on the availability of structural information[92,96]; among those, single-wavelength anomalous diffraction (SAD) phasing has become the preferred structure solution method for many crystallographers[92,97].

3. 6. 7. 1. SAD phasing enabled by heavy metal soaking

For SAD phasing, the presence of *anomalous scatterers* is required to solve the phase problem. The anomalous scattering effect is especially strong for heavier atoms, thus some of the classical compounds brought into crystals for structure determination purposes contain Hg, Pt, U or Au[87], but the use of lanthanides has also proven to be well suited for phase determination[10,98,99].

Anomalous diffraction occurs when heavy atoms are subjected to an X-ray wavelength at or near their absorption maximum. Therefore, the experiments may have to be conducted at tunable synchrotron beamlines, i. e. beamlines where it is possible to alter the X-ray wavelength. Absorption maxima are different for each atom and can either be determined experimentally or extracted from literature (e. g. found in Pike *et al.*[87] or under http://skuld.bmsc.washington.edu/scatter/AS_periodic.html). When anomalous diffraction occurs, *Friedel's law* is broken. Certain reflections are related to each other by inversion through the origin (they occupy the positions h, k, l and –h, -k, -l), these are referred to as a *Friedel pairs*. *Friedel's law* states that these have equal amplitude and opposite phase, hence the intensity of the reflections is equal. When it is not fulfilled, a difference in the intensities of this pair of reflections can be observed, called the *Bijvoet difference*[92].

To be able to determine the phases, the positions of the anomalous scatterers have to be determined first. This is achieved from the *Bijvoet differences* using *Patterson* or *direct methods*[100]. This results in two possible enantiomers, of which the correct one is selected by evaluating which hand provides the better electron density map for the partial structure. The heavy atom parameters are refined, before the starting phases for the protein are deduced from the calculated anomalous model phases. Finally, phases are improved by density modification[92].

To enable phase determination of Awp3A via SAD, crystals were soaked in a drop of mother liquor containing 50 mM Gd(III) acetate for 90 min, before they were frozen in liquid nitrogen. For data collection, the X-ray wavelength was set to 1.71237 Å, which is near the L-III absorption edge of Gd. Crystallographic phases of a SAD dataset were determined using *CRANK2*[101].

3. 6. 7. 2. Native SAD phasing using the anomalous diffraction from sulfur

Native SAD phasing exploits the anomalous diffraction from atoms not heavier than calcium (atomic number 20) for structure solution. Such can occur naturally in the protein, in ligands (e. g. phosphorous in bound DNA or RNA), or in buffers. In many cases, the sulfur atoms from cysteine or methionine residues in the protein are used for native SAD phasing, a practice also referred to as S-SAD phasing[102]. Usually, the wavelength of the X-ray beam wavelength cannot be adjusted to be very close to the X-ray absorption edge of the atom addressed in this phasing approach. This results in only low anomalous signal, so the data has to be collected carefully

to increase the signal to noise ratio of the data by reduction of noise[97]. This is often achieved by collection of several datasets and merging the data[93]. Other approaches are also applied at beamlines specialized for the collection of native SAD datasets, i. e. longer wavelength ranges, vacuum or helium environment, or the usage of special detectors[103].

Datasets of *Ct*Pth11 crystals were collected at the SLS, beamline X06DA (PXIII), together with Dr. Vincent Oliereic. The data collection strategy described by Basu *et al.* in 2019 was applied in this case: Several 360° ω datasets were collected from a single crystal, using a wavelength of 2.25 Å. After measurement of a dataset, Κ and ϕ orientations were incremented 5° and the next 360° dataset was collected[93]. S-SAD datasets were then merged on site using a custom script for *xscale*[54]. The structure of the *Ct*Pth11 CFEM domain was solved using *CRANK2*[101].

3. 6. 7. 3. Molecular replacement (MR)

If a structure of a protein with a low root-mean-square deviation (RMSD) to the target protein is accessible, the *phase problem* might be solved by Molecular Replacement (MR). A low RMSD is generally indicated by a high sequence identity, with a minimal sequence identity of 30% often suggested in literature. The critical point in MR is the model quality; thus models may have to be trimmed – i. e. long loops or other flexible regions are removed, as well as bulky side chains – or adapted otherwise (e. g. a polyalanine model can be used)[92,96].

The structure solution process is essentially a comparison of the measured data with the model data. To enable this process, *Patterson maps* are calculated from both the observed data and the model. The maps are then correlated, whereby 6N parameters must be established to define the solution: three rotation angles and three translations for each molecule (N) in the asymmetric unit. As this six-dimensional search would take very long, it is usually split into two three-dimensional searches: maps are rotated against one another, then translated. However, the correct rotation cannot be calculated with an unknown translation. A scoring algorithm has to be applied at this point to pick a small number of solutions to go on with[104]; in *Phaser* this is the *maximum likelihood* method[105]. If the searches are completed successfully, the initial phases can be calculated and an electron density map is generated[104].

In this thesis, *Phaser*[105] was used to solve the structure of Awp1A, with Awp3A serving as a search model. The initial MR result was then subjected to 20 cycles of model building using the model mode in the *ARP/wARP* Web Service (running *ARP/wARP* version 8.0)[55] to obtain a complete structural model. *Phaser*[105] is also implemented in the *DIMPLE* pipeline, which was used to analyze the *Ct*Pth11 fragment screen data (see chapter 3. 6. 8.).

3. 6. 8. Analyzing fragment screen data – the DIMPLE pipeline

DIMPLE (DIfference Map PipeLinE) is an automated software pipeline designed to analyze crystals of a known protein that may have bound a ligand. It has been developed by the CCP4 software group and the Diamond light source and can be run in *CCP4*[106]. A detailed description of the pipeline can be found under: https://ccp4.github.io/dimple/. The workflow applies *rigid-body* refinement to obtain the electron density map of the target structure; MR is done only when necessary.

DIMPLE requires several input files: The model of the apo structure (pdb) and the corresponding reflection data (merged mtz) have to be given, as well as the target structure data (merged mtz). The target structure data is then prepared for *rigid-body* refinement in several steps: if the unit cell constants do not match the apo structure data, reindexing is required; this is done using *POINTLESS*. The data is then truncated (*TRUNCATE*) and *FREERFLAG* is run. When comparing data, it might be advisable to use consistent flags. *DIMPLE* therefore automatically assigns the same flags when the same pdb file is used. Alternatively, external reference flags may be given or the existing flags from the input mtz can be used. After these preparations, *rigid-body* refinement is done by *REFMAC5*, followed by a few more rounds of restrained refinement. Sometimes, MR is required before restrained refinement, this is done using *Phaser*. Finally, *unmodelled blobs* are identified.

In this work, a custom script for running *DIMPLE* on a large amount of datasets measured at the SLS has been used (see Appendix III). The script is written for execution in the Unix shell, using the programming language *Bash*. The input of multiple datasets from the SLS is facilitated by implementation of a step for identifying "XDS_ASCII.HKL" files within a set of given directories. *POINTLESS* and *AIMLESS* are then run to obtain the merged mtz files from the *XDS* output files. Then *DIMPLE* is executed, with Free-R flags derived from the input mtz of the apo structure.

3. 6. 9. Structure refinement

Structure refinement is done to achieve agreement between the structural models obtained in the structure solution process and the experimental data. This is necessary because the initial structural model usually contains errors, i.e. deviations from the electron density map or chemical or physical flaws. During refinement, water molecules and ligands are added as well.

The refinement is carried out in iterative cycles of manual model building and computational refinement; the data are continuously evaluated by the examination of certain parameters during the process. Manual model building is done by inspecting the fit of the model to the electron density map and adjusting it appropriately. Computational refinement is done by statistical improvement of the model to better fit the diffraction data, commonly applying two different methods: *maximum likelihood refinement* (used in *REFMAC*[107]) or *simulated*

annealing (*phenix.refine*[52]). Both use restraints in respect to bond distances, bond angles, torsion angles, and temperature factors (*B-factors*).

The main indicators for the progress and quality of a refinement are the R-factors, R_{work} and R_{free}. These serve as a measure of the agreement between the structural model and the experimental data and are calculated as follows:

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

F_{obs} = structure factor amplitudes of the experimental data, *Fcalc* = structure factor amplitudes calculated from the model

R_{work} is calculated from the working model, whereas R_{free} is calculated from reflections excluded from the refinement process (by default 5% of reflections), providing a tool for cross-validation. R_{work} is always higher than R_{free}, but large differences between the values indicate that the model is over-refined[92].

Most structures in this thesis were refined via iterative cycles of model building, performed in *phenix.refine* (part of the *PHENIX* crystallographic software suite[52]) and *WinCoot*[53]. The refinement of Awp3A-Gd was done in *REFMAC5*[107] (run within the *CCP4* software suite[51]) and *WinCoot*[53].

4. Results

4. 1. The cell wall proteome of *Chaetomium thermophilum*

4. 1. 1. Prediction of GPI-anchored proteins

For prediction of GPI-anchored proteins in *C. thermophilum* several features were considered. Firstly, GPI-anchored proteins have an N-terminal signal peptide, which targets them to the ER[108]. SignalP[61] was used for identification of these signal peptides. The annotated *C. thermophilum* proteome contains 7165 protein sequences; an N-terminal signal peptide was identified in 562 sequences. Typically, GPI-anchored proteins do not contain any transmembrane helices[108], absence of those was analyzed via TMHMM[63]. However, it must be considered that the GPI-anchor attachment sequence is recognized as a transmembrane helix[108], therefore C-termini were ignored in this analysis. Among the 562 proteins with a signal peptide, transmembrane helices were not identified in 473 sequences. Finally, the Big-PI Fungal Predictor was used for detection of C-terminal GPI anchor attachment sequences[12]. 61 GPI-anchored proteins were predicted in this way. As an alternative approach for identification of GPI anchor attachment sequences, a pattern search was conducted using the sequence described by de Groot *et al.*[11]. This search lead to a set of 76 predicted GPI-anchored proteins. By combining the Big-PI positives and the proteins identified by pattern search, a total of 79 predicted GPI-anchored proteins were derived. Assignment to different protein families was then done by consulting the UniProt database in combination with BLAST searches.

Table 4 shows a list of 46 proteins, for which an assignment of either protein family or contained domains could be made. Proteins without any assignments are shown in Table 5.

Table 4: Predicted GPI-anchored proteins with family or domain assignments

UniProt-ID	Description	Family/Domains	Big-PI	Pattern
G0S879	hypothetical protein CTHT_0037870	Agglutinin-like	-	+
G0S3D9	alpha-amylase-like protein	Alpha-amylase-like	+	+
G0SAA8	hypothetical protein CTHT_0041610	Alpha-carbonic anhydrase, zinc-ion binding	-	+
G0RYL2	hypothetical protein CTHT_0007090	CFEM	+	+
G0SBA5	hypothetical protein CTHT_0049520	CFEM	+	+
G0SDR6	hypothetical protein CTHT_0052730	CFEM	+	-
G0S3S8	hypothetical protein CTHT_0030500	CFEM	+	+
G0S002	hypothetical protein CTHT_0008240	CFEM, Mad1-like	+	+
G0S223	hypothetical protein CTHT_0015720	ChpA-C/DUF320	+	+
G0SEJ6	putative covalently-linked cell wall protein	Contains PIR-repeat	+	+
G0S1Y6	hypothetical protein CTHT_0015310	Cupredoxin	+	+
G0S9D8	hypothetical protein CTHT_0045490	Cupredoxin	-	+

G0SEF6	putative cell wall protein	Ecm33	+	+
G0SEN2	hypothetical protein CTHT_0064350	Endonuclease/exonuclease/phosphatase-like	-	+
G0SEQ3	hypothetical protein CTHT_0064570	FAD-binding, false positive result?	-	+
G0S7F5	hypothetical protein CTHT_0027960	Ferritin-like superfamily, Rds1	-	+
G0SG17	hypothetical protein CTHT_0064700	GH catalytic core, ASL-like	-	+
G0S4P0	hypothetical protein CTHT_0023010	GH16	+	+
G0SFX7	putative cell wall protein	GH16	+	+
G0S5R2	hydrolase-like protein	GH16, ConA-like domain	+	+
G0SCM1	putative cell wall protein	GH16, LamG superfamily	+	+
G0SA20	cell wall glucanase-like protein	GH16, LamG-superfamily	+	+
G0SFR4	hypothetical protein CTHT_0071830	GH17	+	+
G0S1A4	hypothetical protein CTHT_0012900	GH18, chitinase, LysM-domain	+	+
G0SH28	hypothetical protein CTHT_0068470	GH45, cerato-platanin	-	+
G0S1V8	hypothetical protein CTHT_0015000	GH64, thaumatin-like	+	+
G0S6S8	1,3-beta-glucanosyltransferase-like protein	GH72/Gel1	+	+
G0S249	1,3-beta-glucanosyltransferase-like protein	GH72/Gel2	+	+
G0S7C3	chitosanase-like protein	GH75, Endo-chitosanase	-	+
G0SFA3	mannan endo-1,6-alpha-mannosidase DCW1-like protein	GH76/Dcw1	+	+
G0SFW3	putative UPF0619 GPI-anchored membrane protein	Kre9/Knh1	+	+
G0SHT5	hypothetical protein CTHT_0073300	Kre9/Knh1	+	+
G0SF37	phospholipase-like protein	Lysophospholipase	+	+
G0S1H4	aspartic-type endopeptidase-like protein	Peptidase A1 family/aspartic-type endopeptidase	+	+
G0S4R8	hypothetical protein CTHT_0023290	Peptidase A1 family/aspartic-type endopeptidase	+	+
G0S3I8	hypothetical protein CTHT_0021410	Peptidase A1/pepsin-like	+	+
G0SAA2	hypothetical protein CTHT_0041530	Peptidase A1-domain/aspartic peptidase	-	+
G0S6I1	phosphoric diester hydrolase-like protein	PLC-like phosphodiesterase, TIM beta/alpha-barrel domain superfamily	+	-
G0SDH5	phosphoric diester hydrolase-like protein	PLC-like phosphoric diesterase, TIM barrel	-	+
G0SI08	hypothetical protein CTHT_0074060	Polyampholyte	+	+
G0S1M2	hypothetical protein CTHT_0014100	SAP-like domain-containing protein/Aspartic peptidase A1 family	+	+
G0SGS6	Cu/Zn superoxide dismutase-like protein	SOD-like Cu/Zn-domain	+	+
G0S667	hypothetical protein CTHT_0034360	SUN family	+	+
G0S3B5	hypothetical protein CTHT_0020420	SurE-like	+	+
G0SAZ2	hypothetical protein CTHT_0048310	Tetratrico peptide repeat	-	+
G0SDV4	hypothetical protein CTHT_0053120	Wsc-domain	+	+
G0RXT8	guanyl-specific ribonuclease-like protein	false positive result?	+	+

Table 5: Uncharacterized or unknown predicted GPI-anchored proteins

UniProt-ID	Description	Big-PI	Pattern
G0SDD9	putative structural constituent of cell wall protein	+	+
G0RXW9	hypothetical protein CTHT_0004570	+	+
G0S179	hypothetical protein CTHT_0012640	+	+
G0S4A7	hypothetical protein CTHT_0039500	+	+
G0S348	hypothetical protein CTHT_0019640	-	+
G0S193	hypothetical protein CTHT_0012780	+	+
G0S5B3	hypothetical protein CTHT_0024200	+	+
G0S5C3	hypothetical protein CTHT_0024300	+	+
G0S759	hypothetical protein CTHT_0027530	+	+
G0S4Y9	hypothetical protein CTHT_0032150	+	+
G0S6P8	hypothetical protein CTHT_0035150	+	+
G0S6S2	hypothetical protein CTHT_0035400	+	+
G0S8L8	hypothetical protein CTHT_0038540	+	+
G0S8N5	hypothetical protein CTHT_0038740	+	+
G0S8Q3	hypothetical protein CTHT_0039950	+	+
G0S9L3	hypothetical protein CTHT_0046300	+	+
G0SBG4	hypothetical protein CTHT_0050180	+	-
G0SDX5	hypothetical protein CTHT_0053340	+	+
G0SDZ7	hypothetical protein CTHT_0053570	+	+
G0SBN8	hypothetical protein CTHT_0054240	+	+
G0SBT2	hypothetical protein CTHT_0054690	+	+
G0SCA5	hypothetical protein CTHT_0056530	+	+
G0SCN2	hypothetical protein CTHT_0057830	+	+
G0SF62	hypothetical protein CTHT_0060930	+	+
G0SHI8	hypothetical protein CTHT_0070170	+	+
G0SI03	hypothetical protein CTHT_0074010	+	+
G0S306	hypothetical protein CTHT_0019200	-	+
G0S609	hypothetical protein CTHT_0025610	-	+
G0S671	hypothetical protein CTHT_0034410	-	+
G0SCW3	hypothetical protein CTHT_0058590	-	+
G0SFJ0	hypothetical protein CTHT_0071010	-	+
G0S0P3	hypothetical protein CTHT_0010740	+	+

The set of predicted GPI-anchored proteins in *C. thermophilum* contains a variety of commonly encountered cell wall proteins, e. g. an agglutinin-like protein, proteins containing a CFEM domain, several members of GH-families, Ecm33, and a member of the SUN-family[11,12].

4. 1. 2. Proteomic analysis of isolated C. thermophilum cell walls

The prediction of GPI-anchored proteins poses a useful tool to generate an overview of the cell wall proteome and the families represented therein. However, data must be interpreted with some reservations, as it may contain false positive or false negative results. Additionally, it is based on genomic data, thus proteins without any proteomic evidence are included as well. To obtain a more realistic picture of the *C. thermophilum* cell wall proteome, cell wall isolates were analyzed by MS/MS analysis after digestion with proteases (trypsin and LysC). Data analysis was done with Proteome Discoverer 2.4 (*ThermoFisher*), using *SEQUEST* as a search engine and the *C. thermophilum* proteome and a list of common contaminations found in MS samples as search libraries.

Because significant differences between samples were observed in previous measurements, three samples were measured to ensure high quality of results. The quality of the three samples was found to be consistent. Sample 1 contained 44 proteins that met the quality criteria employed for data analysis. 14 of those were identified to be contaminants, including 9 proteins from other cellular components. This results in the identification of 30 potential cell wall proteins. Sample 2 contained 41 proteins, with 10 contaminants (5 coming from other cell organelles) and 31 cell wall proteins. 46 proteins were identified in sample 3, 15 of those were classified as contaminants (9 from other cellular components) and 31 as cell wall proteins. In total, 34 potential cell wall proteins were identified in the samples, 26 of those were found in all samples. The differences between the cell wall samples is highlighted in Figure 8, a list of the identified proteins can be found in Table 6.

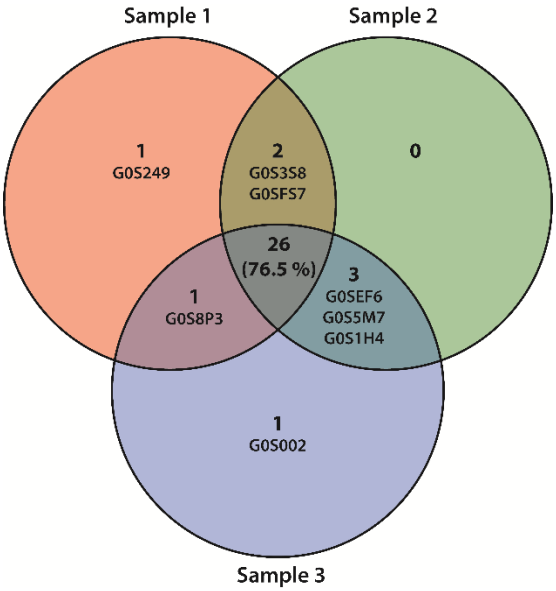


Figure 8: Venn diagram of *C. thermophilum* cell wall samples

The diagram shows the amount of proteins identified in each sample. 26 proteins were found in all three samples. One protein was identified only in sample 1 (UniProt-ID indicated in the figure), and one only in sample 3. Two proteins were found in samples 1 and 2, but not in sample 3, three proteins in sample 2 and 3, but not in sample 1. G0S8P3 was found in sample 1 and 3, but not in sample 2.

Table 6: List of proteins identified in isolated *C. thermophilum* cell walls (sorted by Sequest HT score)

UniProt-ID	Description (UniProt)	GPI predicted	Family/Domains/Orthologues
G0SDK5	Endo-1,3(4)-beta-glucanase-like protein	-	GH16, peptidase M48 and ConA-like domain
G0SEU4	Hydrolase-like protein	-	GH17
G0RZV2	SH3b domain-containing protein	-	GH24, endolysin T4 type, lysozyme-like, SH3-like bact type
G0SFR4	Uncharacterized protein CTHT_0071830	+	GH17
G0SDZ7	Uncharacterized protein CTHT_0053570	+	
G0SH48	1,3-beta-glucanosyltransferase	-	GH72, X8 domain, probably anchored to PM via helix
G0SFX7	Putative cell wall protein	+	GH16, ConA-like domain
G0SA20	Glycosidase	+	GH16, LamG-superfamily
G0S5W8	LysM domain-containing protein	-*	Cyanovirin-N, Gly zipper, LysM domain – probable adhesin
G0S763	Uncharacterized protein CTHT_0027570	-	**Bys1**, osmotin/thaumatin-like
G0RZV3	Uncharacterized protein CTHT_0004320	-	SH3b-like bac type, peptidase C51, CHAP domain
G0SCM1	Glycosidase	+	**Crh1**, GH16, ConA-like domain
G0SG36	SH3b domain-containing protein	-	Hcy domain, SH3b domain, Papain-like - similar to NlpC/P60-like protein
G0SD45	Probable alpha/beta-glucosidase agdC	-	GH31, Gal mutarotase
G0SF37	Lysophospholipase	+	PLA2c
G0SBL0	Glyoxal oxidase-like protein	-	5 x Wsc-domain, galactose oxidase

G0S9L3	Uncharacterized protein CTHT_0046300	+	
G0SFW3	Putative UPF0619 GPI-anchored membrane protein	+	**Kre9/Knh1**
G0S2U2	C3H1-type domain-containing protein	-	contains C3H1-type Zn-finger domain
G0S1A4	Chitinase	+	GH18, Chitinase, LysM-domain
G0SA61	Uncharacterized protein CTHT_0041120	-	6-blade b-propeller TolB-like, quinoprot gluc/sorb DH
G0SB94	Exo-1,4-beta-D-glucosaminidase	-	GH2, Mannosidase, Ig GlcNase
G0RZA2	Glucoamylase	-	6-hairpin glycosidase, CBM20, GH15
G0S8P3	Serine-type endopeptidase-like protein	-	Fn3, Peptidase S8/S53, subtilisin - annotated as cell wall protein in the UniProt
G0SCA5	Uncharacterized protein CTHT_0056530	+	
G0S3S8	CFEM domain-containing protein	+	**CFEM**
G0SFS7	Uncharacterized protein CTHT_0071970	-	similar to *Neurospora crassa* Acw12
G0S6S8	1,3-beta-glucanosyltransferase	+	**GH72/Gel1**
G0S3D9	Alpha-amylase	+	Alpha-amylase
G0S249	1,3-beta-glucanosyltransferase	+	**GH72/Gel2**
G0SEF6	Putative cell wall protein CTHT_0063570	+	**Ecm33**
G0S5M7	Catalase	-	Catalase class 2
G0S1H4	Aspartic-type endopeptidase-like protein	+	Peptidase A1 family/aspartic-type endopeptidase
G0S002	CFEM domain-containing protein	+	**CFEM/Mad1**

* no signal peptide predicted by SignalP

The GPI anchor signal sequence was predicted using the Big-PI Fungal Predictor and the pattern search.

In total, 17 of the predicted proteins were identified in the cell wall isolates. At this point, it has to be considered that not all GPI-anchored proteins are associated to the carbohydrate moiety of the cell wall, some remain at the plasma membrane (e. g. Dcw1). The prediction does not include sorting signals in the ω- region of the GPI-attachment site[12], hence identification of all predicted proteins in the cell wall isolates should not be expected. Interestingly, the analysis of cell wall isolates also revealed 17 proteins that were not included in the prediction. These proteins have a signal peptide, but no GPI anchor attachment sequence was detected by the Big-PI Fungal Predictor - with the exception of G0S5W8, for which no signal peptide was predicted either.

4. 1. 3. Imaging of C. thermophilum cell walls

To provide first insights into the structure of the *C. thermophilum* cell wall, cells were imaged using TEM. Well-grown mycelium from liquid cultures was used for imaging; sample preparations were done by Dr. Thomas Heimerl from the Synmikro Electron Microscopy Facility. Selected images are shown in Figure 9 and Figure 10.

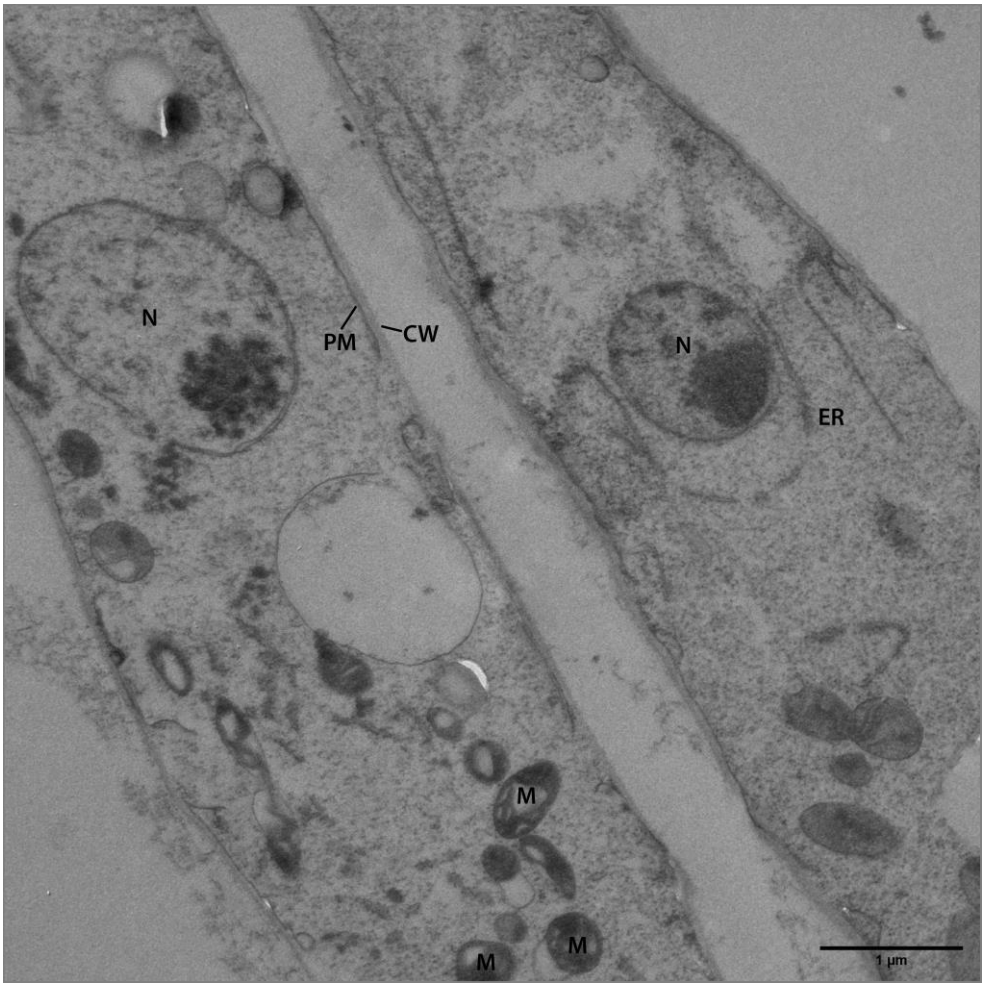


Figure 9: TEM image of *C. thermophilum*

C. thermophilum was imaged using TEM; the image shows fungal hyphae. The identified cellular components are labelled as follows: N = nucleus, PM = plasma membrane, CW = cell wall, ER = endoplasmic reticulum, M = mitochondria.

Several organelles can be identified in the TEM images of *C. thermophilum*, including the nucleus, mitochondria, the endoplasmic reticulum, the plasma membrane and the cell wall. Further components could not be clearly identified and therefore remained unlabeled in Figure 9. The diameter of both hyphae shown were measured using the image analysis software *Fiji*[109], revealing a diameter of ca 2.6 µm.

During sample preparation, the cell wall is partly detached from the plasma membrane. The cell wall is therefore not visible in parts of the image. A closer look on the *C. thermophilum* cell wall is provided in Figure 10, which provides insight into the cell wall structure.

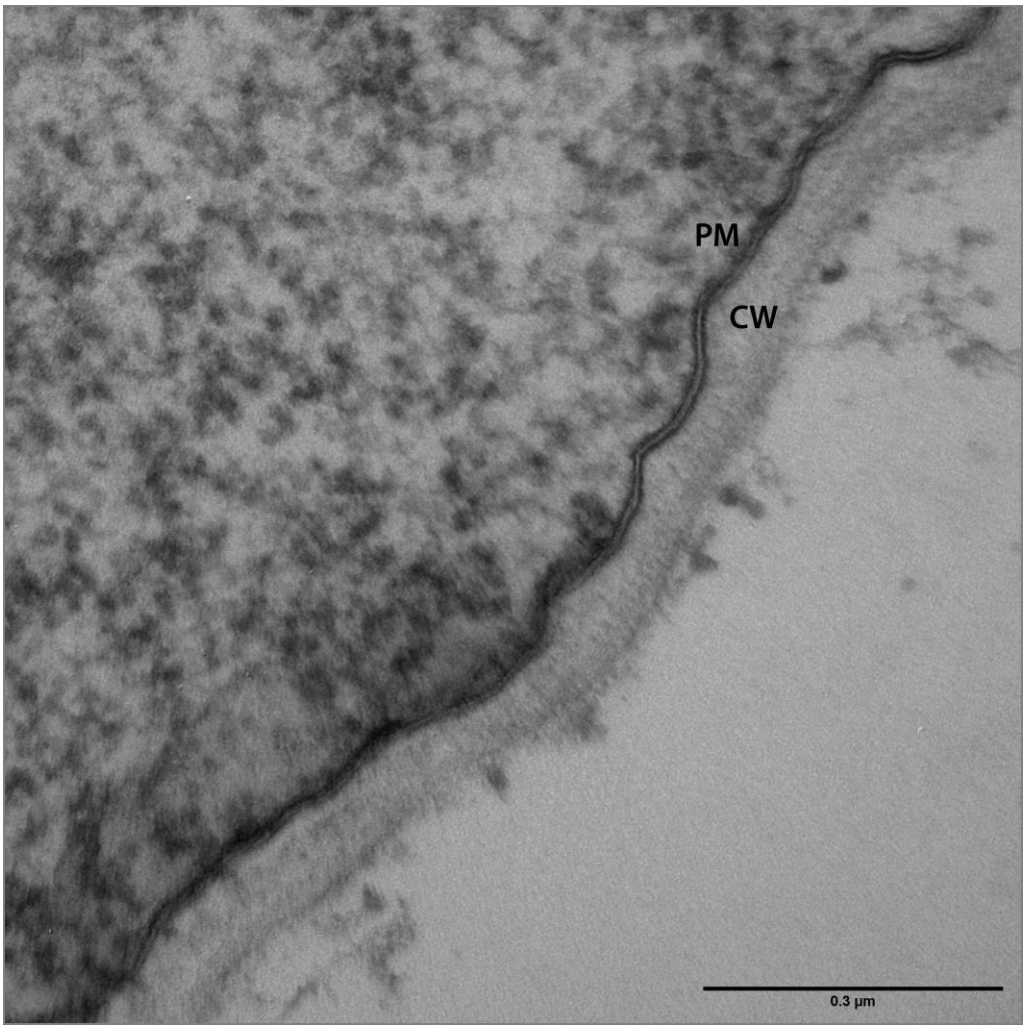


Figure 10: TEM image of the *C. thermophilum* cell wall

An image of the *C. thermophilum* cell wall (CW) reveals the two layers of the cell wall: the less electron dense inner wall is mainly composed of the β-1,3-glucan moiety of the cell wall; the outer more electron dense layer is mainly composed of mannoproteins. The plasma membrane (PM) appears as a very electron dense double layer.

The two layers of the wall, which are described in literature, can be recognized in TEM images of *C. thermophilum*. The carbohydrate-rich inner part of the cell wall appears less electron dense (i. e. lighter) than the protein-rich outer part. The plasma membrane is visible as a bilayer with very high electron density. A cell wall thickness of ca 75 nm was measured.

4. 2. Analysis of cluster VI adhesins from *C. glabrata*

4. 2. 1. Functional classification of Awp's based on a SSN

The SSN can be used as a tool for the identification of isofunctional subfamilies within a set of similar sequences. Sequences within the network are represented as "nodes", their relationship to each other is shown by lines connecting those, referred to as "edges". The similarity between sequences within the network is assessed via all-by-all BLAST. An E-value is specified by the user as a cut-off for drawing edges. This leads to the formation of clusters of nodes that represent protein subfamilies. Additionally, the information gained from a SSN often allows the identification of orthologues, which could not be clearly assigned using a BLAST search alone[110]. In this context, the SSN was used as a tool to re-evaluate previous classifications of Awp's into different subfamilies and to incorporate proteins from other organisms in the analysis and thereby exhibit possible orthologues. An SSN was created that used the β-helical region of the Awp1 and Awp3b A-domains (see below) as search templates for iterative PSI-BLAST searches. After 10 rounds, the aligned sequences were combined and redundant sequences were removed, resulting in a total of 11737 sequences that served as an input for the SSN. Initial analysis was done with a E-value cut-off of 10^{-5}, which was then decreased to 10^{-20} and 10^{-25}, respectively, for edge removal. The resulting SSN contains 4507 nodes with a pair-wise sequence identity greater than 80% for each node.

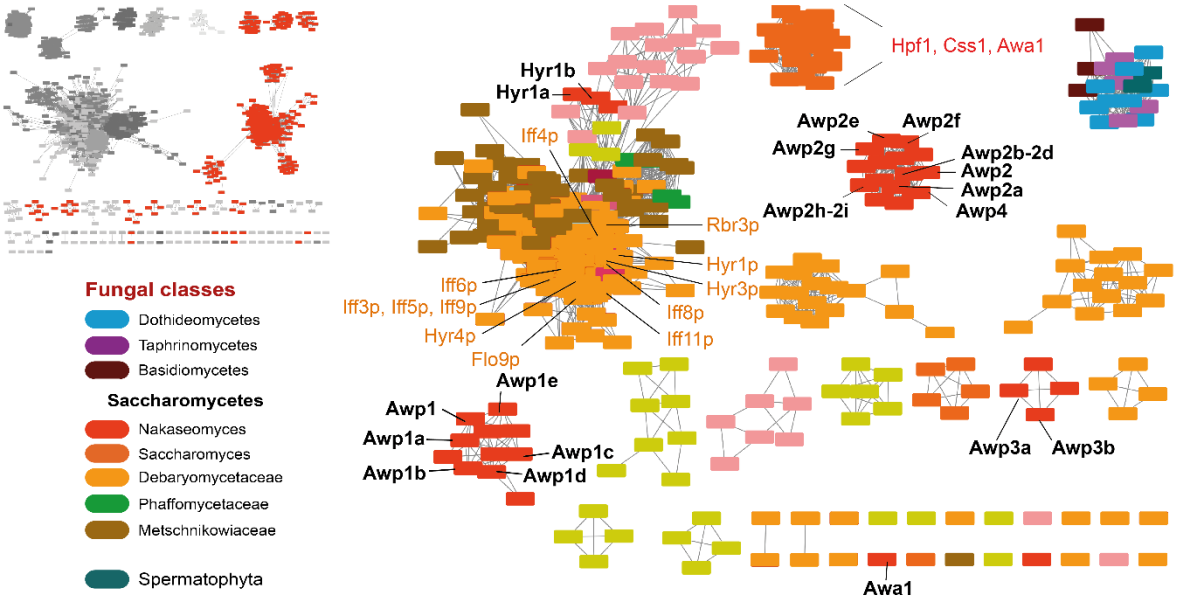


Figure 11: Classes of fungal cell wall proteins of the Awp1/Hyr1/Hpf1-type

The upper left panel shows the SSN of orthologues of Awp1, Awp3a, Awp3b and Awp4 (E-value cut-off 10^{-20}), which were identified via 10 repeated rounds of PSI-BLAST and then merged. Bacterial classes are shown in different shades of grey; fungal classes are colored red. Fungal classes (color scheme shown on the left) are enlarged in the right panel. Awp1 and Awp3b are located in two different clusters. Also, the Iff family of adhesins forms a large cluster, as well as the numerous paralogs of Awp2.

The SSN consists of protein sequences from bacteria (colored grey in the upper left panel of Figure 11) and fungi (shown in red in the same panel). Among the 11737 sequences in the network (4507 nodes), 625 sequences (445 nodes) are of fungal origin. The majority of fungal sequences are from *Saccharomycetes* (colored in different shades of orange and green in Figure 11). Awp1/3 orthologues from *Dothideomycetes*, *Taphrinomycetes* and *Basidiomycetes* form their own subcluster. The largest cluster is the IFF4 subcluster, containing several members of the Iff family of adhesins, as well as Hyr1 and Hyr3 paralogs. *S. cerevisiae* Hpf1, Css1 and Awa1 form a common subcluster. Concerning the *C. glabrata* Awp proteins, Awp1 and Awp3 form their own subclusters. The Awp1 subcluster contains 17 paralogs of Awp1 (17 sequences forming 11 nodes) and the Awp3 subcluster is made up from four nodes (five sequences). Interestingly, Awp2 and Awp4 are members of the same subcluster, consisting of 15 nodes (31 sequences).

4. 2. 2. Structural analysis of Awp3A

4. 2. 2. 1. Cloning, expression and purification of Awp3A

The Ser/Thr rich region of adhesins is subject to heavy glycosylation and therefore expected to be a flexible region. To increase the chance of crystallization, only the A-domain of Awp3 (Awp3A) was expressed in *E. coli* and used for further experiments. A plasmid containing the domain was received by Prof. Dr. Piet de Groot (pRSETa-Thr_Awp3A). Because some features of pET28a(+) were regarded more favorable than certain features of the pRSETa vector (e. g. kanamycin resistance instead of ampicillin resistance), Awp3A was cloned into pET28a(+). This was achieved by digestion of pRSETa-Thr_Awp3A with *Bam*HI and *Hind*III, followed by ligation into the pET28a(+) vector, which was also digested with named restriction enzymes beforehand. The resulting recombinant Awp3A contains an N-terminal His_6-Tag to facilitate IMAC that is removable by thrombin cleavage. Theoretical properties of Awp3A were computed using the *ProtParam* tool (accessible via the ExPASy server)[59].

Name	UniProt-ID	Native amino acid range	Length	pI	MW	Extinction coefficient (280 nm)*
Awp3A	A0A6C0A1R4	20 – 345	360 aa	5.67	38.7 kDa	28.225 mM^{-1} cm^{-1}

* assuming all cysteine residues form disulfide-linked cystines

The expression strain *E. coli* SHuffle T7 Express was included in the test expression experiment, because the sequence of Awp3A contains 6 cysteine residues that may form disulfide bonds. The strain is engineered to support formation of disulfide bonds in the cytoplasm of the cells[111] and proved to be well suited for production of Awp3A. The protein was overproduced for 72 h at 12 °C; expression was induced with 0.1 mM IPTG.

For purification, 2 – 8 L liquid culture were used, depending on the application. Cells were lysed either by sonication or with the microfluidizer. The lysate was cleared by centrifugation, sterile-filtered and applied on a 5 mL Ni-NTA column. Washing steps with phosphate buffer containing 10 and 20 mM imidazole were done and the protein was eluted with buffer containing 500 mM imidazole. Resulting fractions, analyzed by SDS-PAGE, are shown in Figure **12**. For further purification by SEC, the elution fraction from Ni-NTA chromatography was concentrated and applied on either a HiLoad 26/600 Superdex 200pg column (320 mL column volume) or a HiLoad 16/600 Superdex 200pg column (120 mL column volume), depending on the expected quantity of Awp3A. SDS-PAGE analysis of the SEC, as well as the chromatogram, are depicted in Figure 12. Pure fractions from SEC were pooled and brought to the concentration required for further experiments, usually via concentrating the sample. Large scale expression of Awp3A resulted in a yield of approximately 6 mg of pure protein per L of culture.

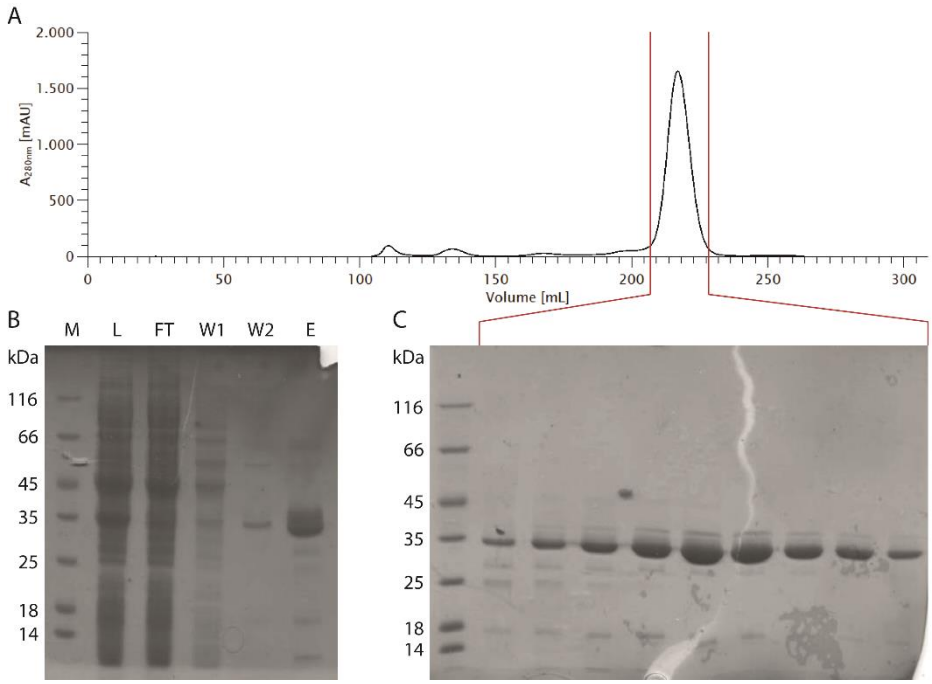


Figure 12: Purification of Awp3A

A) The SEC chromatogram of the Ni-NTA elution fraction is shown. B) 12% SDS-PAGE analysis of the Ni-NTA purification of Awp3A. M: marker, L: lysate, FT: flow-through, W1: wash 1 (10 mM imidazole), W2: wash 2 (20 mM imidazole), E: elution (500 mM imidazole) C) The SDS-PAGE analysis of the SEC purification of Awp3A is pictured. A red marking indicates the fractions that have been used for the SDS-PAGE.

4. 2. 2. 2. Crystallization, soaking and structure solution

Crystal growth could be observed at a protein concentration of 24 mg/mL in 0.2 M magnesium chloride, 0.1 M Tris pH 7.0, 2.5 M sodium chloride after two to three weeks of incubation at 18 °C. The condition was optimized using a hanging-drop vapor diffusion setup, resulting in larger crystals (see Figure 13). Awp3A crystals were harvested from an optimized condition containing 0.2 M magnesium chloride, 0.1 M Tris pH 7.0, 3.0 M sodium chloride. As no protein with over 30% sequence identity could be found in the PDB, SAD phasing was chosen as an approach to solve the structure of Awp3A. Therefore, some crystals were soaked in Gd(III) acetate before taking them to the synchrotron. Awp3A crystals were also observed in an initial crystallization screen, growing in 0.1 M sodium phosphate, 0.1 M potassium phosphate, 0.1 M MES pH6.5, 2.0 M sodium chloride after several months. These were directly frozen without any additional cryoprotection and taken to the ESRF.

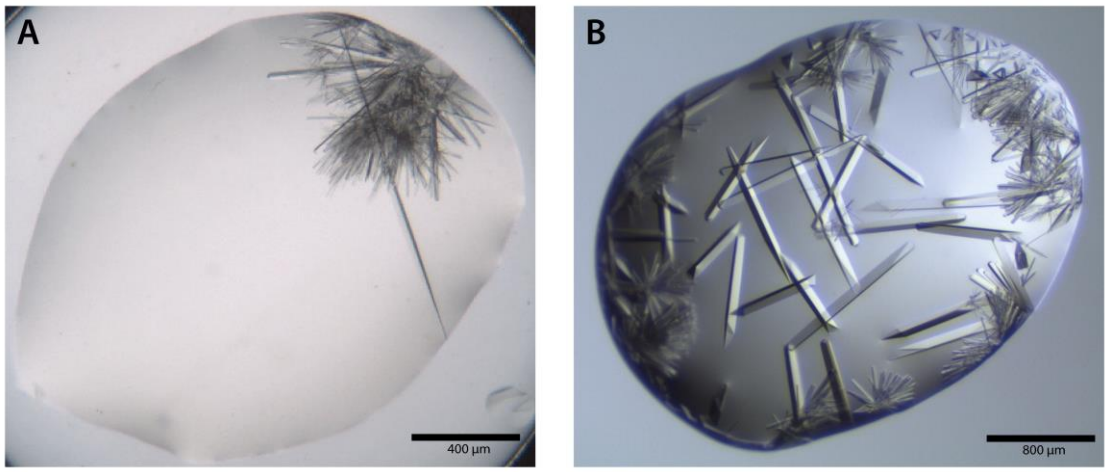


Figure 13: Awp3A crystals

A) Crystals of Awp3A that grew in a sitting-drop vapor diffusion setup at 18 °C in 0.2 M magnesium chloride, 0.1 M Tris pH 7.0, 2.5 M sodium chloride after few weeks. B) The optimized crystallization condition: a hanging-drop vapor diffusion setup with larger drop size (1.2 µL) was used, crystals were grown in 0.2 M magnesium chloride, 0.1 M Tris pH 7.0, 3.0 M sodium chloride at 20 °C.

Awp3A crystallized in space group H 3 2 with following unit cell constants: a = b = 147.34 Å, c = 117.44 Å, α = β = 90°, γ = 120°. Gd-soaked crystals diffracted to a resolution of 2 Å; processing was done in *iMOSFLM*[112], scaling and data reduction were done in *AIMLESS*[95], run in the *CCP4i2* software suite[51]. Crystallographic phases of the SAD dataset were determined using the *CRANK2* pipeline[101], followed by refinement in *REFMAC5*[107] and model building in *ARP/wARP*[55]. The structure was further refined in *Coot*[53] and *PHENIX*[52]. The Gd-derivate of Awp3A is referred to as Awp3A-Gd hereafter. Data collection and refinement statistics are shown in Table 7.

Table 7: Data collection and refinement statistics of Awp3A and Awp3A-Gd (values in the parenthesis are for the outer shell)

	Awp3A	**Awp3A-Gd**
Dataset name	2017_06_25-CC189A_x3	2017_06_25-CC213A_x2
Data collection		
X-ray source	ESRF, ID23-1	ESRF, ID29
Wavelength (Å)	0.97625	1.71237
Space group	*R* 3 2	*R* 3 2
Unit cell parameters (Å)	*a* = *b* = 147.97, *c* = 117.77	*a* = *b* = 144.4, *c* = 113.95
Resolution range (Å)	53.51 - 1.55 (1.61 - 1.55)	27.41 - 1.99 (2.06 - 1.99)
Total no. of reflections	134731 (13321)	62641 (6200)
No. of unique reflections	69278 (6889)	31321 (3100)
R_{merge} (%)	3.627 (42.99)	3.672 (12.86)
$I/\sigma(I)$	10.68 (1.84)	18.42 (4.90)
Completeness (%)	96.96 (97.30)	99.92 (100.00)
Multiplicity	1.9 (1.9)	2.0 (2.0)
$CC_{1/2}$	0.999 (0.431)	0.997 (0.924)
Refinement		
R_{work}/R_{free} (%)	15.93/18.79	19.03/22.78
No. of atoms	3117	2658
Average *B* factor (Å^2)	28.78	37.13
R.m.s. deviations		
Bond length (Å^2)	0.008	0.014
Bond angles (°)	0.99	2.02
Ramachandran plot (%)		
Favoured	96.93	96.68
Allowed	3.07	2.99
Outliers	0.00	0.33
Rotamer outliers (%)	0.35	3.09

The asymmetric unit of Awp3A-Gd contains one molecule of Awp3A and 42 Gd^{3+} ions (see Figure 14). The A-domain of Awp3b consists of 33 β-strands and a short α-helix between strands 31 and 32. It can be divided into two domains: a parallel right-handed β-helix with three faces, and an α-crystallin domain. Due to uninterpretable electron density, the following residues could not be modelled in the structure of Awp3A-Gd: S75 – D82 and S320 – E322. These gaps are both located in loop regions.

Crystals of Awp3A obtained in 0.1 M sodium phosphate, 0.1 M potassium phosphate, 0.1 M MES pH6.5, 2.0 M sodium chloride diffracted to 1.55 Å resolution. The structure was solved in *Phaser*[105], using Awp3A-Gd as a model. Iterative cycles of real space and reciprocal space refinement were done in *WinCoot*[53] and via *phenix.refine*[52]. In contrast to the heavy atom derivate (Awp3A-Gd), the electron density of the native structure of the Awp3 A-domain (Awp3A) is clearly defined in all parts of the structure. The conformation of some loops is different in both structures, indicating flexibility of these regions.

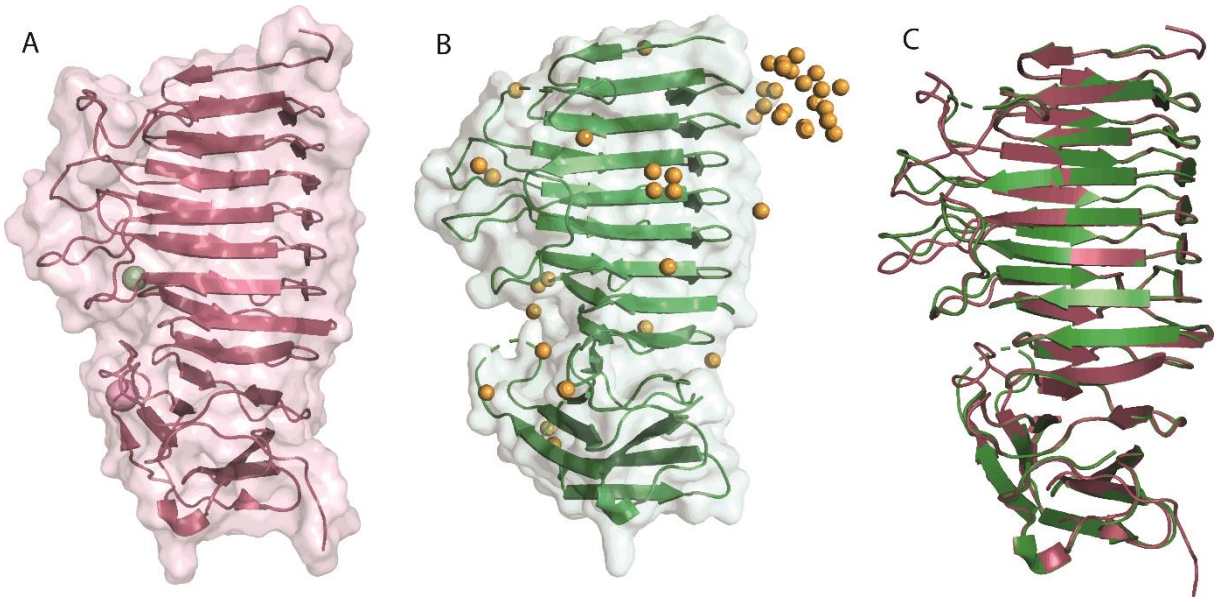


Figure 14: Comparison between native Awp3A and the Gd-derivate

A) Cartoon and surface representation of Awp3A. Na (shown in light pink) and Cl (colored green) are bound to the structure. B) Cartoon and surface representation of Awp3A-Gd; the 42 Gd^{3+} ions are represented as orange spheres. C) Comparison between the protein moieties of the two structures. Awp3A is shown in red, Awp3A-Gd is shown in green. The structures are highly similar to each other, with minor changes in the conformation of some loops. Additionally, a few residues (S75 – D82 and S320 – E322) could not be modelled in Awp3A-Gd due to unclear electron density.

4. 2. 3. Structural analysis of Awp1A

4. 2. 3. 1. Cloning, expression and purification of Awp1A

A plasmid containing the A-domain of Awp1 (Awp1A) was received from Prof. Dr. Piet de Groot (pRSETa-Thr_Awp1A). Just as Awp3A, Awp1A was cloned into pET28a using the restriction enzymes *Bam*HI and *Hind*III. The recombinant Awp1A with an N-terminal His_6-Tag was produced in *E. coli* SHuffle T7 Express at 12 °C for 72 h; induction was done with 0.1 mM IPTG. The theoretical properties of Awp1A were computed with *ProtParam*[59].

Name	UniProt-ID	Native amino acid range	Length	pI	MW	Extinction coefficient (280 nm)*
Awp1A	Q6FPN0	18 – 325	341 aa	4.94	35.7 kDa	16.515 mM^{-1} cm^{-1}

* assuming all cysteine residues form disulfide-linked cystines

2 – 8 L liquid culture were used for purification; cells were broken either by sonication or with the microfluidizer. After clearing and sterile-filtering, the lysate was applied on a 5 mL Ni-NTA

column. A washing step was done with phosphate buffer containing 30 mM imidazole and Awp1A was eluted using buffer with 250 mM imidazole. Analyzed fractions are shown in Figure 15. SEC was done as a polishing step of the purification process, using either a HiLoad 26/600 Superdex 200pg column or a HiLoad 16/600 Superdex 200pg column, depending on the expected quantity of purified protein. Fractions from SEC that contained pure Awp1A were pooled and brought to the concentration required for further experiments. A yield of approximately 7 mg pure protein per L of culture could be achieved.

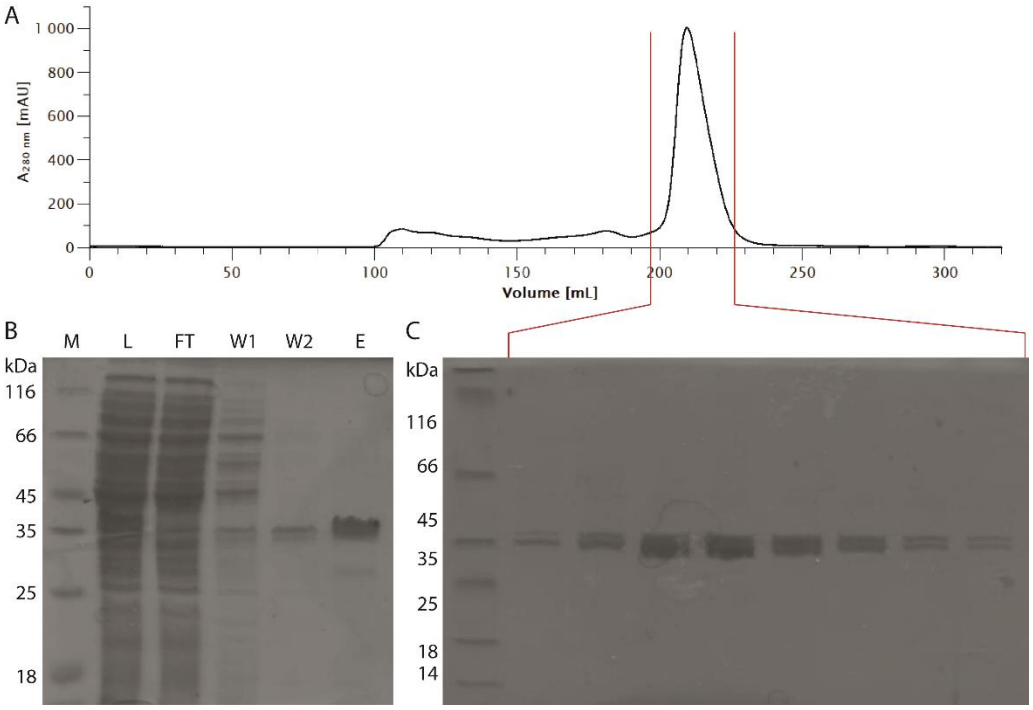


Figure 15: Purification of Awp1A

A) The SEC chromatogram for the purification of Awp1A is shown. B) The 12% SDS-PAGE analysis of the Ni-NTA purification of Awp1A. M: marker, L: lysate, FT: flow-through, W1: wash 1 (10 mM imidazole), W2: wash 2 (30 mM imidazole), E: elution (250 mM imidazole). C) The SDS-PAGE analysis of the purification of Awp1A via SEC is shown. A red marking indicates the fractions in the SEC chromatogram, which have been used for the SDS-PAGE analysis.

4. 2. 2. 2. Crystallization and structure solution

Crystals of the Awp1 effector domain grew in several conditions, using protein concentrations of 48 mg/mL and 24 mg/mL. Mainly thin needle-shaped crystals were observed, forming brushes or sea urchins in many conditions (see Figure 17). Thicker needles were harvested and taken to the ESRF for data collection. Native crystals of Awp1A diffracted to a maximum resolution of around 2.5 Å, showing moderate anisotropy. Because Awp1 and Awp3 belong to the same subfamily of putative adhesins, structural similarity was expected. Additionally, the sequence identity and similarity of the effector domains (ranging from amino acid 19 to 325

in Awp1 and from 20 to 345 in Awp3b) are 25.1% and 40.6%, respectively, indicating a sufficient resemblance for MR. However, structure solution attempts failed.

Optimization of the crystals was conducted to gain higher quality datasets. Additionally, reproduction of these well diffracting crystals for *ab initio* structure solution was expected to be required. An optimized crystallization condition for Awp1A, containing 0.1 M MOPSO/Bis-Tris pH 6.5, 10% (w/v) PEG 8000, 20% 1,5-pentanediol, 0.5 mM erbium (III) chloride, 0.5 mM terbium (III) chloride, and 0.5 mM ytterbium (III) chloride was found. Resulting crystals diffracted to a resolution of up to 1.85 Å with some anisotropy. Crystals of the Awp1 effector domain that were soaked in various heavy atom solutions. However, various attempts of heavy atom phasing did not initially result in structure determination. Most heavy metals soaked into the crystal were either not bound (no anomalous signal could be detected) or the anomalous diffraction was very weak and did not reach a high resolution (in most cases anomalous diffraction to a resolution of only 6 Å could be detected). Also anomalous diffraction from the heavy atoms that were already present in the crystallization condition (erbium, terbium, and ytterbium) could not be observed.

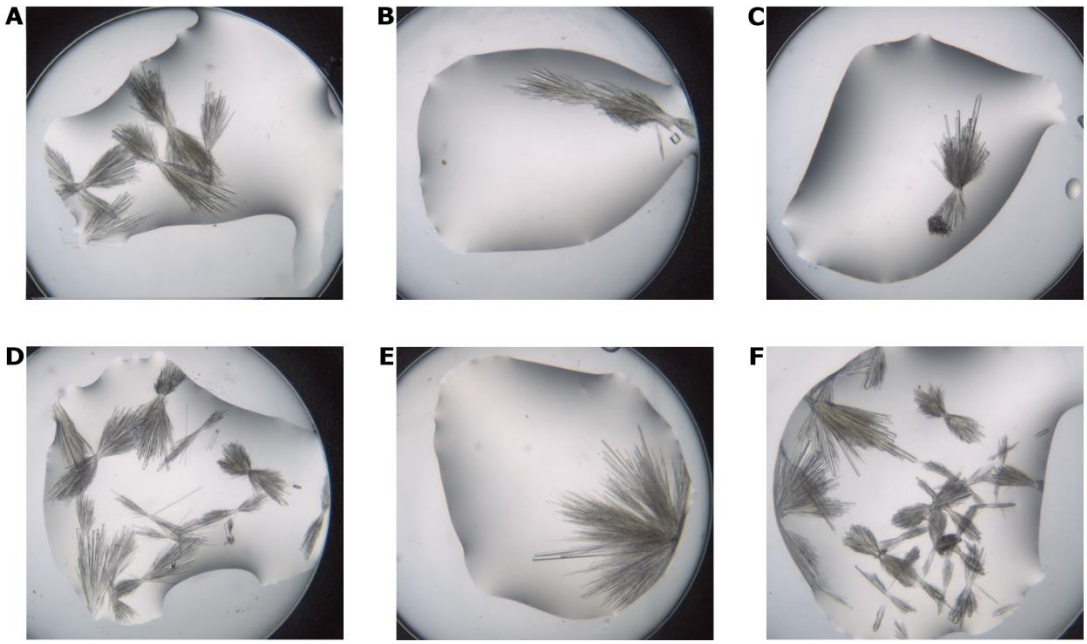


Figure 16: Crystals of the Awp1 effector domain

The effector domain of Awp1 formed needle-shaped crystals in several conditions: (A) 0.1 M MOPSO/Bis-Tris pH 6.5, 12.5% (w/v) PEG 4000, 20% (v/v) 1,2,6-hexanetriol, Amino acid II mix (1:10), (B) 0.1 M MOPSO/Bis-Tris pH 6.5, 10% (w/v) PEG 8000, 20% 1,5-pentanediol, Lanthanide mix (1:10), (C) 0.1 M MOPSO/Bis-Tris pH 6.5, 12.5% (w/v) PEG 4000, 20% (v/v) 1,2,6-hexanetriol, Lanthanide mix (1:10), (D) 0.1 M MOPSO/Bis-Tris pH 6.5, 12.5% (w/v) PEG 4000, 20% (v/v) 1,2,6-hexanetriol, Alkali mix (1:10), (E) 0.1 M HEPES pH 7.5, 10% (w/v) PEG 6000, 5% (v/v) MPD, (F) 0.1 M MOPSO/Bis-Tris pH 6.5, 12.5% (w/v) PEG 4000, 20% (v/v) 1,2,6-hexanetriol, Alkalis mix (1:10)[85] with a protein concentration of 24 mg/mL.

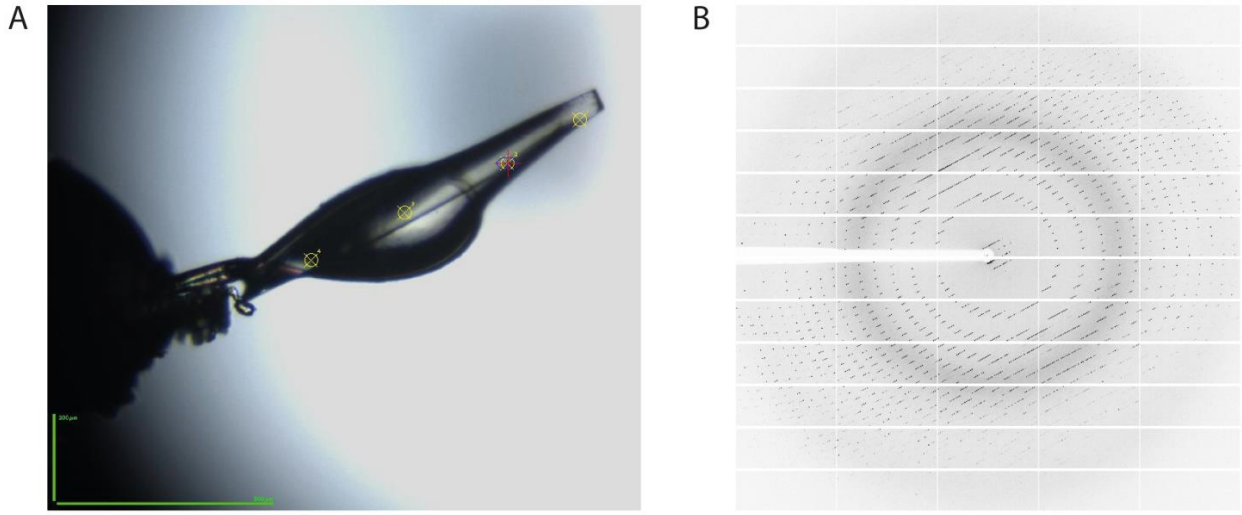


Figure 17 Crystal of the Awp1 effector domain and its diffraction image

A) A photograph of the crystal before its characterization is shown. Its diffraction pattern is shown in B), the detector resolution at the edge is 2 Å. The data collection of 2018_06_28-CC223A_x6 is shown.

Ultimately, the data collection represented in Figure 17 yielded a structure solution. The protein crystallized in space group P 4_3 2_1 2 with two molecules per asymmetric unit; the crystal diffracted to a resolution of 1.85 Å. Structure solution was done by MR in *Phaser*[105] with Awp3A as a template, resulting in an incomplete model of Awp1A, with an R_{free} of 55.1%. Even though the R_{free} does not indicate structure solution, the data was used as an input for model building using the *ARP/wARP* Web Service[55]. After 10 cycles of model building, which is the default setting in *ARP/wARP*, the R_{free} dropped to 40.9% and 46% of expected residues were build. Finally, running 10 additional cycles lead to a decrease of R_{free}/R_{work} to 26.2/22.8% and 609 of 648 amino acids were modelled. Statistics after refinement using *Coot*[53] and *phenix.refine*[52] are presented in Table 8; the structure of Awp1 is shown in Figure 18. Just as Awp3A, Awp1A consists of 33 β-strands and a short α-helix between β-strands 31 and 32. It contains a triangular right-handed parallel β-helix and an α-crystallin domain as well.

4. 2. 4. Structures of the A-domains of cluster VI adhesins Awp1 and Awp3

Both, Awp1A and Awp3A consist of a β-helix domain and an α-crystallin domain. They are structurally highly similar to each other, with a root mean square deviation (RMSD) of 1.466 Å with 1300 aligned atoms (calculated via structure-based alignment in *PyMOL*). One could presuppose this, as both proteins belong to the same subfamily of putative adhesins, i. e. cluster VI (see chapter 1.3). In this context the sequence identity and similarity of 22.1% and 32.4% (aligned via EMBOSS Needle), respectively, should be mentioned. Sequence identity and similarity of these effector domains (ranging from amino acid 19 to 325 in Awp1 and from 20 to 345 in Awp3b) are 25.1% and 40.6%, respectively. The initial difficulties in solving the structure of Awp1A were therefore unexpected.

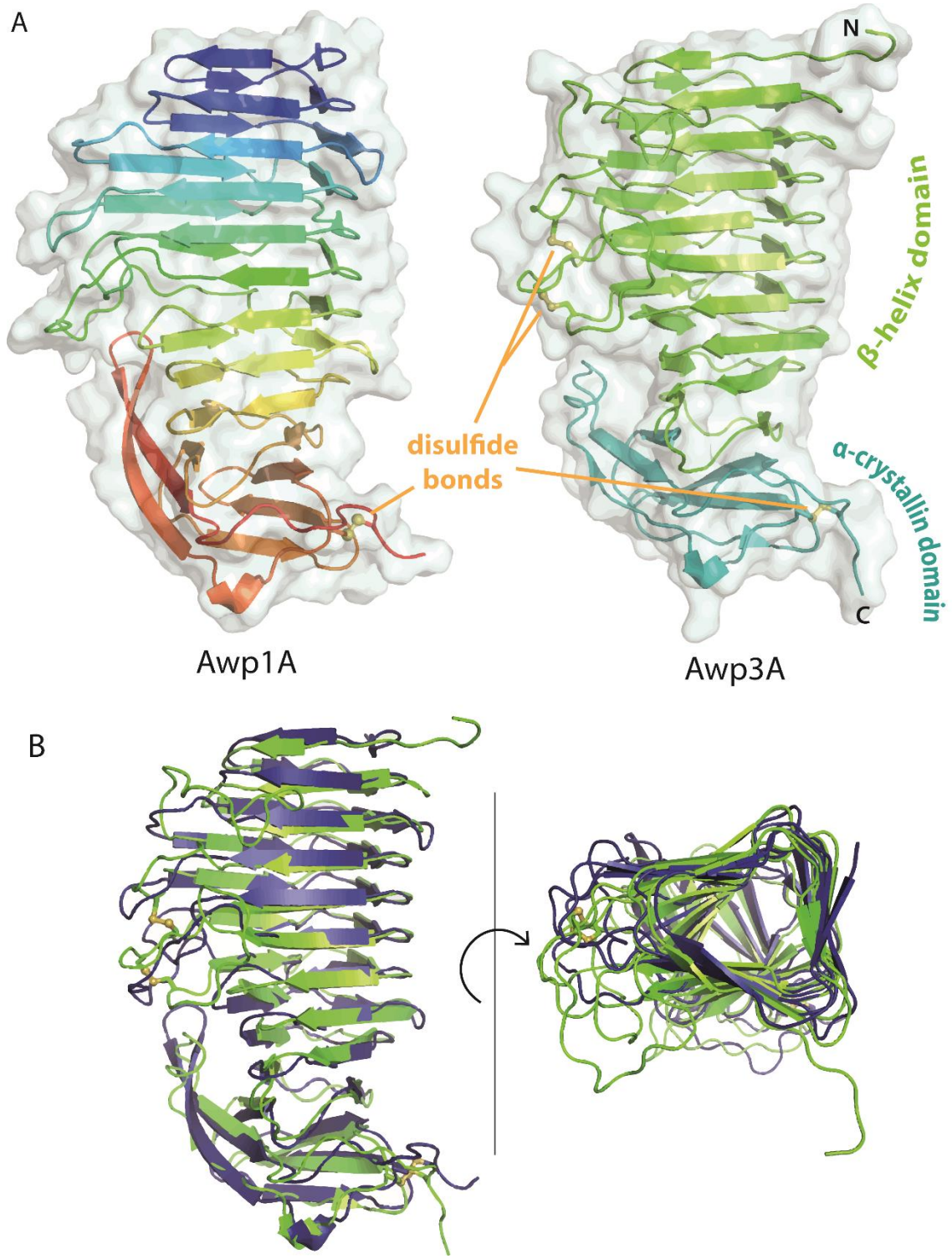


Figure 18: Overall structures of Awp1A and Awp3A

A) Structure of Awp1A is shown on the left in cartoon representation, colored in a rainbow-scheme. The N-terminus is colored blue, the C-terminus is shown in red. Disulfide bonds are indicated by yellow sticks and spheres. The two domains both proteins consist of are indicated in different colors on the right, on Awp3A. The N-terminal β-helix domain is shown in green, the α-crystallin domain is pictured in cyan. Awp3A contains 3 disulfide bonds (shown in yellow), of which only one is also present in Awp1A. B) Structural comparison between Awp1A (blue) and Awp3A (green), represented in two different orientations. The A-domains are structurally similar to each other, with an RMSD of around 1.4 Å.

Table 8: Data collection and refinement statistics for Awp1A (values in the parenthesis are for the outer shell)

	Awp1A
Dataset name	2018_06_28-CC223A_x6
Data collection	
X-ray source	ESRF, ID29
Wavelength (Å)	0.97717
Space group	P 4_3 2_1 2
Unit cell parameters (Å)	$a = b = 83.28$, $c = 274.24$
Resolution range (Å)	45.81 - 1.85 (1.92 - 1.85)
Total no. of reflections	165589 (16103)
No. of unique reflections	83156 (8101)
R_{merge} (%)	2.97 (34.94)
$I/\sigma(I)$	12.64 (1.62)
Completeness (%)	99.16 (97.48)
Multiplicity	2.0 (2.0)
$CC_{1/2}$	0.998 (0.95)
Refinement	
R_{work}/R_{free} (%)	18.79/20.83
No. of atoms	5262
Average B factor (Å^2)	51.43
R.m.s. deviations	
Bond length (Å^2)	0.004
Bond angles (°)	0.67
Ramachandran plot (%)	
Favoured	96.89
Allowed	2.62
Outliers	0.49
Rotamer outliers (%)	1.31

The major structure motif of the A-domains of Awp1 and Awp3 is the N-terminal right-handed parallel β-helix with three faces. According to the nomenclature for β-helices introduced by Yoder & Jurnak[113], the three β-strands forming a single turn are referred to as PB1, PB2, and PB3; loops between them are labeled T1 (connecting PB1 and PB2), T2 (PB2 and PB3), and T3 (PB3 and PB1 of the next turn), as indicated in Figure 19 A. Along the whole β-helix, T2 and T3 loops are very short, whereas T1 loops are more extended. Awp3A contains three disulfide bonds, of which two are placed within the latter loop regions. The third one is placed near its C-terminus and is not resolved in Awp3A-Gd. The disulfide bonds within the T1 loops are not observed in Awp1A, only the one near the C-terminus is preserved.

Both structures display several features that are well conserved in parallel β-helix proteins. Within the β-helix domains of Awp1A and Awp3A, stacks of hydrophobic amino acids can be observed (see Figure 19 B). These are not perfectly aligned but slightly offset, which is achieved by twisting the β-helix. This prevents an energetically unfavorable alignment of the

aromatic side chains, in which the π-electron clouds would repel one another. In addition, the asparagine ladder, which can be detected in the T3 turns of both structures, has also been described as a typical feature of β-helix proteins. In Awp1A, it is composed of five asparagines, where each side chain forms a hydrogen bond to the next one. Both features provide additional stability and rigidity to the β-helix[114,115]. In case of the Awp1 A-domain, further amino acid stacks can be observed on the motif's surface. These are serine/threonine ladders, which are not typical for β helix proteins (see Figure 19 D).

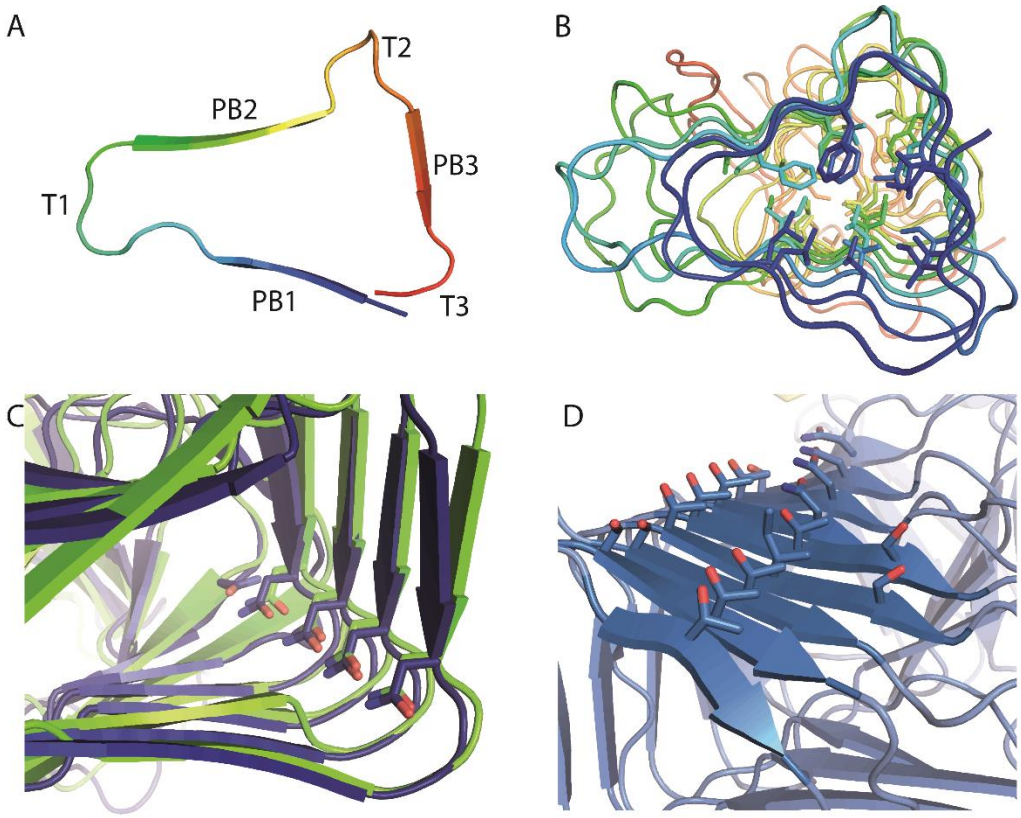


Figure 19: Features of the β-helix domains of Awp1A and Awp3A

A) A single turn of the β-helix domain is shown from above and labeled according to the standard nomenclature used for β-helices[113]. The β-strands PB1 and PB2 are packed against each other, while PB3 is placed perpendicular in relation to them. The T1 loops are elongated in both structures, whereas T2 and T3 loops are very short. B) Awp1A is depicted from the top of the β-helix. Stacking of hydrophobic residues can be observed inside the domain, involving leucine, isoleucine, and phenylalanine residues. C) A stack of asparagine residues inside the β-helix is found in both, Awp1A and Awp3A. Similar stacking interactions were also observed in other β-helix proteins[114]. D) Ladders of similar residues are also placed on the outer face of the Awp1A β-helix, namely stacks of serine, threonine, and asparagine.

4. 2. 5. Binding studies on Awp1A and Awp3A

Due to the clear structural similarity between Awp1A and Awp3A and various polysaccharide binding proteins, carbohydrate binding studies were conducted: A TSA served as quick and easy screening method for analyzing potential binding of smaller polysaccharides that were available in the lab. Additionally, both proteins were sent to the Consortium of Functional Glycomics (CFG) to analyze binding properties on a glycan array.

4. 2. 5. 1. Ligand screening via TSA

The TSA provides a convenient screening method that can be done in a short time with a relatively low amount of sample. The determination of a protein's melting temperature can serve various purposes, usually an increase of protein stability is looked for by screening the thermal stability in various buffers or in presence of potential ligands. As proteins tend to be more thermally stable when their cognate ligand is bound, a thermal shift, i. e. a shift of melting temperature, is an indication for protein-ligand-interaction[116].

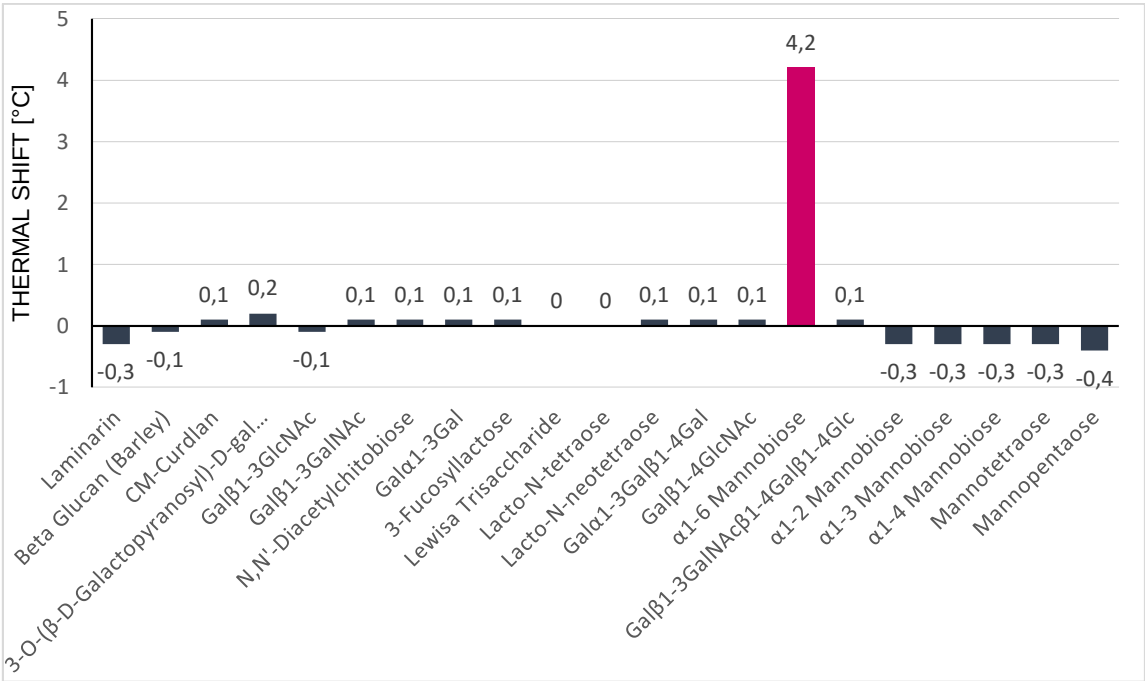


Figure 20: Results of the TSA for Awp3A

The melting temperature of Awp3A without any potential ligands added is used as a base line. The bars show the deviation of the melting temperatures of Awp3A in presence of the indicated glycan. A significant increase in melting temperature by 4.2 °C could be observed in the presence of 50 mM Manα1-6Man.

A melting temperature of 59.5 °C was measured for Awp3A in SEC buffer without any potential ligands added. In the ligand discovery experiment, addition of α-1,6-mannobiose revealed an increase in melting temperature by 4.2 °C. Other mannobiose components (α-1,2-mannobiose, α-1,3-mannobiose, α-1,4-mannobiose) did not induce any significant changes in melting temperature, which is common for glycan binding proteins, as these proteins tend to be very specific and a change in the connection of the mannose units can alter the glycan structure significantly. Interestingly, also mannotetraose and mannopentaose did not induce any significant changes in melting temperature, although both compounds contain α-1,6-mannobiose. A TSA conducted with Awp1A with the same set of potential ligands did not show any thermal shifts (Appendix IV).

The possible interaction of Awp3A and α-1,6-mannobiose was also examined via ITC, where no binding event could be observed (Appendix V). However, ITC experiments with carbohydrates may not reveal any binding, although present, because the release of ordered water molecules may compensate for the temperature change that is generated via the binding process. To further investigate a possible interaction, Awp3A crystals were soaked with highly concentrated α-1,6-mannobiose solution (1 M). In the structure obtained in this experiment, additional electron density is present (see Figure 21). However, this density can be unambiguously assigned to the His_6-Tag of Awp3A. In conclusion, the addition of α-1,6-mannobiose leads to conformational stabilization of the His_6-Tag, but no specific binding of Awp3A to this glycan could be determined.

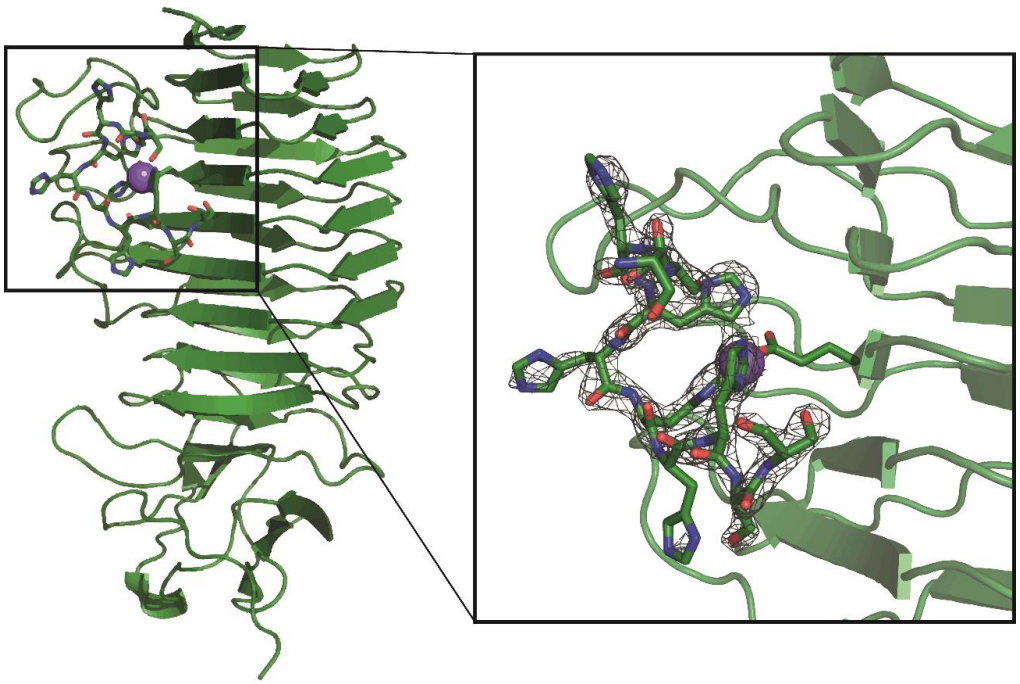


Figure 21: Structure of Awp3A, soaked with 1 M α-1,6-mannobiose

The overall structure of Awp3 is shown in cartoon style, with the ordered His_6-Tag depicted as sticks. E121 interacts with a sodium ion (purple sphere), to which the His_6-Tag is bound. The $2mF_{obs}$-DF_{calc} map at a contouring level of 2.0 σ is depicted for the His_6-Tag.

4. 2. 5. 2. Glycan array screening

Due to the high structural similarity to various glycan binding proteins and the lack of availability of a wide variety of potential ligands in the lab, the binding properties of both, Awp1A and Awp3A, were analyzed via the Mammalian Glycan Array version 5.2 from the CFG. Glycan arrays have proven to be an efficient tool for determination of ligand binding patterns of glycan-binding proteins. The method requires only a small amount of sample, while a large library of glycans can be screened against[117]. Therefore, purified protein samples with a concentration of 50 µg/mL in SEC buffer were sent to the CFG, where the experiment was conducted. Detection was done via an anti-His antibody, coupled to AlexaFluor 488, to enable detection without masking any residues that may be involved in the binding process. The data is deposited under cfg_rRequest_3531. No binding to the glycans presented on the chip was detected (Appendix VI).

4. 3. Analysis of the cluster III adhesin Awp14

4. 3. 1. Cloning, expression and purification of Awp14A

pET28b(+) containing the A-domain of Awp14 (Awp14A) has been received from Prof. Dr. Piet de Groot (pET28b_Awp14). An N-terminal His_6-Tag enables protein purification via IMAC. Theoretical properties of Awp14A were computed via *ProtParam*[59].

Name	Candida database ID	Native amino acid range	Length	pI	MW	Extinction coefficient (280 nm)
Awp14A	CAGL0L00157g	22 – 400	413 aa	5.49	45.95 kDa	45.27 mM^{-1} cm^{-1}

As Awp14A only contains one cysteine residue, formation of disulfide bonds can be excluded. Accordingly, the protein was produced in *E. coli* BL21 (DE3) Gold at 12 °C for 72 h. Induction of protein expression was done by addition of 0.1 mM IPTG. 2 – 8 L of expression culture were used for purification of Awp14A. The cells were lyzed by sonication or using the microfluidizer. The lysate was cleared and sterile-filtered and loaded on a 5 mM Ni-NTA column, equilibrated with phosphate buffer. A washing step was performed with buffer containing 15 mM imidazole, before the protein was eluted with phosphate buffer containing 250 mM imidazole. Awp14A was further purified via SEC, either with a HiLoad 26/600 Superdex 200pg column or a HiLoad 16/600 Superdex 200pg column. Fractions containing pure Awp14A were pooled and concentrated. Approximately 10 mg of pure protein were produced per L of culture.

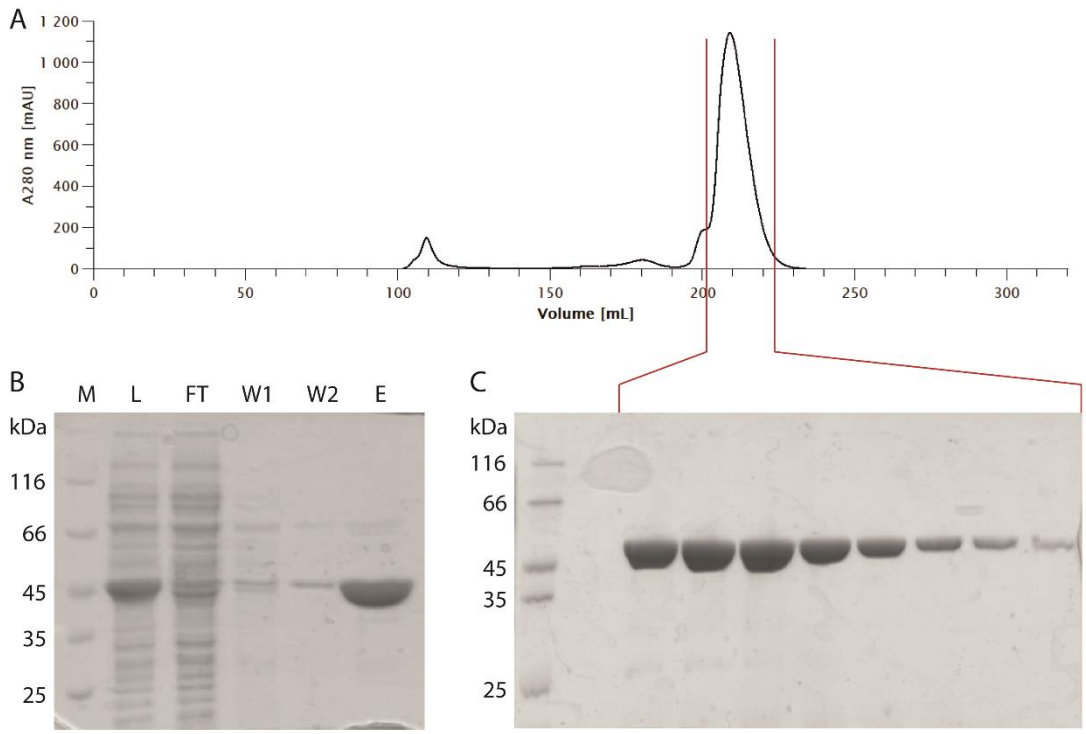


Figure 22: Purification of Awp14A

A) The chromatogram of the SEC, which was used as a polishing step for the purification of Awp14A. B) 12% SDSPAGE analysis of fractions from the Ni-NTA purification of Awp14A. M: marker, L: lysate, FT: flow-through, W1: wash 1 (10 mM imidazole), W2: wash 2 (15 mM imidazole), E: elution (250 mM imidazole). C) SDS-PAGE from the SEC purification of Awp14A. The red marking indicates the fractions in the SEC chromatogram, which have been used for the SDS-PAGE analysis.

4. 3. 2. Crystallization of Awp14A

The Awp14 A-domain crystallized at a protein concentration of 22.5 mg/mL in four different crystallization conditions, all part of the Morpheus II crystallization screen. The plates that grew in 0.1 M BES/TEA pH 7.5, 10% (w/v) PEG 8000, 20% 1,5-pentanediol, “Amino-acid II” (1:10)[85] were harvested and sent to the ESRF for data collection. The crystals diffracted to a resolution of approximately 2.5 Å, data collection statistics are given in Table 9. Awp14A crystallized in space group *C* 2 2 2_1. Calculation of the Matthews coefficient indicates a unit cell content of two molecules with a solvent content of 52.63%. Unfortunately, an appropriate model for MR is not available and the reproduction of Awp14A crystals could not be achieved in this work.

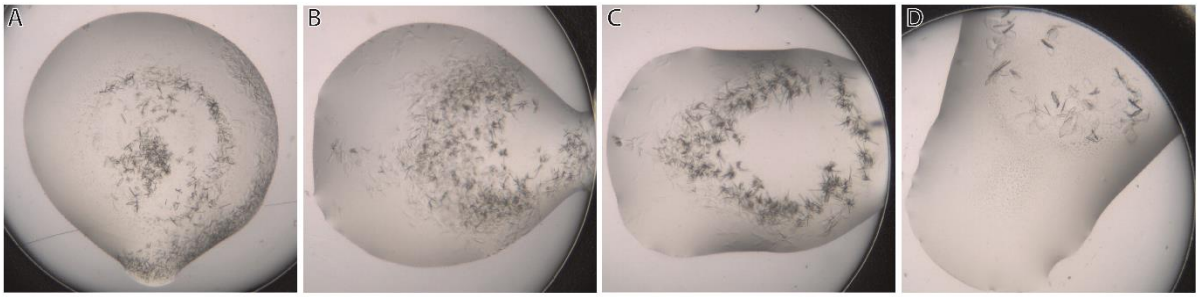


Figure 23: Awp14A crystals

Crystals of Awp14A that grew in different crystallization conditions, all part of the Morpheus II crystallization screen. Each crystallization condition in the Morpheus II screen is composed of precipitant mix, buffer mix and additive mix (1:10). The conditions leading to crystallization all contain the same mix of precipitants, namely 10% (w/v) PEG 8K, 20% (v/v) 1,5-pentanediol. A) shows the crystals in condition A3, which additionally contains a MOPSO/Bis-Tris buffer system (0.1 M MOPSO and 0.1 M Bis-Tris, mixed in a ratio that produces a pH of 6.5), as well as the "LiNaK" additive mix (0.3 M lithium sulfate, 0.3 M sodium chloride and 0.3 M potassium sulfate). B) Crystals from the condition C3, containing the MOPSO/Bis-Tris buffer mix and the "Alkalis" additive mix (10 mM rubidium chloride, 10 mM strontium acetate, 10 mM cesium acetate, 10 mM barium acetate). C) Condition G3, containing the MOPSO/Bis-Tris buffer mix and the "Amino-acids II" additive mix (0.2 M DL-arginine HCl, 0.2 M DL-threonine, 0.2 M DL-histidine HCl H2O, 0.2 M DL-5-hydroxylysine HCl, 0.2 M trans-4-hydroxy-L-proline). D) Crystals grown in condition G7, composed with a BES/TEA buffer mix (0.1 M BES/TEA pH 7.5) and the "Amino-acids II" additive mix[85].

Table 9: Data collection statistics for Awp14A

Dataset name	2017_05_17-CC172A_x6
X-ray source	ESRF, ID23-1
Wavelength (Å)	0.972
Space group	$C\,2\,2\,2_1$
Unit cell parameters (Å)	a = 78.23, b = 172.94, c = 140.99
Resolution range (Å)	46.41 – 2.5 (9.01 – 2.5)
Total no. of reflections	142870 (16343)
No. of unique reflections	32560 (3687)
R_{merge} (%)	0.276 (1.685)
$I/\sigma(I)$	5.9 (1.1)
Completeness (%)	97.8 (99.3)
Multiplicity	4.4 (4.4)
$CC_{1/2}$	0.982 (0.337)

4. 4. Analysis of the CFEM domain of the GPCR *Ct*Pth11

4. 4. 1. Cloning, expression and purification of CtPth11

First work on the GPCR Pth11 has already been conducted by Dr. Vitali Kalugin. In this work, no overproduction of soluble protein could be achieved using the CFEM-domain of *M. grisea* Pth11. The *C. thermophilum* orthologue of Pth11 was therefore identified via a SSN[41] and cloned into the vector pET28a(+). The transmembrane helices, as well as the signal peptide and a few residues predicted to be unstructured at the N-terminus of the protein, were removed for this purpose. The generated plasmid (pET28a_*Ct*Pth11) contains V24 – S105 with an N-terminal His_6-Tag to enable purification via IMAC. Theoretical properties of the domain were calculated via the ExPASy *ProtParam* tool and are as follows:

Name	UniProt-ID	Native amino acid range	Length	pI	MW	Extinction coefficient (280 nm)*
*Ct*Pth11	G0SBE2	24 – 105	105 aa	6.62	11.0 kDa	1.99 mM^{-1} cm^{-1}

* assuming all cysteine residues form disulfide-linked cystines

For overexpression of the *Ct*Pth11 CFEM-domain, the strain *E. coli* SHuffle was chosen, because the domain is proposed to contain four disulfide bonds[36]. The overproduction was done by growing the cells to an OD_{600} of approximately 0.6, induction by addition of 0.1 mM IPTG and further incubation at 18 °C for 48 h.

8 L of liquid culture were used for quantitative preparation of *Ct*Pth11 CFEM protein. The cells were broken by sonication or with the microfluidizer and the lysate was cleared via centrifugation at 18000 rpm, 4 °C. After sterile-filtering, the lysate was applied on a 5 mL Ni-NTA column, which was then washed with Ni-NTA buffer containing 20 mM imidazole. The *Ct*Pth11 CFEM domain was eluted with 500 mM imidazole. The elution fraction was concentrated and applied on a HiLoad 26/600 Superdex 75pg column, which was equilibrated with SEC buffer (in this case phosphate buffer). The fractions from the peak containing the target protein were analyzed by SDS-PAGE and pooled. The sample was concentrated and glycerol was added to a final concentration of 10% (v/v). The purified protein sample was then divided into several 1.5 mL Eppendorf cups, shock-frozen in liquid nitrogen and stored at -80 °C for further use.

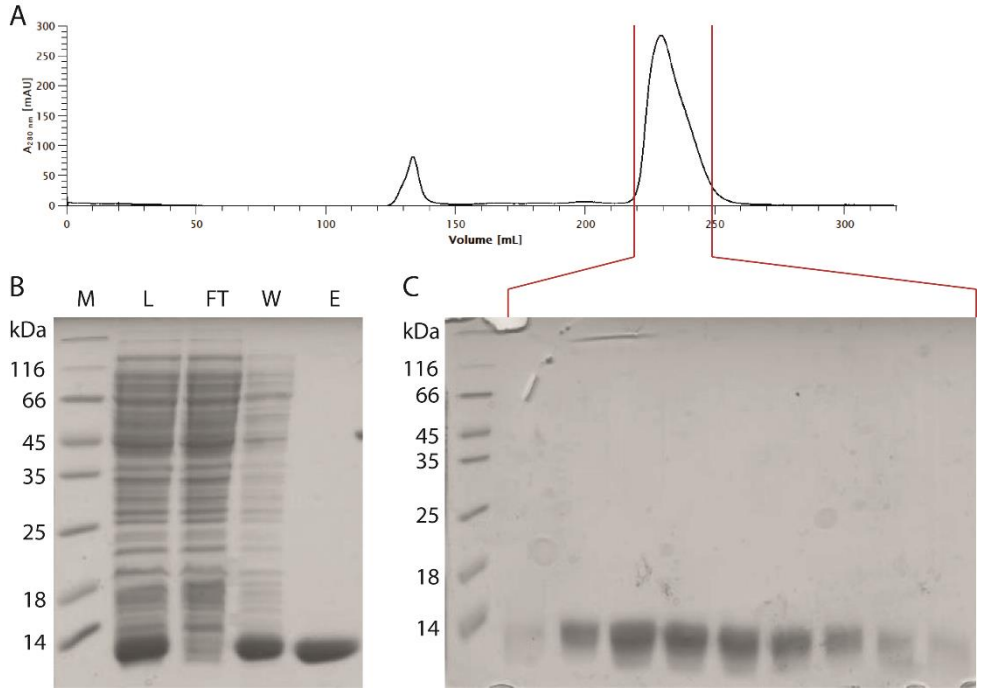


Figure 24: Purification of *Ct*Pth11

A) The chromatogram of the SEC purification of *Ct*Pth11. Only low absorption at 280 nm can be detected in this case, because CtPth11 does not contain any tryptophans. B) The 15% SDS-PAGE analysis after Ni-NTA purification of *Ct*Pth11. M: marker, L: lysate, FT: flow-through, W: wash (20 mM imidazole), E: elution (500 mM imidazole). C) SDS-PAGE analysis of the SEC purification of *Ct*Pth11. A red marking was used to indicate the fractions in the SEC chromatogram, which were analyzed via SDS-PAGE.

4. 4. 2. Crystallization of CtPth11

In initial crystallization experiments, crystal growth could be observed at a protein concentration of 5.4 mg/mL in 0.1 M citrate pH 5.6, 0.2 M K-Na tartrate, 2.0 M ammonium sulfate after several weeks of incubation at 18 °C (shown in Figure 25). These crystals diffracted to a resolution of approximately 2.4 Å. Around two months after setting up the crystallization screens, *Ct*Pth11 crystals were observed in following conditions as well:

- 0.5 M ammonium sulfate, 0.1 M tri-Na citrate pH 5.6, 1.0 M lithium sulfate
- 0.1 M citric acid pH 4.0, 1.6 M ammonium sulfate
- 2.0 M ammonium sulfate.

As these crystallization conditions contain medium to high concentrations of ammonium sulfate, the AmSO4 crystallization suite was used for further crystallization experiments. Crystal growth could be observed in several conditions of the screen, diffraction to a resolution of up to 1.8 Å was measured. *Ct*Pth11 crystallized in space group $P\,4_1\,2_1\,2$ with the unit cell constants $a = b = 68.68$, $c = 176.78$, $\alpha = \beta = \gamma = 90$ or $a = b = 71.59$, $c = 141.93$, $\alpha = \beta = \gamma = 90$, depending on the crystallization condition.

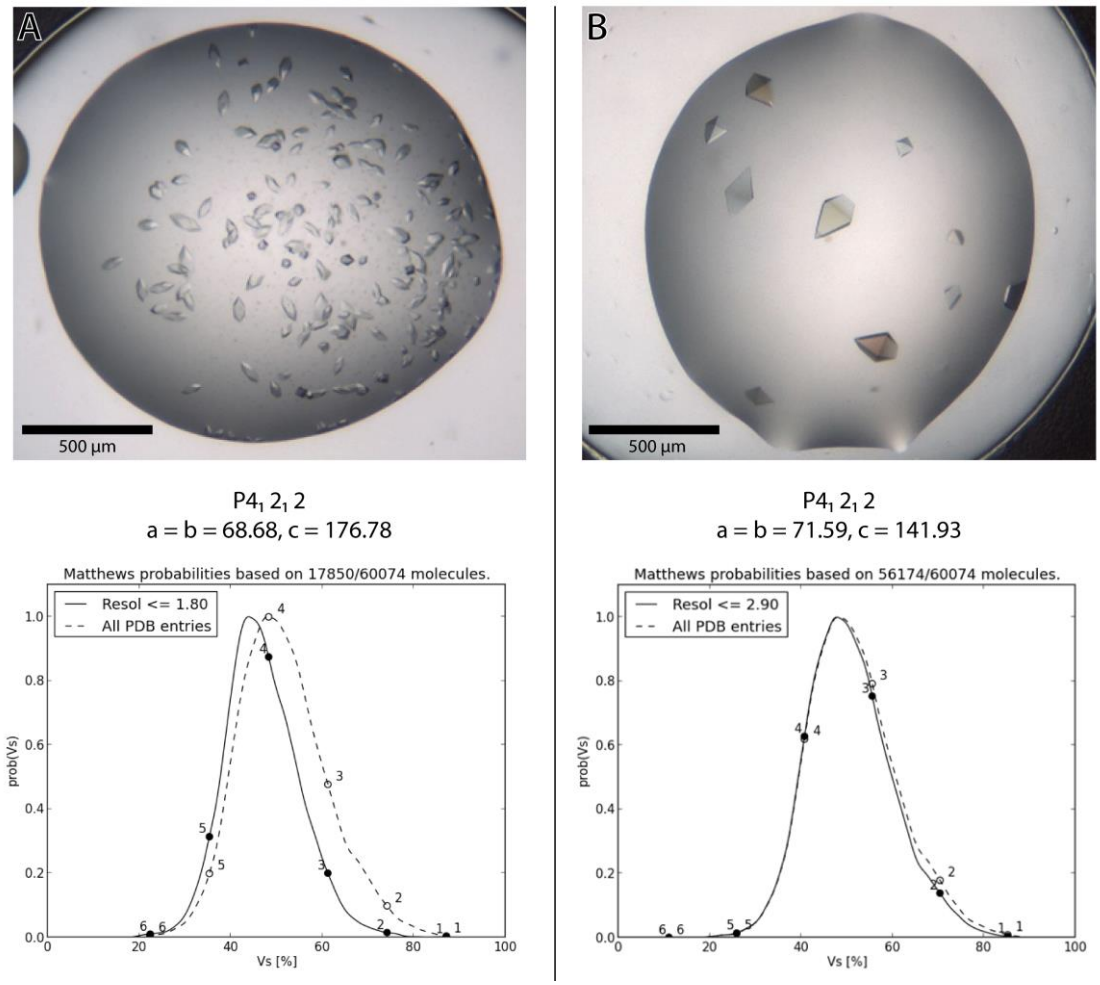


Figure 25: Crystals of the CtPth11 CFEM domain and corresponding unit cells

A) Crystals observed in the initial crystallization experiments are shown. These grew after several weeks of incubation at 18 °C in 0.1 M citrate pH 5.6, 0.2 M K-Na tartrate, 2.0 M ammonium sulfate; space group and unit cell constants are given below. The calculation of the Matthews probability (http://www.ruppweb.org/mattprob/default.html) determined a content of 4 molecules per asymmetric unit. B) Crystals of CtPth11 used for collection of S-SAD datasets – grown in 3.0 M ammonium sulfate, 1%(w/v) MPD – have different unit cell constants. The Matthews probability was calculated to determine the number of molecules per asymmetric unit, but does not show a clear result. The asymmetric unit accordingly contains either 3 or 4 molecules.

4. 4. 3. Structure solution

MR was attempted using the structure of Csa2 (PDB: 4Y7S) as a model. For this purpose, the full-length structure and several shortened variants of the structure were used. Additionally, a model of Pth11 was created using MODELLER[118] with Csa2 as a template to generate a model structure for MR. Despite all attempts, the structure could not be determined using this approach. Subsequently, heavy metal soaking and SAD phasing were pursued, but crystal quality was found to be massively affected by the soaking procedure.

As an alternative method for structure solution, native SAD-phasing using the anomalous diffraction from sulfur atoms was chosen. The method seemed feasible for *Ct*Pth11 because the CFEM-domain contains eight cysteines. As they are predicted to form four disulfide bridges they might be treated as "super-sulfurs" during the site-detection step if required. Furthermore, the high-symmetry space group the CFEM domain crystallized in is favorable for SAD phasing, as high multiplicity can be easily achieved. The data-collection strategy commonly applied for native SAD-phasing at beamline X06DA at the Swiss Light Source has been described by Basu *et al.* in 2019: The maximum wavelength achievable at the beamline (5.5 keV/2.25 Å) is used for data collection[93]. As this wavelength is still remote from the sulfur absorption edge (K-edge) of 2.472 keV/5.0155 Å[87], several 360° datasets are collected and merged. This approach generates high multiplicity, thus the low anomalous signal originating from the sulfur atoms is significantly enhanced. The data collection strategy is described to be suitable for crystals with anomalous signal extending to a wavelength of up to ~2.8 Å[93].

Table 10: Data collection and refinement statistics for the structure of the *Ct*Pth11 CFEM domain (outer shell values are given in the parenthesis)

	***Ct*Pth11**
Dataset name	2018_06_28-CC220A_x3
Data collection	
X-ray source	ESRF, ID
Wavelength (Å)	
Space group	$P\,4_1\,2_1\,2$
Unit cell parameters (Å)	$a = b = 68.68$, $c = 176.78$
Resolution range (Å)	54.23 - 1.822 (1.887 - 1.822)
Total no. of reflections	77555 (7568)
No. of unique reflections	38778 (3784)
R_{merge} (%)	0.02256 (0.239)
$I/\sigma(I)$	10.73 (2.56)
Completeness (%)	99.96 (99.92)
Multiplicity	2.0 (2.0)
$CC_{1/2}$	0.999 (0.97)
Refinement	
R_{work}/R_{free} (%)	19.1/23.76
No. of atoms	2726
Average *B* factor (Å^2)	49.72
R.m.s. deviations	
Bond length (Å^2)	0.012
Bond angles (°)	1.08
Ramachandran plot (%)	
Favoured	98.79
Allowed	1.21
Outliers	0.00
Rotamer outliers (%)	4.47

Four datasets were merged on site using a custom script for *xscale*[54], anomalous diffraction was observed to a resolution of around 3.5 Å. This does not meet the requirements for native SAD-phasing that were previously described[93]. Substructure determination was attempted on site using the *SHELXD* procedure[119], but the substructure could not be detected. Evaluation of the same merged datasets was done using *CRANK2*[120] and lead to structure solution. The native SAD structure was then used as a template to solve another dataset from a *Ct*Pth11 crystal that diffracted to 1.8 Å resolution (see Table 10 for data collection and refinement statistics).

4. 4. 4. The structure of the CtPth11 CFEM domain

The structure of the *Ct*Pth11 CFEM domain is shown in Figure 26. It consists of five α-helices and is stabilized by four disulfide bonds, which are formed between following residues: C43 and C83, C47 and C78, C57 and C64, C66 and C99. Chain B from the asymmetric unit was chosen here to examine the surface of the CFEM-domain of *Ct*Pth11. The surface examination reveals a large positively charged cleft on one side of the protein and a smaller negatively charged indentation on the other side. Analysis of surface electrostatics was done using the APBS Plugin for PyMOL. The positive charge is caused by three lysine residues (K80, K92, K104) on the cleft's entrance. Deeper inside, it is predominantly composed of hydrophobic and uncharged amino acids. In the crystal structure the cleft is occupied by two sulfates, which are part of the crystallization condition. The smaller indentation on the other side is negatively charged due to a glutamic acid (E49) on its entrance. On the inside, hydrophobic amino acids can be observed (I52, F48). F48 seems to divide the two cavities from each other. Interestingly, a different orientation of this residue can be observed in each molecule of the asymmetric unit or alternative side chain conformations are present. The *Ct*Pth11 CFEM domain thereby has either a hole or two cavernous surface invaginations.

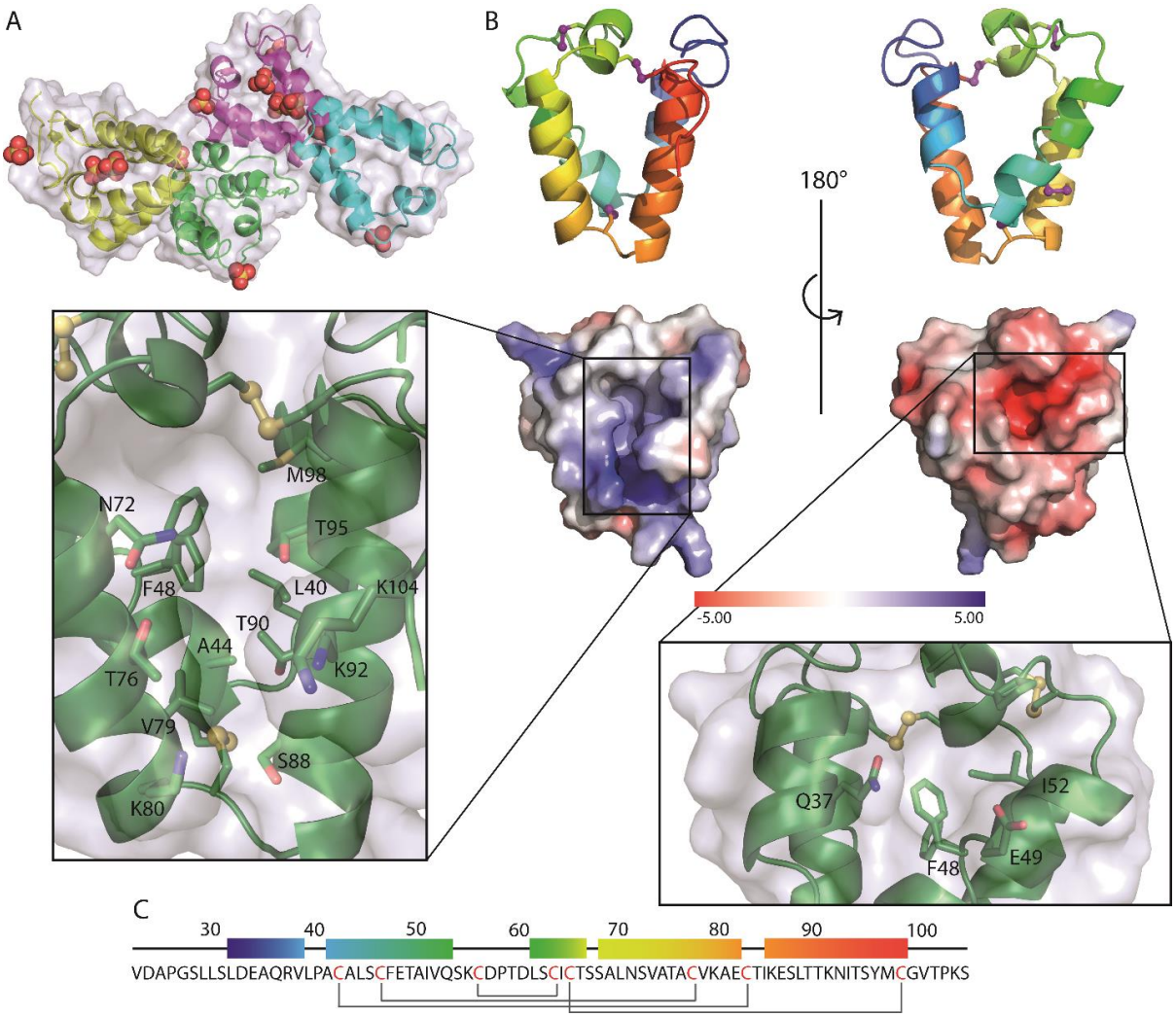


Figure 26: Structure of the *Ct*Pth11 CFEM domain

A) The asymmetric unit of crystals of the *Ct*Pth11 CFEM domain. It contains (in this case) four molecules of the CFEM domain and 13 sulfurs originating from the crystallization condition. B) The overall structure of a single molecule (chain B) is shown in two orientations. Two cavities can be identified when observing the protein surface, the APBS Tool in PyMOL was used to generate a surface electrostatic potential map. The larger positively charged cavity is mainly composed of hydrophobic and uncharged residues. F48 is shown in its two alternative side chain conformations, which indicates a certain flexibility of this residue. Three lysine residues at the entrance of the pocket provide a positive charge to the potential binding pocket. A smaller indentation can be found on the other side of the molecule. F48 also plays a role in this cavity, as well as the negatively charged E49. C) The sequence of the Pth11 CFEM-domain. The α-helices are indicated by boxes above the sequence; disulfide bonds are marked below the sequence.

4. 4. 5. Fragment screen

Although several studies aimed for the identification of the ligand of *M. oryzae* Pth11, it still remains unknown[40,44]. To gain hints on a putative ligand of Pth11, a fragment screen was conducted against its CFEM domain. Given the important role of Pth11 in appressorium formation and plant infection, Pth11 represents a promising target for agrotechnological applications.

Fragments were soaked into protein crystals in concentrations of either 100 mM or 50 mM, depending on the solubility of the fragment in the crystallization condition. The protein crystals were protected from ice crystal formation by addition of glycerol in the soaking conditions. Soaking times were extended as long as possible, up to 26 h. However, in many cases, the soaked crystals dissolved rather quickly and the soaking times had to be kept very short. Crystals were then frozen in liquid nitrogen and brought to the synchrotron for data collection. In many cases, a significant decrease of crystal quality could already be anticipated during the soaking procedure. Crystals cracked, slowly dissolved or showed other signs of disintegration. For those conditions, soaking times were kept very short (e. g. only 1 min). In total, 87 different fragments were used in the experiments. As multiple soaking durations were used for most fragments, in total 163 *Ct*Pth11 crystals were soaked and analyzed. 35 crystals did not show sufficient diffraction for data collection. The automatic data analysis software at the SLS (*DA+*) was able to automatically process 80 datasets; 48 datasets had to be processed manually using *XDS*, which failed for 10 of those. All datasets obtained from the fragment screen are listed in Appendix II.

A custom script was then used for data reduction (using *AIMLESS*) and running the *DIMPLE* software pipeline. 21 datasets could not be handled by the script, data reduction, structure solution (using *Phaser*) was done manually for those. Also, all structures were manually evaluated to detect any bound fragments that may have been overlooked by *DIMPLE*. Additional electron density was observed for four fragments (see Figure 27).

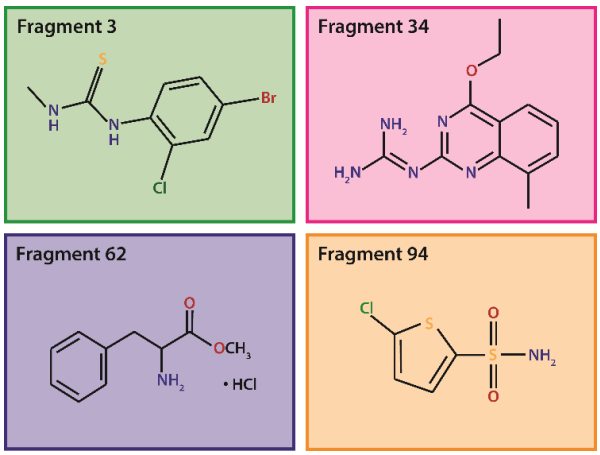


Figure 27: Fragments that were bound by the *Ct*Pth11 CFEM domain

The chemical structures of the fragments and their number in the Frag Xtal Screen (*Jena Biosciences*) are shown.

Table 11: Data collection and refinement statistics for fragment-bound *Ct*Pth11 (outer shell values written in parentheses)

	***Ct*Pth11-Frag3**	***Ct*Pth11-Frag34**	***Ct*Pth11-Frag62**
Dataset name	VR_139	VR_219	VR_171
Data collection			
X-ray source	SLS, X06SA (PXI)	SLS, X06SA (PXI)	SLS, X06SA (PXI)
Wavelength (Å)	1.0	1.0	1.0
Space group	$P\,4_1\,2_1\,2$	$P\,4_1\,2_1\,2$	$P\,4_1\,2_1\,2$
Unit cell parameters (Å)	$a = b = 68.76$, $c = 175.59$	$a = b = 68.87$, $c = 175.74$	$a = b = 69.22$, $c = 176.26$
Resolution range (Å)	46.86 - 2.0 (2.07 - 2.0)	48.7 – 2.12 (2.2 – 2.12)	48.95 – 2.1 (2.18 – 2.1)
Total no. of reflections	59061 (5754)	49658 (4796)	51722 (5017)
No. of unique reflections	29536 (2881)	24831 (2398)	25879 (2516)
R_{merge} (%)	0.01416 (0.5417)	0.01988 (0.8066)	0.01316 (0.5151)
$I/\sigma(I)$	15.64 (1.22)	11.85 (0.92)	21.08 (1.36)
Completeness (%)	99.71 (98.96)	99.39 (95.48)	99.71 (99.80)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
$CC_{1/2}$	1 (0.713)	1 (0.498)	1 (0.739)
Refinement			
R_{work}/R_{free} (%)	18.66/22.36	20.89/23.49	19.51/25.13
No. of atoms	2589	2507	2553
Average *B* factor (Å^2)	52.64	82.61	74.65
R.m.s. deviations			
Bond length (Å^2)	0.014	0.01	0.007
Bond angles (°)	1.33	1.24	0.82
Ramachandran plot (%)			
Favoured	98.47	99.07	98.17
Allowed	1.22	0.93	1.52
Outliers	0.33	0.0	0.3
Rotamer outliers (%)	2.82	3.62	3.25

4.4.5.1. CtPth11 CFEM domain – Fragment 3

A crystal of the *Ct*Pth11 CFEM domain was soaked in mother liquor containing 50 mM fragment 3 (SMILES code: CNC(=S)NC1=C(C=C(C=C1)Br)Cl) for 23 h, then directly flash-frozen in liquid nitrogen and brought to the synchrotron for data collection. The crystal diffracted to a resolution of 2.0 Å. Data analysis using the *DIMPLE* pipeline did not identify any unmodelled blobs, but upon manual examination of the data additional electron density was found. Fragment 3 is bound to three of the four CFEM domains in the asymmetric unit (see Figure 28). Placement and conformation of the bound fragment are the same in each of the three molecules with occupancies of 0.68 in chain A, 0.59 in chain B and 0.75 in chain D. The electron density for each ligand molecule in the structure is clearly defined. At the location of the Br-ion, negative difference electron density can be observed, caused by increased radiation damage at this specific location.

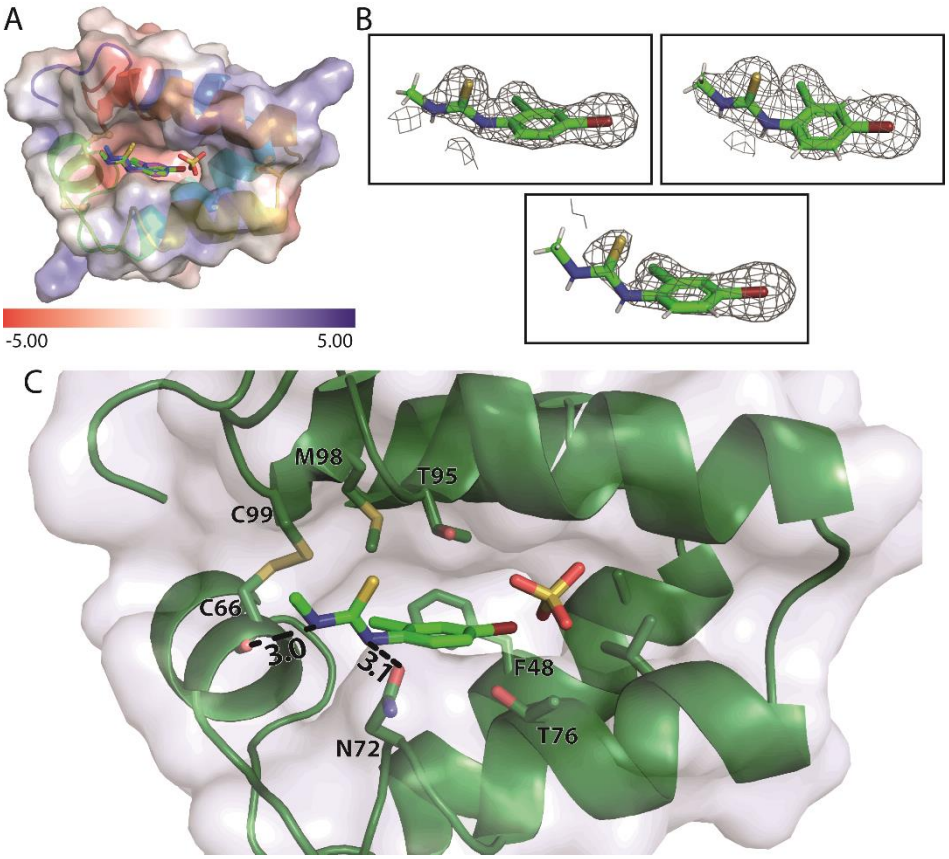


Figure 28: Interaction between fragment 3 and the *Ct*Pth11 CFEM domain

A) Surface representation of a single *Ct*Pth11 CFEM domain. The fragment is bound in the larger, negatively charged cleft of three from the four molecules in the asymmetric unit. The orientation of the bound fragment is the same in all three. B) $2mF_{obs}$-DF_{calc} maps (contoured at 2.0 σ) of the ligands. The electron density is nicely defined for all three bound fragments. C) Binding mode of fragment 3. The hydrophobic fragment is placed in the hydrophobic cleft. Only two weak electrostatic interactions are formed, involving C66 and N72.

Two weak electrostatic interactions are formed between the CFEM domain and the ligand: first, between the hydroxyl group of the side chain of N72 and the fragment with a distance of 3.1 Å; second, between the hydroxyl group O of the peptide bond of C66 and the ligand with 3.0 Å distance. Further specific interactions between the *Ct*Pth11-CFEM domain and fragment 3 cannot be observed. It is rather the case that the hydrophobic fragment is placed in the hydrophobic region of the larger cavity of the domain.

4.4.5.2. CtPth11 CFEM domain – Fragment 34

*Ct*Pth11 CFEM-domain crystals were soaked in mother liquor containing 50 mM fragment 34 (SMILES code: CCOc1nc(NC(N)=N)nc2c(C)cccc12) for 3 h and 24 h. The corresponding datasets have resolutions of 2.0 Å and 2.1 Å, respectively. The 24 h dataset was successfully processed by automatic data analysis software *DA+* and unmodelled blobs were identified by *DIMPLE*. Contrarily, the 3 h dataset had to be evaluated manually. Fragment 34 is bound to all four molecules in the asymmetric unit in both cases. Due to better data quality, the 24 h dataset was chosen for refinement and interpretation of the structure.

The placement of the ligand in the electron density is unambiguous. A part of the fragment is not visible in the electron density map. This is the same in all four molecules of the asymmetric unit and could be caused by a certain degree of flexibility of the fragment in this area. However, it is more likely that the fragment has broken apart because the O atom in the vicinity of the aromatic rings is not expected to be flexible and should therefore be visible in the electron density map. This could have happened during storage of the fragment, dissolving it in the crystallization condition or during soaking.

The side chains of three residues interact with the guanidine group of fragment 34: N72, T76 and T95. These form electrostatic interactions with following distances: 2.6 Å between the hydroxyl group O of N72 and the fragment, 3.2 Å between T76 and the fragment, and 2.9 Å between T95 and the ligand. Additionally, the hydrophobic aromatic rings of the fragment are placed in the hydrophobic cavity.

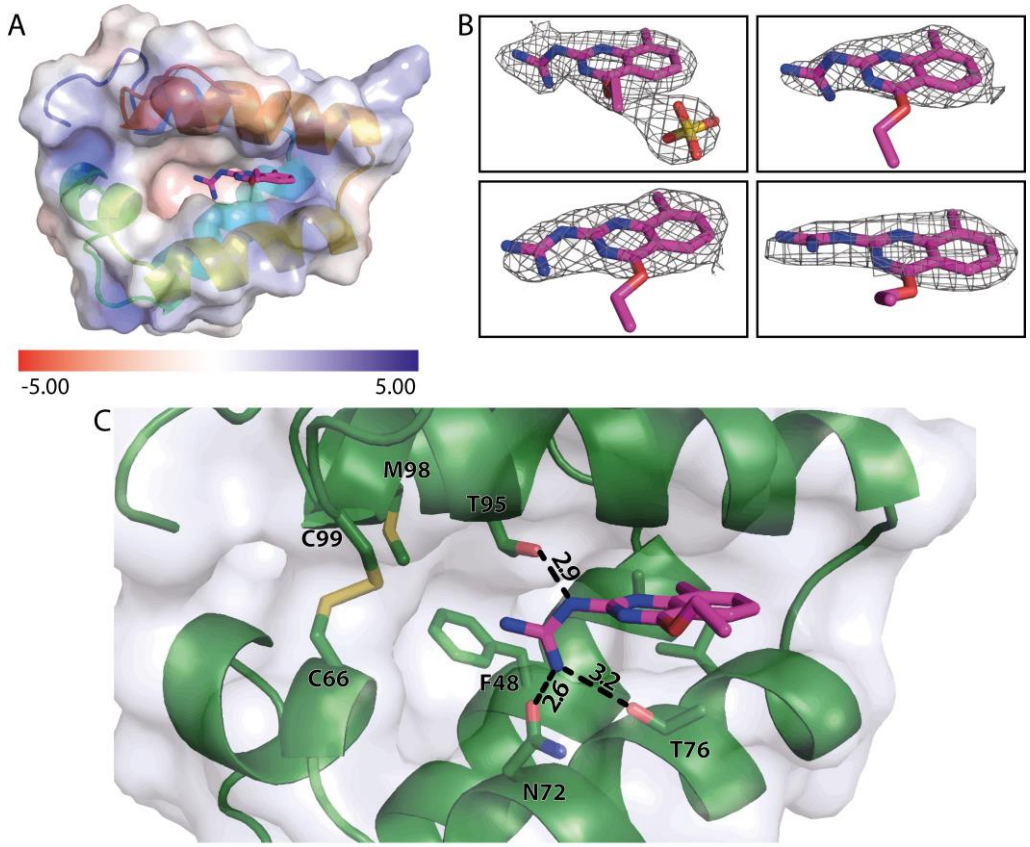


Figure 29: Interaction between the CtPth11 CFEM domain and fragment 34

A) Cartoon and surface representation of a single *Ct*Pth11 CFEM domain. Fragment 34 is located in the negatively charged cleft, with the same orientation in all four molecules in the asymmetric unit. B) $2mF_{obs}$-DF_{calc} maps (contoured at 2.0 σ) of the bound fragments reveal that the electron density is clearly interpretable. The same part of the ligand is not visible, indicating that the ligand has disintegrated. The electron density of the fragment bound to chain A merges into the density of a sulfate, which is located in its vicinity. C) Interactions between the CFEM domain and fragment 34 are shown. There are three residues involved: N72, T76 and T95.

4.4.5.3. CtPth11 CFEM domain – Fragment 62

100 mM fragment 62 (SMILES code: COC(=O)C(CC1=CC=CC=C1)N.Cl) were soaked into a *Ct*Pth11 CFEM domain crystal for 6 min, which diffracted to 2.1 Å. The acquired dataset was successfully processed by the automatic data analysis software *DA+* and by the *DIMPLE* pipeline. However, no unmodelled blobs were identified by *DIMPLE*. Only upon manual examination of the data, additional electron density was found in the cavities of all four CFEM domains in the asymmetric unit.

All four unmodelled regions in the electron density map are clearly interpretable and each one reveals a good fit of fragment 62. The occupancies are 0.95, 1.0, 0.76, and 0.84, for chain A, B, C, and D, respectively. Interestingly, the fragment is positioned slightly different in each *Ct*Pth11 CFEM domain in the asymmetric unit. The aromatic ring is analogously placed in all four cavities; the other portion of the fragment is positioned differently in each one. Specific

interactions between protein and fragment cannot be observed. Rather, the hydrophobic fragment is bound within the hydrophobic cavity.

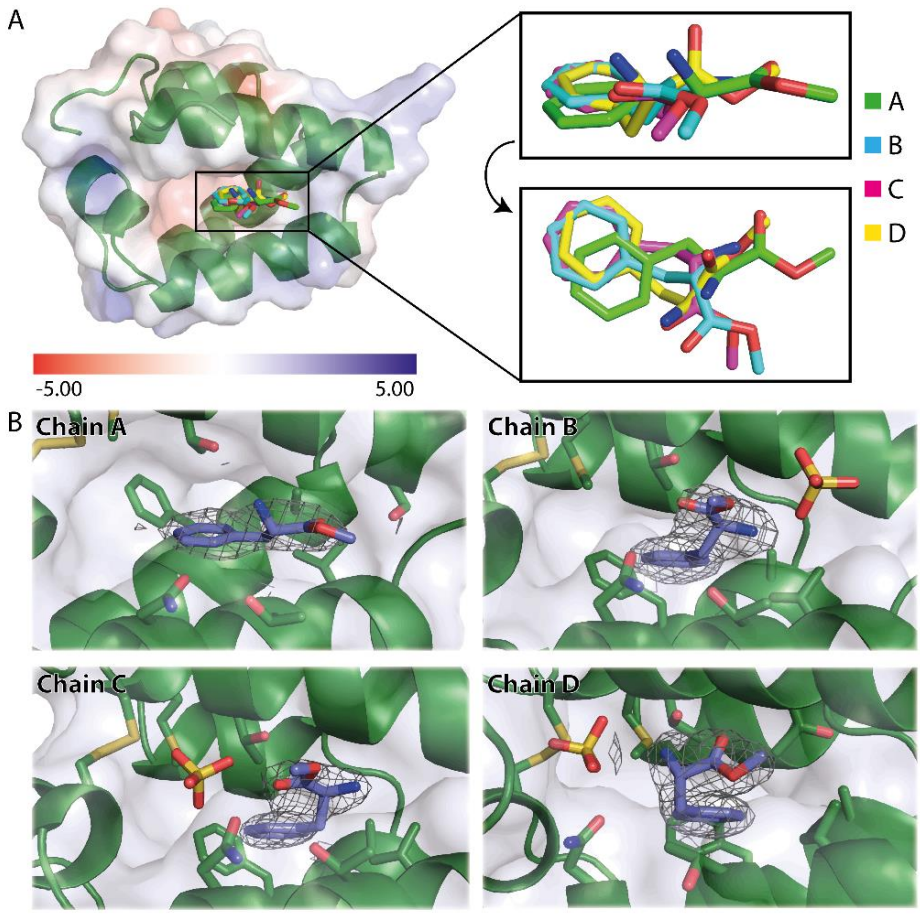


Figure 30: Interaction between the CtPth11 CFEM domain and fragment 62

A) A CFEM domain is shown in cartoon and surface representation, with fragment 62 depicted in all orientations. Slight differences can be observed between the fragments bound to each molecule in the asymmetric unit. The aromatic ring is in the same position in chains B, C, and D, but slightly displaced in chain A. B) The electron densities of each bound ligand. All electron densities are defined very well. Interactions with specific residues in the CFEM domain cannot be observed.

4.4.5.4. CtPth11 CFEM domain – Fragment 94

Soaking experiments with crystals of the *Ct*Pth11 CFEM domain and 100 mM fragment 94 (SMILES code: NS(=O)(=O)c1ccc(Cl)s1) were conducted for approximately 2 min. Crystals were quickly disintegrating during the soaking process. Nevertheless, a dataset with a resolution of 2.0 Å could be collected. The dataset was successfully handled by *DA+*, as well as by the *DIMPLE* pipeline, but unmodelled blobs were only identified upon manual examination of the dataset.

Additional electron density has been detected in all four molecules, to different extent. The most striking unmodelled region is located in chain C of the asymmetric unit and shown in Figure 31. However, the fragment does not fit into the densities. Data collection statistics for the corresponding dataset are shown in Table 12.

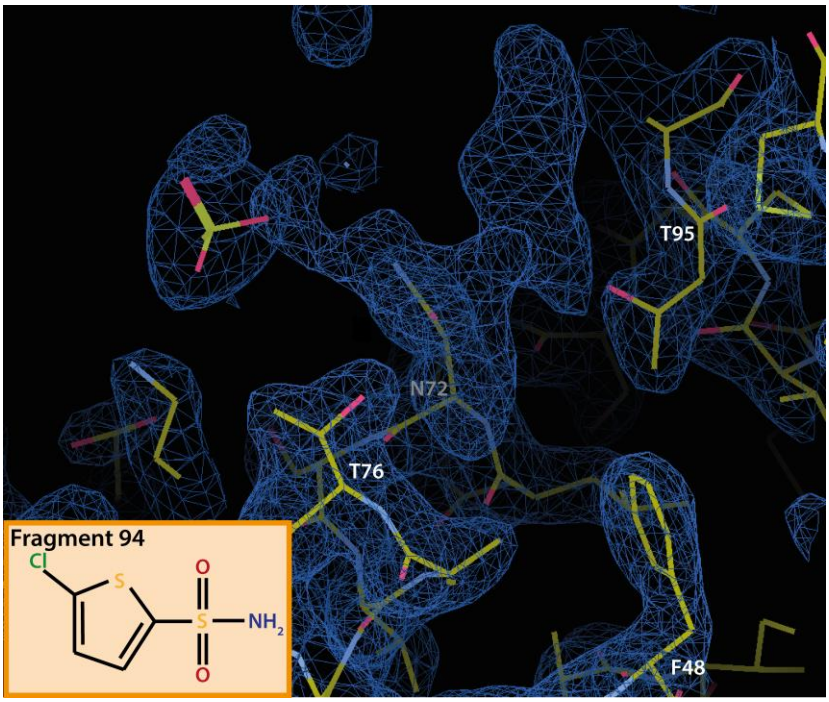


Figure 31: Additional electron density observed in VR_225

After soaking CtPth11 crystals with 100 mM fragment 94, unmodelled blobs could be observed. The additional electron density located in chain C of the asymmetric unit is shown here. The soaked fragment is depicted in the lower left corner of the image.

Table 12: Data collection statistics for the CtPth11 CFEM domain, soaked with fragment 94

Dataset name	VR_225
X-ray source	SLS, X06SA (PXI)
Wavelength (Å)	1.0
Space group	$P\,4_1\,2_1\,2$
Unit cell parameters (Å)	$a = b =$, $c =$
Resolution range (Å)	48.67 – 2.0 (2.06 – 2.0)
Total no. of reflections	420935 (29716)
No. of unique reflections	29439 (2103)
R_{merge} (%)	0.112 (1.599)
$I/\sigma(I)$	13.6 (2.0)
Completeness (%)	99.8 (98.3)
Multiplicity	14.3 (14.1)
$CC_{1/2}$	0.998 (0.722)

5. Discussion

5. 1. The cell wall of the thermophilic fungus *C. thermophilum*

Proteins from thermophilic organisms are generally considered more stable than their mesophilic counterparts[14]. This feature is not only favorable for the expression and biochemical characterization of a protein, but also for the crystallization process[121]. In recent years, the thermophilic fungus *C. thermophilum* has proven to be an excellent model organism for analysis of eukaryotic proteins[10,19,20] and it may also serve as a well suited model organism for the study of fungal cell wall proteins. However, information on the cell wall proteome of *C. thermophilum* is still lacking. This thesis aimed to fill this gap by bioinformatic prediction of GPI-anchored proteins and mass spectrometric analysis of GPI-CWPs. Additionally, the cell wall structure was analyzed via TEM.

5. 1. 1. Prediction of the C. thermophilum cell wall proteome

In total, 79 GPI-anchored proteins were predicted in *C. thermophilum* (see chapter 4. 1. 1, Table 4) using combination of signal peptide detection (using SignalP[62]), rejection of proteins with transmembrane helices (TMHMM[63]), and identification of GPI anchor signal sequences via the Big-PI Fungal Predictor[12] and a pattern search (as described by de Groot *et al.*[11]). The annotated *C. thermophilum* proteome, which was used as an input for the prediction, is derived from its genome and contains 7165 protein sequences. The number of predicted GPI-anchored proteins in *C. thermophilum* therefore represents 1.1% of its proteome. This fraction varies significantly in different fungi. For example, only 28 proteins in *Schizosaccharomyces pombe* (0.56 % of the proteome) contain both, an N-terminal signal peptide and a GPI anchor attachment sequence. In *S. cerevisiae* 59 GPI-anchored proteins (0.93 % of the proteome) were predicted; and 169 proteins (1.19 %) in *C. albicans*. Within the proteome of the filamentous fungus *Aspergillus nidulans*, 74 (0.78 %) proteins were predicted to be GPI-anchored[12]. With 79 predicted GPI-anchored proteins in *C. thermophilum*, the absolute number is very similar to *A. nidulans*, but in relative terms they represent a higher percentage of the genome.

The prediction of GPI-anchored proteins has proven to be very robust[12], but several limitations have to be considered. First, the Big-PI Fungal Predictor and the pattern search do not include the ω- region of the GPI anchor signal sequence, which has been shown to be associated with the final location of GPI-anchored proteins[11,13]. The amino acids located right upstream the GPI anchor attachment site (ω site) are considered determinants for the final localization of a GPI protein, i. e. the plasma membrane or the cell wall[9,13,122,123]. In yeast, the sequences of GPI-plasma membrane proteins (GPI-PMPs) are proposed to contain a dibasic motif just before the ω site (ω-4 to ω-1)[13]. However, the dibasic motif can be overridden by certain sequence features, e. g. long Ser/Thr-rich regions[124]. The sorting signal was shown to be

slightly different in *A. fumigatus*, where only one basic residue at the ω-1 or ω-2 site was identified in many GPI-PMPs[123]. The exact sequence requirements for discrimination between GPI-PMPs and GPI-CWPs are therefore elusive. Also, the localization of proteins to either the cell wall or the plasma membrane is not considered absolute[124]. Nevertheless, discrimination between GPI-PMPs and GPI-CWPs based on the features of the ω- region should provide a first insight on the distribution of the predicted GPI proteins. The Big-PI Fungal Predictor was used for the identification of the ω site before manual examination of the ω- regions in the protein sequences was done. If no potential GPI modification site was found by Big-PI, the residue indicated to be most likely the ω site was used. Proteins were assigned as GPI-PMPs, if a dibasic motif in the region from ω-4 to ω-1 was identified, or if a single basic residue at positions ω-2 or ω-1 was found. Unexpectedly, only few GPI-PMPs could be determined using this method; these are listed in Table 13.

Table 13: Predicted GPI-PMPs in *C. thermophilum*

UniProt-ID	Description (UniProt)	Family/Domains	Recognition as GPI-PMP
G0SEQ3	hypothetical protein CTHT_0064570	FAD-binding	Dibasic motif (ω-4 and ω-3)*
G0S249	1,3-beta-glucanosyltransferase-like protein	GH72/Gel2	Single basic residue (ω-2, alternative GPI-modification site)
G0SDH5	phosphoric diester hydrolase-like protein	PLC-like phosphoric diesterase	Single basic residue (ω-2)*
G0SDV4	hypothetical protein CTHT_0053120	Wsc-domain	Single basic residue (ω-1)
G0S5C3	hypothetical protein CTHT_0024300		Single basic residue (ω-1)
G0SCA5	hypothetical protein CTHT_0056530		Single basic residue (ω-1)
G0SHI8	hypothetical protein CTHT_0070170		Single basic residue (ω-1)
G0SFJ0	hypothetical protein CTHT_0071010		Dibasic motif (ω-2 and ω-1)*

* No GPI-modification site predicted by Big-PI, residue with the best score used

A dibasic motif is only contained in two *C. thermophilum* GPI proteins (G0SEQ3 and G0SFJ0). Identification of GPI-PMPs using only a single basic amino acid as an indicator for the final localization resulted in a list of 6 more proteins, including Gel2 (G0S249). However, certain proteins considered as typical GPI-PMPs, such as Gel1 or Ecm33, were not detected as such[122,123]. This may have two reasons: firstly, some of the GPI proteins that are located at the plasma membrane in other fungi are transferred to the cell wall in *C. thermophilum*. Secondly, the conditions for retention of GPI-proteins in the plasma membrane might be different in the thermophilic fungus. The transfer of GPI proteins from the plasma membrane into the cell wall is catalyzed by the transglycosidase Dfg5. It was shown that the removal of an ethanolamine-phosphate (EtN-P) group at the first mannose of the GPI-core glycan is required for successful cell wall transfer[10]. This group is proposed to be removed by Cdc1[125], a process which might be dependent on the amino acids in the ω- region of a protein. An analysis of Cdc1 may therefore provide the missing link in determining whether a GPI protein ends up in the plasma membrane or in the cell wall.

Another limitation of the prediction of GPI-anchored proteins is associated with the input itself. When ORFs are used to predict an organism's proteome, the analysis does not confer a realistic picture of the proteome and the relevance of a part of the hits may be debatable. To elucidate the importance of the hits, they were compared to the proteomic study conducted by Bock *et al.*[18]. The predicted GPI proteome contains 28 proteins, for which there is proteomic evidence, indicating biological relevance of these proteins. These are expressed in the organism upon growth in the standard media conditions described by the German Collection of Microorganisms and Cell Cultures (DSMZ)[18] and listed in Table 14.

Table 14: Predicted proteins with proteomic evidence in Bock *et al.*[18]

UniProt-ID	Description (UniProt)	Family/Domains
G0S879	hypothetical protein CTHT_0037870	Agglutinin-like
G0S3D9	alpha-amylase-like protein	Alpha-amylase-like
G0SAA8	hypothetical protein CTHT_0041610	Alpha-carbonic anhydrase, zinc-ion binding
G0S3S8	hypothetical protein CTHT_0030500	CFEM
G0SEF6	putative cell wall protein	Ecm33
G0SG17	hypothetical protein CTHT_0064700	GH catalytic core, ASL-like
G0SFX7	putative cell wall protein	GH16
G0S5R2	hydrolase-like protein	GH16, ConA-like domain
G0SCM1	putative cell wall protein	GH16, LamG superfamily
G0SA20	cell wall glucanase-like protein	GH16, LamG-superfamily
G0SFR4	hypothetical protein CTHT_0071830	GH17
G0S1A4	hypothetical protein CTHT_0012900	GH18, Chitinase, LysM-domain
G0S6S8	1,3-beta-glucanosyltransferase-like protein	GH72/Gel1
G0S249	1,3-beta-glucanosyltransferase-like protein	GH72/Gel2
G0SFW3	putative UPF0619 GPI-anchored membrane protein	Kre9/Knh1
G0SHT5	hypothetical protein CTHT_0073300	Kre9/Knh1
G0SF37	phospholipase-like protein	Lysophospholipase
G0S1H4	aspartic-type endopeptidase-like protein	Peptidase A1 family/aspartic-type endopeptidase
G0S3I8	hypothetical protein CTHT_0021410	Peptidase A1/pepsin-like
G0SAZ2	hypothetical protein CTHT_0048310	Tetratricopeptide repeat
G0S8N5	hypothetical protein CTHT_0038740	
G0S8Q3	hypothetical protein CTHT_0039950	
G0S9L3	hypothetical protein CTHT_0046300	
G0SDX5	hypothetical protein CTHT_0053340	
G0SDZ7	hypothetical protein CTHT_0053570	
G0SCA5	hypothetical protein CTHT_0056530	
G0SI03	hypothetical protein CTHT_0074010	
G0SCW3	hypothetical protein CTHT_0058590	

5. 1. 2. Mass-spectrometric analysis of C. thermophilum GPI-cell wall proteins

The limitations of the cell wall proteome prediction were addressed by mass spectrometric determination of the GPI-CWPs in *C. thermophilum*. The fungus was grown in liquid culture until spherical aggregates of mycelium had formed. The cell walls were then isolated as described by de Groot *et al.*, with only cell wall carbohydrates and GPI-CWPs supposed to be remaining in the sample. Cytosolic contaminants were removed by extensive washing with 1 M NaCl, and a boiling step with β-mercaptoethanol and SDS was conducted to remove PIR and disulfide linked proteins[66]. Regardless of the isolation steps, contamination by non-cell wall proteins cannot be completely prevented. Obvious contaminations were removed before the analysis.

34 GPI-CWPs were identified in *C. thermophilum* cell walls, with only few variations between the analyzed samples (see chapter 4. 1. 2, Table 6). Surprisingly, only 17 proteins were already included in the list of predicted GPI-anchored proteins. Among those, two are in the list of GPI-PMPs, namely G0S249 (Gel2) and G0SCA5 (uncharacterized). Accordingly, 17 proteins were found in the cell wall samples, but not predicted. These unpredicted proteins were all not recognized as GPI-anchored proteins by the Big-PI Fungal Predictor and via the pattern search. The identified proteins were sorted according to their putative function and are summarized in Table 15.

Table 15: Functional annotation of the *C. thermophilum* cell wall proteome

Category and UniProt-ID	Description (UniProt)	Family	Properties, proposed function
Carbohydrate-active enzymes			
G0SB94	Exo-1,4-beta-D-glucosaminidase	GH2	SP, 897 aa Involved in chitin degradation
G0RZA2	Glucoamylase	GH15	SP, 667 aa Hydrolyzes α-1,4-glycosidic bonds of starch
G0SDK5	Endo-1,3(4)-beta-glucanase-like protein	GH16	SP, 1104 aa Contains GH16-domain and Zn^{2+} dependent metallopeptidase (Peptidase M48) domain
G0SFX7	Putative cell wall protein	GH16	SP, GPI, 445 aa Involved in carbohydrate metabolism, acting on O-glycosyl components; **Crh**
G0SA20	Glycosidase	GH16	SP, GPI, 383 aa Involved in chitin metabolism, similar to **Crh1**
G0SCM1	Glycosidase	GH16	SP, GPI, 423 aa Involved in chitin metabolism, similar to **Crh1**
G0SFR4	Uncharacterized protein CTHT_0071830	GH17	SP, GPI, 394 aa Involved in carbohydrate metabolism, probable β-1,3-endoglucanase
G0SEU4	Hydrolase-like protein	GH17	SP, 552 aa Involved in carbohydrate metabolism

G0S1A4	Chitinase	GH18	SP, GPI, 908 aa Chitinase
G0RZV2	SH3b domain-containing protein	GH24	SP, 263 aa Lysozyme activity, Endolysin T4 type
G0SD45	Probable alpha/beta-glucosidase agdC	GH31	SP, 926 aa Involved in carbohydrate metabolism, α- and β-glucosidase activity
G0SH48	1,3-beta-glucanosyltransferase	GH72	SP, 514 aa Transglycosidase, also contains X8 domain
G0S6S8	1,3-beta-glucanosyltransferase	GH72	SP, GPI, 453 aa **Gel1**
G0S249	1,3-beta-glucanosyltransferase	GH72	SP, GPI, 482 aa **Gel2**
G0SFW3	Putative UPF0619 GPI-anchored membrane protein		SP, GPI, 218 aa **Kre9/Knh1**
G0S3D9	Alpha-amylase		SP, GPI, 533 aa Alpha-amylase
Other enzymatic activity			
G0S8P3	Serine-type endopeptidase-like protein		SP, 919 aa Subtilisin
G0S5M7	Catalase		SP, 723 aa Clade 2 catalase
G0S1H4	Aspartic-type endopeptidase-like protein		SP, GPI, 470 aa Pepsin
G0SBL0	Glyoxal oxidase-like protein	Wsc	SP, 1111 aa Contains 5 Wsc-domains and annotated glyoxal oxidase function
G0RZV3	Uncharacterized protein CTHT_0004320		SP, 237 aa Papain-like
G0SF37	Lysophospholipase		SP, GPI, 676 aa Lysophospholipase
G0SG36	SH3b domain-containing protein		SP, 253 aa Papain-like
Potential adhesins			
G0S002	CFEM domain-containing protein	CFEM	SP, GPI, 601 aa **Mad1**
G0S5W8	LysM domain-containing protein		327 aa Probably contains sequencing errors, Cyanovirin-N domain
Unknown proteins			
G0SDZ7	Uncharacterized protein CTHT_0053570		SP, GPI, 195 aa
G0S763	Uncharacterized protein CTHT_0027570		SP, 155 aa **Bys1**
G0S9L3	Uncharacterized protein CTHT_0046300		SP, GPI, 162 aa
G0S2U2	C3H1-type domain-containing protein		SP, 162 aa
G0SA61	Uncharacterized protein CTHT_0041120		SP, 507 aa
G0SCA5	Uncharacterized protein CTHT_0056530		SP, GPI, 200 aa
G0S3S8	CFEM domain-containing protein	CFEM	SP, GPI, 170 aa Contains CFEM domain, unknown function
G0SFS7	Uncharacterized protein CTHT_0071970		SP, 373 aa similar to *Neurospora crassa* Acw12
G0SEF6	Putative cell wall protein CTHT_0063570	Ecm33	SP, GPI, 400 aa **Ecm33**

SP: signal peptide detected by SignalP; GPI: GPI anchor attachment signal predicted

With exception of G0S002 (Mad1) and G0S2U2, all proteins identified in this study were also found in the proteomic analysis conducted by Bock *et al.*[18]. This is hardly surprising, because similar growth conditions were used.

More surprisingly, half of the detected proteins are not included in the list of predicted GPI-anchored proteins (see chapter 4. 1. 1, Table 4 and Table 5). There are two possible explanations for this outcome: First, the cell wall samples could be contaminated with material from other cellular components. Secondly, the GPI anchor signal sequence in *C. thermophilum* may not be recognized by the applied methods.

Obviously, the isolated cell walls are not completely free of contaminations with cytosolic proteins or plasma membrane proteins and the samples contain several proteins that are described to be GPI-PMPs, such as members of the GH72 family[126], as well as Gel1, Gel2, and Ecm33[123]. However, only very few transmembrane proteins were identified and the detection of GPI-PMPs in cell wall samples does not seem to be uncommon. An example for this is Ecm33: plasma membrane localization was described to be important for its function[127], but Ecm33 is still commonly identified in isolated cell walls[24,28,66,128]. The purity of the samples analyzed in this work is therefore considered appropriate.

The large amount of unpredicted proteins may be caused by the Big-PI Fungal Predictor not being perfectly suited for the prediction of GPI proteins in thermophilic fungi. The learning set of the algorithm consists of 254 entries, originating from following organisms: *S. cerevisiae*, *C. albicans*, *Neurospora crassa*, and *Schizosaccharomyces pombe*. The algorithm was then tested on sequences from *A. nidulans*, *C. albicans*, *N. crassa*, *S. cerevisiae*, and *S. pombe*, as well as several mutants of Gas1 and found to be reliable for these. But while filamentous fungi have been implemented in both the learning set and algorithm testing, this does not apply to thermophilic fungi. Even the pattern search failed to detect the 17 unpredicted proteins found in the isolated cell walls. This method is considered a much simpler tool for identifying GPI anchor signal sequences, which normally reveals a larger amount of potentially GPI-anchored proteins, but is also more unspecific. Nevertheless, the method proved to be compatible with the results of the Big-PI Fungal Predictor[12].

To obtain a clearer picture of the identified proteins, the 17 unpredicted cell wall proteins were analyzed via SignalP and the ω- region was examined for plasma membrane retention signals. A dibasic motif between ω-4 and ω-1 or single basic residues at positions ω-2 or ω-1 were regarded as such. The results are listed in Table 16.

Table 16: Unpredicted GPI proteins in the isolated C. thermophilum cell walls

Category and UniProt-ID	Description (UniProt)	Recognition as GPI-PMP
Carbohydrate-active enzymes		
G0SB94	Exo-1,4-beta-D-glucosaminidase	Single basic residue (ω-1)
G0RZA2	Glucoamylase	CWP
G0SDK5	Endo-1,3(4)-beta-glucanase-like protein	Dibasic motif (ω-4 and ω-3)
G0SEU4	Hydrolase-like protein	Single basic residue (ω-2)
G0RZV2	SH3b domain-containing protein	Dibasic motif (ω-4 and ω-3)
G0SD45	Probable alpha/beta-glucosidase agdC	Three basic residues (ω-4 to ω-2)
G0SH48	1,3-beta-glucanosyltransferase	CWP
Other enzymatic activity		
G0S8P3	Serine-type endopeptidase-like protein	Dibasic motif (ω-2 and ω-1), ω site is R! (GPI signal sequence maybe false)
G0S5M7	Catalase	Single basic residue (ω-2)
G0SBL0	Glyoxal oxidase-like protein	CWP
G0RZV3	Uncharacterized protein CTHT_0004320	Single basic residue (ω-1)
G0SG36	SH3b domain-containing protein	Single basic residue (ω-1)
Potential adhesins		
G0S5W8	LysM domain-containing protein	CWP*
Unknown proteins		
G0S763	Uncharacterized protein CTHT_0027570	CWP
G0S2U2	C3H1-type domain-containing protein	CWP
G0SA61	Uncharacterized protein CTHT_0041120	CWP
G0SFS7	Uncharacterized protein CTHT_0071970	Dibasic motif (ω-2 and ω-1)

* no signal peptide detected by SignalP

Approximately one third of the proteins identified in *C. thermophilum* cell walls (11 out of 34) contain a plasma membrane retention signal. Nine of these were not recognized in the prediction of GPI-anchored proteins. Accordingly, 23 proteins that were detected in the isolated cell walls could be assigned as GPI-CWP based on their sequence properties in the ω- region. Seven of these were not predicted and in one (G0S5W8), no signal peptide could be detected. That being said, the final localization of a particular GPI protein is not only dependent on the ω- region of the protein sequence. The plasma membrane retention signal was shown to be overridden by certain sequences, such as Ser/Thr-rich regions[124], similar to the ones often observed in adhesins[22,124]. Also the presence of additional unknown sequence properties influencing GPI protein localization cannot be excluded. In addition, the final localization of a particular GPI-anchored protein is not considered as being exclusive, i. e. it is regarded as a predominant localization[9,124].

Several issues concerning GPI-anchoring are highlighted in the analysis of the *C. thermophilum* cell wall proteome: Firstly, many proteomically identified cell wall proteins were not recognized as such by the identification of the GPI anchor signal sequence via the Big-PI Fungal Predictor and the pattern search. This indicates that the GPI anchor signal sequence may be

slightly different in *C. thermophilum* and possibly also in other thermophilic fungi. Secondly, the conditions for GPI sorting in fungi need further investigation, as it clearly is not solely dependent on the ω- region. The presence of Ser/Thr-rich regions has already been described to override the sorting signal in the ω- region[124], but additional properties may also be involved. In this context, it should also be considered GPI-anchored proteins are not strictly localized at either the plasma membrane or the cell wall, but rather predominantly[124,129].

Concerning GPI-sorting, Cdc1 is an attractive target for further research, as it is involved in GPI-anchor processing and thereby promotes the transfer of GPI-anchored proteins to the cell wall[10,125]. But also the interaction between the GPI-anchor and the plasma membrane itself has to be considered. Contrary to the widely held notion that GPI-anchors simply protrude from the plasma membrane (also referred to as the "lollipop" model), the glycan part of the GPI-anchor has been shown to interact with the membrane, so that the anchor is lying on the membrane ("flop down" model). This is thought to be caused by an interaction between amine groups (from EtN-Ps on the GPI-anchor) and the negatively charged phosphate groups of the membrane[130]; but also the presence of positively charged residues in the vicinity of the GPI-anchor, such as those commonly found in the ω- regions of GPI-PMPs, may contribute. This interaction between GPI-anchor and plasma membrane might be weakened by higher temperature, explaining the increased occurrence of proteins regarded as GPI-PMPs in the cell wall isolates of *C. thermophilum*.

5. 1. 3. The structure of the C. thermophilum cell wall

The ultrastructure of the cell walls of different fungi varies dramatically depending on their cell wall composition. In this regard, TEM presents a well-suited method to gain first insights into the cell wall properties of a fungus[1]. TEM is also commonly used to investigate the morphological effects of certain treatments or mutations on the cell wall (see for example Pardo *et al.*[131] and Popolo et al.[132]).

A few examples of different cell walls are described by Gow *et al.*[1] and are shown in Figure 32, including an image of the *C. thermophilum* cell wall that was obtained in this work. TEM reveals long fibrils of mannoproteins in the outer wall of *C. albicans*; in contrast the *A. fumigatus* cell wall does not contain any fibrils[1]. However, the *C. albicans* fibrils were shown to differ significantly in length, depending on strains and methodologies[133]. *Cryptococcus neoformans* is an example for a fungus, which is surrounded by a capsule, which can be imaged nicely using TEM[1,6]. In the *C. thermophilum* cell wall, the two layers of the cell wall – i. e. the inner and the outer wall – can clearly be distinguished. Also, short mannoprotein fibrils can be identified. A cell wall width of ca 75 nm was measured in *C. thermophilum*; this is in accordance with the cell wall thickness of *A. fumigatus*[134].

The cell wall width and morphology depends on several factors, such as the strain, growth conditions and sample preparation. Nevertheless, this work provides a first insight on the

C. thermophilum cell wall. It shows that the fungus does contain mannoprotein fibrils in the outer layer of the cell wall, which is in contrast to the fibril-free cell wall of *A. fumigatus*. Noticeably, *C. thermophilum* is not surrounded by a capsule.

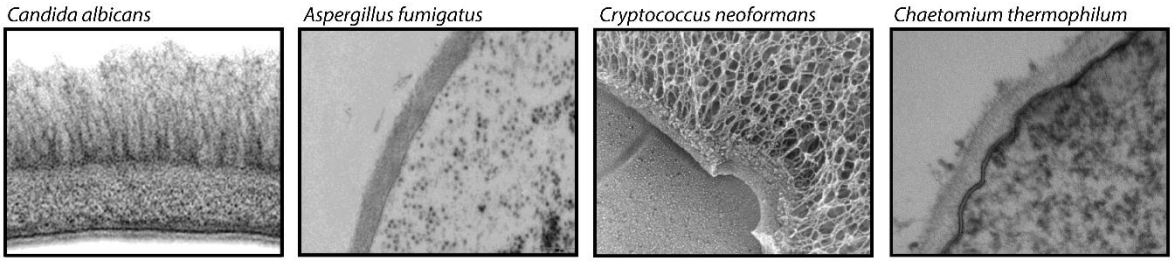


Figure 32: Ultrastructure of different fungal cell walls, adapted from Gow *et al.*[1]

TEM images of the cell walls of *C. albicans*, *A. fumigatus*, *C. neoformans*, and *C. thermophilum*. In all cases, inner and outer cell wall can be distinguished, but the ultrastructure of the walls differs significantly. The outer wall of *C. albicans* contains long mannoprotein fibrils, whereas *A. fumigatus* seems to lack these. *C. neoformans* in enveloped by a capsule, comprised of glucoronoxylomannan and galactoxylomannan. Mannoprotein fibrils can also be observed in the outer cell wall of *C. thermophilum*, but these appear very short compared to the ones in the *C. albicans* cell wall.

5. 1. 4. Targets for structural and biochemical studies on cell wall proteins

Since the fungal cell wall differs significantly from the cell walls of plants or bacteria and the cell membranes of mammalian cells, it has long been described as a promising target for antifungal drugs[4]. Especially cell wall biosynthesis has been shown to be an adequate target in this respect, as demonstrated by the effectiveness of the echinocandin class of antifungal drugs, which act on the β-1,3-glucan synthase Fks1[135,136]. However, the presence of resistances against echinocandins has already been described[137]. The demand for novel antifungal drugs is therefore a matter of concern. Fungal cell wall proteins are also used for the development of vaccines against fungal infections. For example, the recombinant A-domains of the *C. glabrata* adhesins Als1 and Als3 were shown to be effective in animal models[138,139]. But the development of antifungal drugs should not be the only focus of the analysis of cell wall proteins. Also the process of cell wall biosynthesis and the function of some essential cell wall proteins are not fully elucidated yet[1].

Proteins with a high Sequest HT score in all *C. thermophilum* cell wall samples are the glycoside hydrolases G0SDK5 (GH16), G0SEU4 (GH17), and G0RZV2 (GH24). Interestingly, homologs or orthologues of these could not be identified, thus their roles remain undetermined. In addition, many proteins known to be required for biosynthesis, remodeling, and integrity of the fungal cell wall were detected in the samples, including two homologs of Crh1 (G0SA20 and G0SCM1), orthologues of Gel1 (G0S6S8) and Gel2 (G0S249), as well as Kre9/Knh1 (G0SFW3) and Ecm33 (G0SEF6). These are obviously promising targets for further research.

The **Crh** family of transglycosidases is responsible for chitin-chitin and chitin-glucan crosslinking. The number of its members varies in different fungi, with three members in *S. cerevisiae* and *C. albicans* and five members in *A. fumigatus* and *N. crassa*[140], and seven members in *Aspergillus niger*[141]. Three putative Crh family members were identified in the *C. thermophilum* cell wall isolates analyzed in this work, namely G0SFX7, G0SA20, and G0SCM1. Another member of this protein family may be G0S0M3, which was not found in the cell wall samples. Crystal structures of *A. fumigatus* Crh5 are already available (PDB: 6IBU, 6IBW). The Crh family members function redundantly and are not essential for cell wall integrity[140], thus they are not regarded promising targets for the development of antifungal drugs.

Gel1 and **Gel2** are β-1,3-glucanosyltransferases that are orthologous to members of the yeast Gas1 family[142]. The protein family plays a major role in cell wall biogenesis during vegetative growth; it has five members in *S. cerevisiae*[2]. The Gel protein family in *A. fumigatus* consists of seven members[142]. Two obvious members of the Gel family could be identified in *C. thermophilum* cell walls, the Gel1 orthologue G0S6S8 and G0S249, which is similar to Gel2. But also G0SH48, which was detected in the cell wall samples, is similar to Gel1 and may belong to the Gel family.

Kre9 and **Knh1** are functional homologues involved in β-1,6-glucan metabolism, with Kre9 taking the dominant role. Deletion of Kre9 leads to slower cell growth and reduction and defects in the β-1,6-glucan moiety of the cell wall. The phenotype of the Kre9 mutant can be rescued by overexpression of Knh1[2]. Recently, *Candida tropicalis* Kre9 has been shown to possess β-1,6-glucanase activity and has been identified as the target of the antifungal peptide CGA-N12[143]. Kre9 is therefore known to be an excellent target for antifungal drugs and a first biochemical analysis has been conducted; the structure of Kre9 remains unknown. The *C. thermophilum* cell wall isolates contain two proteins similar to Kre9/Knh1: G0SFW3 and G0SHT5. G0SBY7 is also similar to Kre9, but has not been identified in both, the prediction of GPI-anchored proteins (as no GPI-anchor attachment sequence could be identified) and the cell wall isolates.

Ecm33 (Extracellular Mutant 33) and its paralog Pst1 (Protoplasts-Secreted) have been characterized in several fungi (*S. cerevisiae*[131], *Candida albicans*[144], and *A. fumigatus*[145,146], among others[147]), but their function remains elusive. Deletion of Ecm33 results in cell wall defects, including a thin or even absent mannoprotein layer and defects in N-glycosylation, particularly affecting the elongation of N-linked outer chains. Ecm33 contains a receptor L-domain, which is characteristic for certain mammalian receptors, such as insulin receptor[131]. Ecm33 is one of the most common cell wall proteins and is considered to be of major importance for cell wall integrity and biosynthesis. It is regarded a promising target for further characterization. In this respect, especially structural and biochemical analysis of Ecm33 are required for understanding its function[2].

The *C. thermophilum* cell wall analysis revealed two potential adhesins: G0S5W8 and G0S002. G0S002 is an orthologue of the CFEM domain containing adhesin Mad1, which has been

shown to be involved in the adhesion to insect cells in *Metarhizium anisopliae*[148,149]. Some identified cell wall proteins appear a bit unusual on the first sight, such as G0S763, a protein similar to Bys1. Such proteins could also be identified in the cell walls of some *Aspergillus* species (*A. fumigatus*, *Aspergillus flavus*, *A. nidulans*)[65]. The function of Bys1 is unknown, it is expressed at high temperatures in the pathogenic fungus *Blastomyces dermatitidis*[150]. The *C. thermophilum* cell wall also contains an α-amylase (G0S3D9) and a glucoamylase (G0RZA2). These are commonly found in thermophilic fungi and hydrolyze α-1,4-glycosidic linkages[151].

Several proteins have been described as relevant targets for biochemical characterization within the fungal cell wall by Orlean (2012), including Ccw12, Ecm33, Kre1, and Kre9[2]. Some orthologues of these were identified in *C. thermophilum* cell wall isolates. These may be of use for further biochemical studies and especially for structural studies on named proteins.

5. 2. Analysis of cluster VI adhesins from *C. glabrata*

The structures of Awp1A and Awp3A stand out from known structures of *C. glabrata* adhesins. This opportunistic pathogen harbors various families of adhesins, of which the Epa (epithelial adhesin) family resembles the largest and best characterized one[22,31]. High-resolution structures are available of the A-domains of three members, Epa1, Epa6, and Epa9, in complex with different ligands[26,29,152]. The A-domains of Epa family members contain an anthrax protective antigen (PA14) domain, which mediates glycan binding. Another family of *C. glabrata* adhesins also contains an N-terminal PA14 domain and is therefore called the Pwp (PA14 containing wall proteins) family. However, no structural information from Pwp family members is accessible at present[31]. Other subgroups on *C. glabrata* adhesins are poorly characterized, identification usually relies on the typical domain architecture of adhesins[22].

5. 2. 1. Structural similarity to pectate lyase

Structural similarity of Awp1A and Awp3A to proteins deposited in the PDB was analyzed via a pairwise 3D alignment with PDBeFold v2.59 with the default cut-off of 70 % for lowest acceptable similarity (see Appendix VII)[153]. Various proteins were identified to be similar to Awp3A, including the heme-hemopexin binding HxuA from *Haemophilus influenza*[154], a variety of polysaccharide lyases from different organisms (e. g. the pectate lyase Bsp165PelA from *Bacillus Sp. N165*[155], pectate lyase A from *Erwinia chrysanthemi*[156], alginate lyase from *Paenibacillus Sp. Str. FPU-7*[157]), as well as other polysaccharide binding proteins (e. g. the chitin-binding polysaccharide lyase-like protein Cthe_2159 from *Chaetomium thermocellum*[158], the Vi-antigen lyase VexL from *Achromobacter denitrificans*[159] or the serine-rich repeat protein ($SRRP_{100-23}$) from *Lactobacillus reuteri*[160]). The identified proteins all contain a three-faced right-handed β-helix. In general, sequence conservation was observed to be low, with sequence identities between Awp3A and search results ranging from

4.3 to 14.8 %; and RMSD values ranging from 2.63 to 6.04 Å, which indicates structural similarity. Similar results have also been observed for other β-helix proteins[154,158].

Because the identified structurally similar proteins are all carbohydrate-binding proteins, a similar function was assumed for Awp1 and Awp3. Thus, binding to a wide variety of carbohydrates was analyzed via TSA and Glycan array screening (see chapter 4. 2. 5). The experiments did not detect binding to any of the carbohydrates tested.

5. 2. 2. Potential Ca^{2+} binding properties of Awp1A and Awp3A

Because the structures of Awp1A and Awp3A both contain a parallel β-helix, they pose the question of Ca^{2+} binding. Parallel β-helices were identified in polysaccharide lyase families PL1, PL3, PL6, and PL9[161]. In those enzymes, as well as in the polysaccharide lyase-like Cthe_2159 that was encountered in the PDBeFold search, Ca^{2+} is required for ligand recognition[158,161]. Also in the Epa family of *C. glabrata* adhesins, ligand binding is dependent on the presence of Ca^{2+} at the binding site[31]. The use of lanthanides as probes for Ca^{2+} binding sites has been described on several occasions[162]. Accordingly, potential Ca^{2+} coordination sites in Awp3A should be revealed by binding of the Ca^{2+} mimicking Gd^{3+} ions in the structure of Awp3A-Gd and conservation in Awp1A can be analyzed. A high number of the Gd^{3+} ions in Awp3A-Gd is involved in cluster formation, where they do not directly interact with the protein, or they interact with the protein via a single residue only (glutamic acid or aspartic acid). Obviously, these interactions do not resemble a Ca^{2+} binding site. Several Gd^{3+} ions are coordinated by two residues, amongst those two ions are located in a tetrahedral Gd^{3+} cluster, interacting with Q102, E132 and E134 (see cluster 2 in Figure 35). Interestingly, these Gd^{3+} coordination sites are not conserved in Awp1A, in which the Ser/Thr ladder is located at the corresponding face of the β-helix. Also Q70 and Q106 coordinate a single Gd^{3+} ion, as well as N181 and D183. Also these sites are not conserved in Awp1A. A higher coordination number can be observed for two Gd^{3+} ions in Awp3A-Gd, which are located in the T1 loop region of the β-helix. They interact with the carbonyl groups of K109, R110 and G139, and with E141 and D169 (see Figure 33 A). However, a certain flexibility of these loop regions is implied, as the same regions are different in the native structures of Awp3A. The T1 loops Awp1A are dissimilar from the ones of Awp3A-Gd and Awp3A as well. A structural alignment with the pectate lyase C from *Dickeya chrysanthemi* (PDB: 2EWE) as a representative for the search results from the PDBeFold search indicates that no putative Ca^{2+} binding sites in Awp3A are located at positions equivalent to the the active site Ca^{2+} binding site of pectate lyases and pectate lyase-related enzymes (Figure 33 B). Consequently, Awp3 cannot be considered a Ca^{2+} dependent adhesin.

In Awp1A, no heavy atom binding was observed, although the crystallization condition contained a variety of lanthanides, namely Er, Tb, and Yb. Thus, there is no indication of Ca^{2+} binding in Awp1A too.

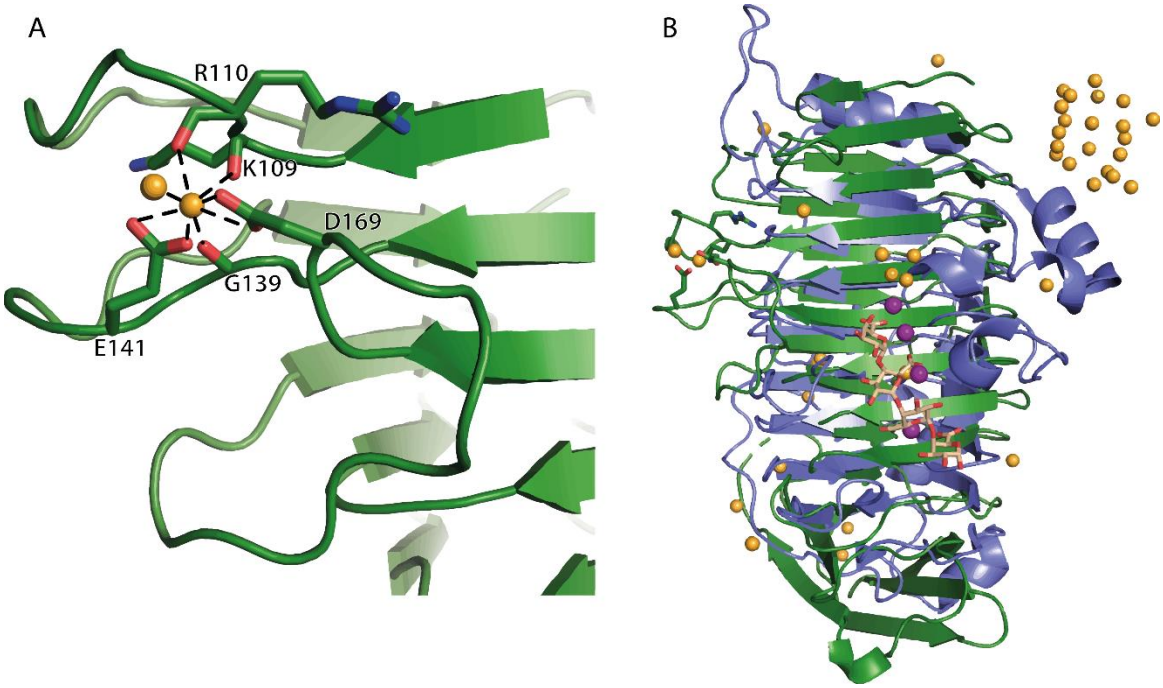


Figure 33: Coordination of Gd^{3+} in a potential Ca^{2+} binding site in Awp3A and comparison to pectate lyase

A) A potential Ca^{2+} binding site in Awp3A, revealed by the Ca^{2+} mimicking Gd^{3+} that was introduced into the protein via soaking during the structure solution process. Among the numerous Gd^{3+} ions identified in the structure, only few are coordinated by more than one residue, therefore resembling a Ca^{2+} site. One of those is shown here, interacting with the side chains E141 and D169 and the carbonyl groups of K109, R110 and G139. B) A superposition of Awp3A-Gd (shown in green) and pectate lyase C from *Dickeya chrysanthemi* (PDB: 2EWE, depicted in blue) reveals that the potential Ca^{2+} binding sites of Awp3A are not located near the expected ligand-binding site. Hence there is no indication for Ca^{2+} dependency of Awp3.

5. 2. 3. Potential glycosylation sites in Awp1 and Awp3

Many fungal CWPs are functionally dependent on glycosylation, which can be divided into N-linked glycosylation and O-linked glycosylation. Upon N-glycosylation sugars are transferred onto asparagine residues in the protein, a process taking place on the cytosolic side of the ER[2]. The consensus sequence N-X-S/T (X can be any amino acid) can be used to recognize potential N-glycosylation sites[2,163]. O-linked glycosylation occurs on serine or threonine residues. However, there is no specific sequence motif associated with O-linked glycosylation in fungi[164]. Rather, the glycosylation seems to depend on a number of factors, including the sequence context (which is significantly different for glycosylated serines and threonines), secondary structure, and surface accessibility. Prediction tools for O-linked glycosylation are available for mammalian proteins (NetOGlyc)[165] and for *Dictyostelium discoideum* (DictyOGlyc)[166].

Because the structure of Awp1A reveals remarkable ladders of serine and threonine residues on the surface of the β-helix domain, a prediction of O-glycosylation sites in Awp1 was done using NetOGlyc 4.0[165]. The tool has been shown to overestimate O-glycosylation sites in fungi; nonetheless it is considered reliable, especially for the identification of highly O-glycosylated regions[167]. The tool predicted numerous glycosylation sides (see Appendix VIII), with the first

one being S235. Additional potentially O-glycosylated residues in Awp1 that are structurally resolved are: S254, S258, S262, T265, T267, T271, T273, T274, S292, T297, S299, S318, T321. Interestingly, none of the predicted glycosylation sites are located within the β-helix part of the protein; they are all located in the α-crystallin domain. Two potential O-glycosylation sites are not surface exposed and therefore not expected to be accessible (S254, T274); S318 and T321 are part of the C-terminal loop region in the structure. NetOGlyc also identified a vast amount of glycosylation sites in the Ser/Thr-rich region of Awp1. This coincides with glycosylation predictions performed on Ser/Thr-rich regions in other fungal cell wall proteins[168]. The last glycosylation site predicted is T845, which may already be part of the ω- region of the GPI-anchor signal sequence.

5. 2. 4. Reclassification of cluster V and cluster VI adhesins via a SSN

Classifications of Awp1-14 have been done by de Groot *et al.* in 2008[24] and by Xu *et al.* in 2020[27]; both classifications are based on a phylogenetic tree. In the phylogenetic analysis of protein sequences the gene tree is combined with the species tree. Resulting subtrees should contain proteins with similar functions, but this is not always the case[169]. In this respect, the SSN provides an additional tool for the classification of protein sequences, which is based on sequence similarity only[170]. Compared to a phylogenetic analysis, sequence similarity based methods perform better in the identification of isofunctional subgroups[169].

The SSN presented in this thesis was generated using the β-helical regions of the Awp1 and Awp3b A-domains for iterative PSI-BLAST searches. Thereby the included number of sequences could be expanded, which also lead to the inclusion of a large amount of bacterial sequences in the network. An E-value cut-off of 10^{-20} was used for SSN generation, hence the formed clusters only contain sequences below this E-value. The clusters in the network contain either bacterial or fungal sequences, no mixed clusters can be observed. In fact, most clusters in the network contain proteins from the same organism, except the Iff/Hyr cluster and a cluster of an unknown protein family, containing sequences from *Dothideomycetes*, *Taphrinomycetes*, *Basidiomycetes* and two plant sequences (from cork oak). Protein families could not be assigned to all clusters in the network.

Various adhesin families contained the network, including the Hyr1 and the Iff family of adhesins from *C. albicans*, which are members of the same cluster. Another cluster is formed by Hpf1, Css1 and Awa1 from *S. cerevisiae*. Interestingly, Hpf1 and Awa1 have been described to be similar to Awp1 and Awp2 by de Groot *et al.*[24]; a relationship that could be confirmed in the SSN. The fact that Awp1 and Awp2 are members of different clusters of adhesins (cluster VI and cluster V, respectively), but are both similar to Hpf1, was not entirely conclusive at that time, but is now confirmed in the SSN. The cluster VI adhesins Awp1 and Awp3 fall into different clusters, both containing sequences from *C. glabrata* exclusively. In contrast to that, the cluster V adhesins Awp2 and Awp4 are members of the same cluster in the SSN, which

also consists of *C. glabrata* sequences. Numerous paralogs of Awp2 were identified (Awp2a-i); the Awp2 paralog originally identified by de Groot *et al.* is named Awp2.

The SSN indicates a high similarity of the cluster V adhesins Awp1/3 to the cluster VI adhesins Awp2/4. This similarity is also expected to be conserved on the structural level, indicating that Awp2 and Awp4 also contain a β-helix motif. Sequence identity and similarity were determined via pair-wise alignment (using EMBOSS Needle) and are shown in Figure 34 A. The sequence identities between the proteins are range from 16.7% to 25.1%, which is generally high, especially for β-helix proteins[114]. Many hydrophobic residues are conserved and a pattern indicating the presence of β-strands can be observed, i. e. in many parts of the sequences every second amino acid is a hydrophobic one. Models of the Awp2 and Awp4 A-domains (ranging from Q26 – Y344 in Awp2 and from Q27 – S231 in Awp4) were generated using *SWISS-MODEL*[171] with Awp1A as a template. This resulted in generation of two different models for Awp2A and one model for Awp4A, which are shown in Figure 34.

A

	Awp1	Awp2	Awp3	Awp4
Awp1		24.6	25.1	18.3
Awp2	38.3		23.2	41.6
Awp3	40.6	36.1		16.7
Awp4	27.6	38.3	26.8	

Sequence identity [%]
Sequence similarity [%]

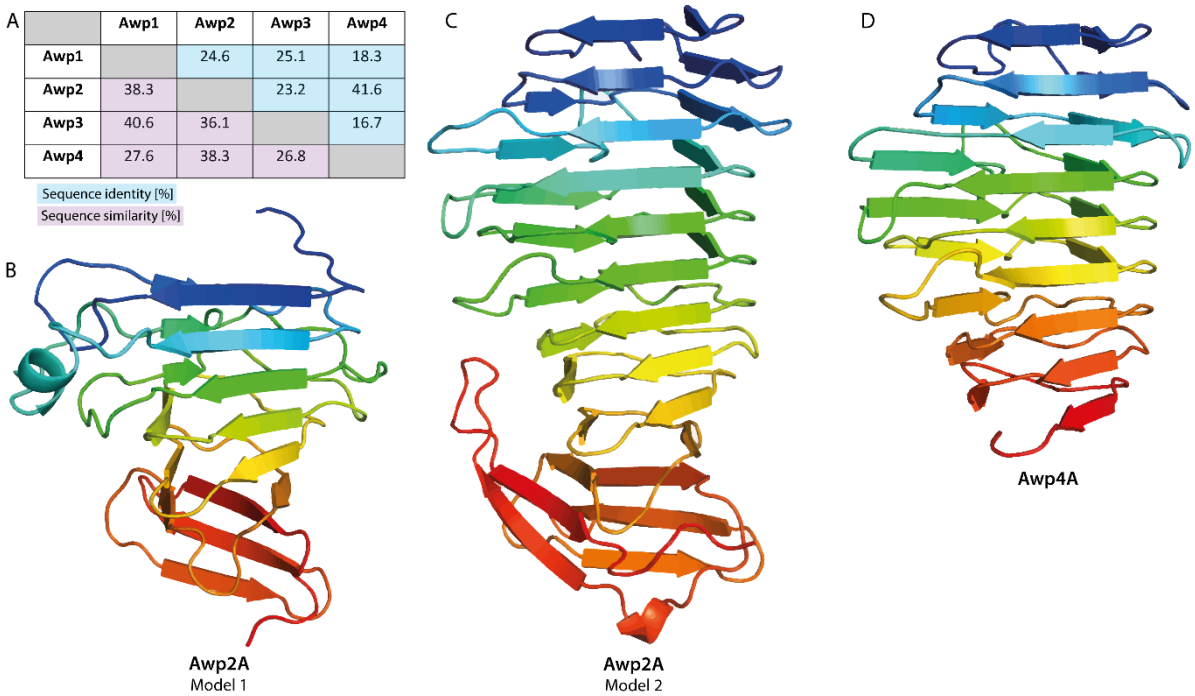


Figure 34: Models of the Awp2 and the Awp4 A-domains

A) Sequence identities and sequence similarities between Awp1-4 are given. B) A model of the Awp2 A-domain, containing P98 – N290. It reveals a β-helix motif with elongated loops at one side, forming a short α-helix. C) The second Awp2A model comprises a larger part of the Awp2 A-domain, namely I27 – N321. The model is highly similar to Awp1A. D) The model of Awp4A contains A28 – S231, which form a three-sided parallel β-helix.

In case of Awp2A, model 1 does not include the full sequence of the A-domain, consisting of P98 – N290. In contrast, model 2 contains almost the full sequence of Awp2A (namely I27 – N321), only 23 residues on the C-terminal end of the domain are missing. The latter is more similar to Awp1A. However, a loop similar to the elongated loop region in model 1 that includes a short α-helix, might also be a part of Awp2A. The "true" structure of Awp2A is expected to be a mixture between the two models, containing a parallel β-helix with extended loops on one side, which eventually form additional secondary structure elements; similar to the structures of some pectate lyases[114,115]. Awp4A was modelled from A28 – S231, thus including the whole A-domain with only one residue missing in the beginning of the sequence. As expected, the model reveals a three-sided parallel β-helix, which is expected to reflect the true structure of the protein very well.

5. 2. 5. Awp3A crystals soaked with Gd^{3+} acetate reveal a lanthanide cluster of three-fold symmetry

Soaking of Awp3A crystals in $Gd(OAc)_3$ resulted in incorporation of 42 Gd^{3+} ions in the asymmetric unit. At present, this is the highest number of lanthanide ions detected in a protein structure. Two Gd^{3+} clusters – which have formed by serendipity – can be identified in Awp3A-Gd, composed of 21 ions and four ions, respectively. The smaller cluster of four Gd^{3+} ions has the shape of a tetrahedron, participating ions are coordinated by Q102, E132 and E134. Distances between the Gd^{3+} ions range from 2.8 – 3.8 Å, they are 2.4 Å apart from the carboxyl group O of the coordinating residues.

The larger cluster is connected to the protein via two residues, D40 and E59. It is composed of four tetrahedral subclusters (A, B, C, D). Subclusters A, B and C reveal distances of 3.3 – 4.1 Å between the ions. They are connected by triangular planar clusters composed of three Gd^{3+}, with which they form a basket-like shape with three-fold symmetry. Distances of ions participating in composing those triangles range from 3.5 to 4.4 Å. Subcluster D is associated to the basket-like shape via a single Gd^{3+} ion, atoms in this subcluster are a bit further apparat from each other when compared to the other subclusters, namely 3.7 – 4.4 Å.

The formation of lanthanide clusters – also in protein crystals – has been described on several occasions[98,172–175]. Ma *et al.* described a tetrahedral Gd_4O_4 cluster, in which they measured distances of approximately 3.7 – 3.9 Å between Gd atoms[172]. Gd-Gd distances observed in Awp3A-Gd are similar to those, but the tetrahedral clusters are more distorted. In case of cluster 2 this may be caused by the coordination via three residues that push the ions into their positions. A distance of ca 2.4 Å between Gd and the carboxyl group O of a valine ligand was described in the Gd_4O_4 cluster[172]; this coincides with the distances measured between Gd^{3+} ions of cluster 2 and coordinating residues E134, E132, and Q102, as well as with the distances measured between D40 and E59 to ions from cluster 1. Clustering of heavy atoms could also be observed in other protein structures. For example, a heptanuclear Gd^{3+} cluster was detected on the surface of the A-domain of the yeast flocculin Flo5, Flo5A. In this case,

the cluster could be divided into two subclusters; one exhibiting the tetrahedral shape described above, the other one having the triangular shape that was also observed in Awp3A-Gd[98].

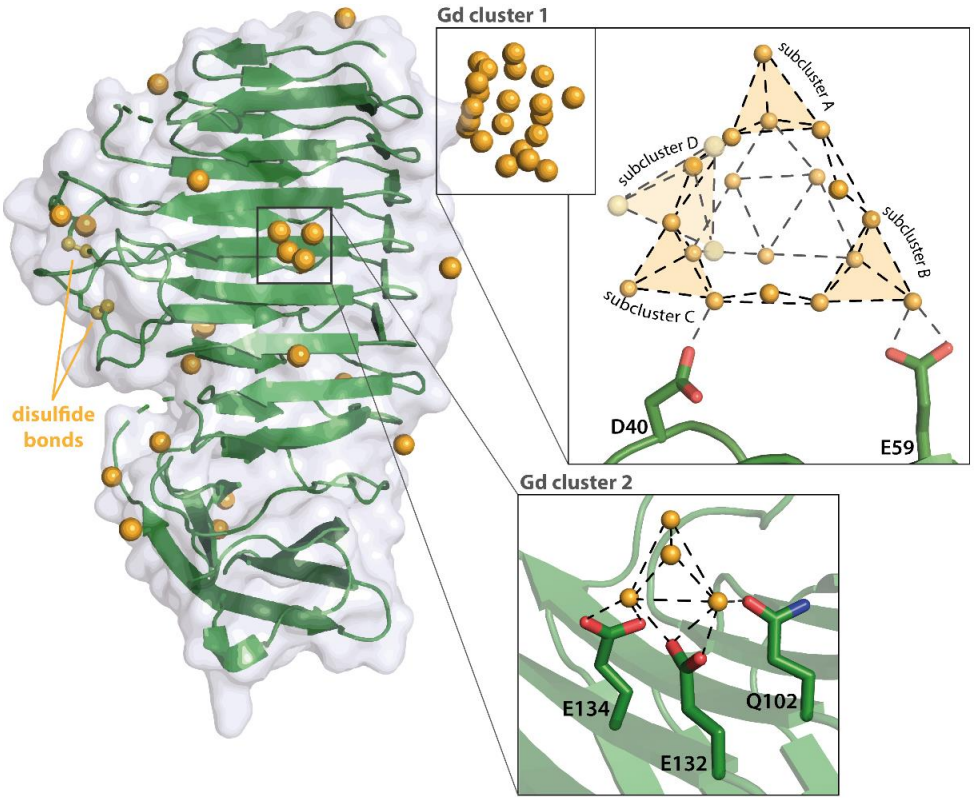


Figure 35: Overall structure of the Awp3 A-domain and coordination of Gd^{3+} clusters

The asymmetric unit of the Awp3A-Gd complex contains one molecule of the Awp3 A-domain (shown in cartoon representation in green), as well as 42 Gd^{3+} ions (orange spheres). Several single Gd^{3+} ions are associated to the protein's surface, as well as two Gd^{3+} clusters, one containing 21 Gd^{3+}, the other one consisting of 4 ions. Cluster 1 is connected to the protein via residues D40 and E59. The tetrahedral subclusters A, B, and C form a basket-like shape of three-fold symmetry, subcluster D is connected to the compartment on a corner of the basket. Cluster 2 is a tetrahedral cluster of 4 Gd^{3+} ions, coordinated by Q102, E132, and E134.

The paramagnetic properties of certain transition metals and lanthanides are commonly exploited for use as contrast agents in magnetic resonance imaging (MRI). Especially Gd complexes are widely used, in approximately 25 - 30 % of all MRI scans (as of 2005)[176]. Although the compounds are designed to be completely excreted from the human body, the accumulation of Gd in different tissues has been described. In patients with compromised renal function, which increases the plasma elimination half-life, Gd is deposited in the skin and various internal organs after administration of certain Gd-based contrast agents. But also patients with normal renal function get accumulations of Gd in the brain and in the bones. Cumulative and long-term effects of these are still unknown[176,177]. Gd clusters, such as the ones observed in the structures of Awp3A-Gd or Flo5A, may provide a basis for the design of novel protein-based contrast agents for MRI[98].

5. 3. Analysis of the CFEM domain of the GPCR *Ct*Pth11

Pth11 is a GPCR that is essential for appressorium formation in several fungal plant pathogens, including the rice blast fungus *M. oryzae*[40,44] and the causative of Fusarium Head Blight, *F. graminearum*[47]. The receptor has a CFEM domain on its N-terminus, which is proposed to contain the binding site for an unknown ligand, seven transmembrane helices, and an unknown cytoplasmic domain. Pth11 is regarded a relevant target for the development of antifungal agents for agriculture[40].

5. 3. 1. Structure of the CtPth11 CFEM domain

The structure of the *Ct*Pth11 CFEM domain was solved via S-SAD, after initial attempts using MR. For latter, the structure of the CFEM protein Csa2 from *C. albicans* (PDB: 4Y7S)[37] was used as an MR model. The Csa2 structure is the only structure of a CFEM domain currently contained in the PDB. The sequence identity and similarity of the CtPth11 CFEM domain (A36 – G100) and the one from Csa2 (Y56 – A119) are 18.5% and 33.8%, respectively. Considering the short length of the sequence and the presence of eight cysteines, which are a characteristic of the CFEM domain, these are very low numbers. In fact, only four more residues were found to be identical. It is therefore not particularly unexpected that the MR attempts failed, even though trimmed versions of the Csa2 structure and models of the Pth11 CFEM domain were used.

Structure solution was achieved via S-SAD, which uses the anomalous scattering originating from sulfurs naturally occurring in the protein for structure solution. The high amount of cysteines in the CFEM domain is advantageous in this regard, as is the high-symmetry space group that allows collection of data with high multiplicity. However, the protein was crystallized in a condition containing a high concentration of ammonium sulfate, thus it is hard to predict how many heavy atoms sites to expect and the presence of unordered sulfur atoms might be unfavorable during the phasing process. Four datasets collected from a single crystal were used for solving the structure of the *Ct*Pth11 CFEM-domain, using *CRANK2*[120].

The *Ct*Pth11 CFEM domain consists of five α-helices, connected to each other via four disulfide bonds (C43 – C83, C47 – C78, C57 – C64, C66 – C99). These are in accordance with the ones of the *C. albicans* Csa2 CFEM domain (PDB: 4Y7S; see Figure 36). *Ca*Csa2 belongs to the Pga7 family of CFEM proteins and is described to be involved in heme-iron acquisition from hemoglobin[37]. When comparing the structures of both CFEM domains – the one of *Ct*Pth11 and the one of *Ca*Csa2 – four helices align very well. This is reflected by the RMSD of 1.976 Å over 509 atoms of the superimposed structures. Only the most N-terminal helix of the *Ct*Pth11 CFEM domain is tilted when compared to the equivalent helix in the *Ca*Csa2 structure. The structure of *Ca*Csa2 does not only contain the CFEM domain, but also two additional α-helices, of which one is placed N-terminal of the domain and the other one C-terminal. The N-terminal helix is involved in ligand binding by being placed over the bound heme molecule like a lid.

Accordingly, the ligand binding site of *Ca*Csa2 is proposed to be on top of the CFEM domain, where the heme interacts with D80 in the CFEM domain and Y36 from the lid helix[37]. No other features indicating further ligand binding sites can be observed in the structure.

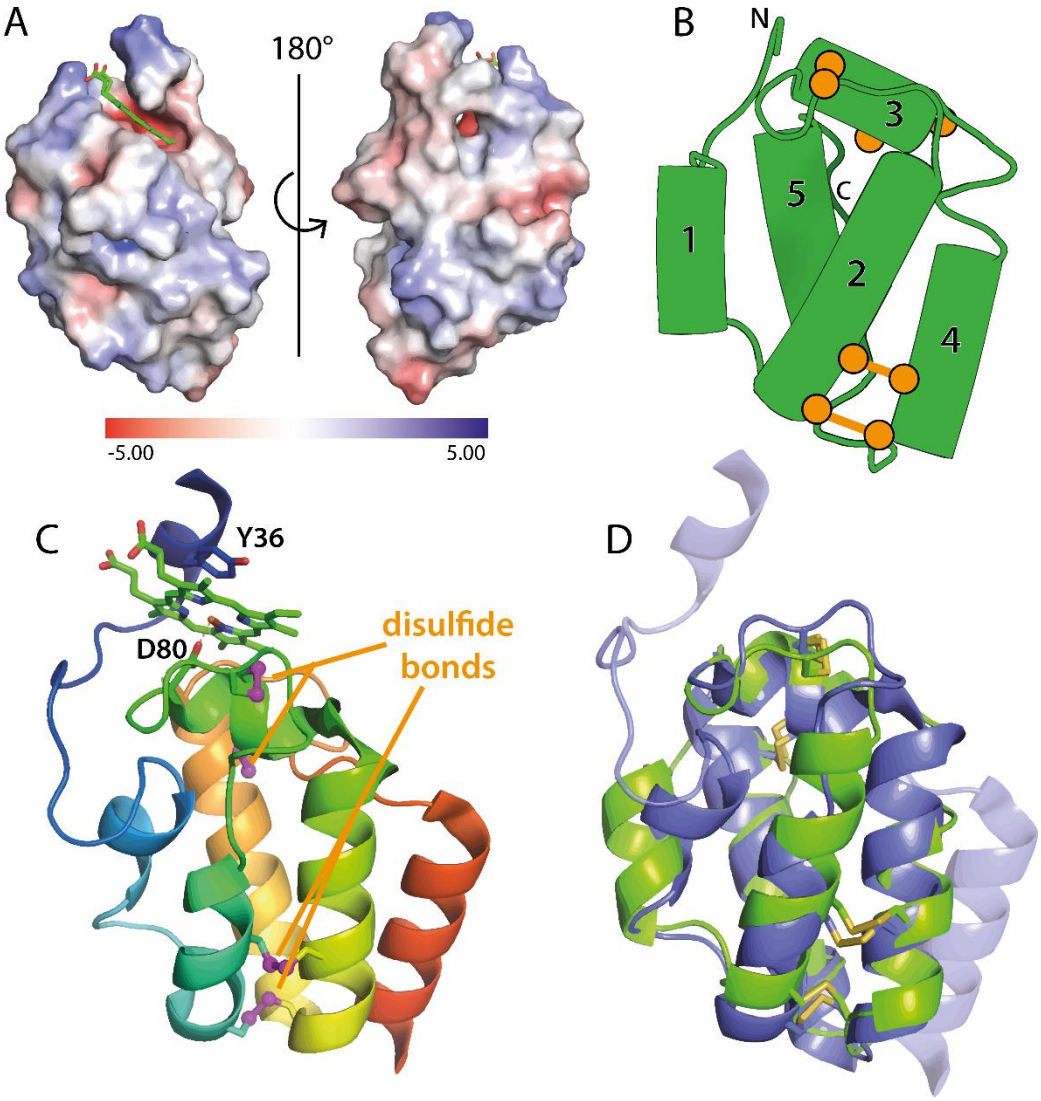


Figure 36: Comparison between the *Ca*Csa2 and the *Ct*Pth11 CFEM domains

A) Surface electrostatic potential of the CaCsa2 CFEM domain. The heme molecule is placed on top of the domain and buried by a lid formed by an α-helix N-terminally from the CFEM domain. B) Schematic representation of the CFEM domain (generated with *Protein Imager*[178]). The helices forming the domain are numbered and the disulfide bonds are indicated by the connected orange spheres. The arrangement of the disulfide bonds is conserved between the two proteins. C) Cartoon representation of the structure of *Ca*Csa2 in complex with heme, colored in rainbow scheme (N-terminus blue, C-terminus red). The disulfide bonds are shown in magenta and labeled. D) Superposition of the structures of *Ca*Csa2 (colored blue) and the *Ct*Pth11 CFEM domain (shown in green). The structure of the CFEM domain is conserved; only the most N-terminal α-helix of the domain is tilted.

In contrast, the structure of the *Ct*Pth11 CFEM domain reveals two potential ligand binding sites, placed vis-à-vis each other (see chapter 4. 4. 4, Figure 26). The potential binding sites are both hydrophobic in their inside; the larger one has some positively charged residues at its entrance (K80, K92, K104), the smaller one a negatively charged one (E49). Fragment screening revealed that the bound fragments are all located in the larger cavity in the CFEM domain (see chapters 4. 4. 5 and 5. 3. 3). Depending on the orientation of F48, the cavities are either divided or a tunnel through the molecule is formed. The properties of the tunnel were analyzed using *MOLEonline*[179] (see Figure 37).

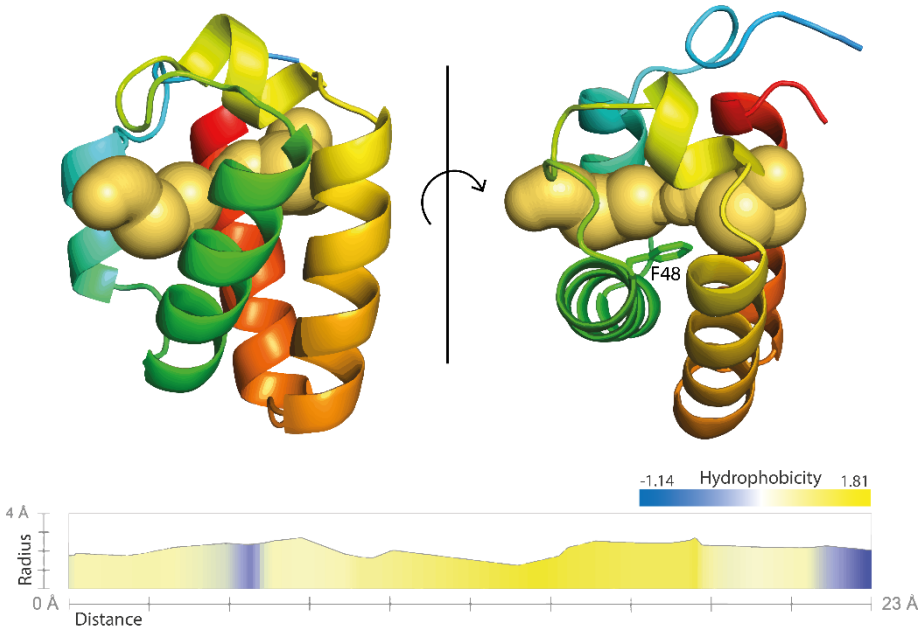


Figure 37: Analysis of the tunnel through the *Ct*Pth11 CFEM domain

A cartoon representation of the *Ct*Pth11 CFEM domain in two different orientations is shown (N-terminus blue, C-terminus red). The tunnel through the domain depicted as yellow spheres, it was analyzed using *MOLEonline*[179]. Chain D was chosen for the analysis; F48 is orientated in a way that does not divide the two cavities in the domain. A graphical representation of hydrophobicity and diameter along the length of the tunnel is shown in the lower part of the figure. The hydrophobicity is shown as a normalized scale that ranges from the most hydrophilic residue (E with -1.14) to the most hydrophobic one (I with 1.81), as described by Cid *et al.*[180].

The tunnel through the *Ct*Pth11 CFEM domain has a length of 23 Å and is mostly lined by hydrophobic residues. The calculation of its diameter – performed on chain D of the asymmetric unit – reveals a bottleneck of 1.2 Å, located at F48. This small diameter does not suggest that a molecule would fit through the tunnel, as it is almost as small as the van der Waals radius of a hydrogen atom (1.09 Å[181]). However, a change in the placement of the F48 side chain might widen the tunnel enough to allow some ligands, such as fatty acids, to fit through the tunnel. Pth11 is suggested to induce the differentiation of appressoria upon sensing either specific plant surface cues or hydrophobicity. This was shown by detection of increased appressorium formation of *M. grisea* on polystyrene and Teflon supplemented with

1,16-hexadecanediol (which contains a fatty acid chain), compared to the unsupplemented surfaces. The effect could not be observed in *M. grisea* mutants with a defective pth11 gene[44]. Fatty acids are common components in the plant cuticle[182] and the question if these are the unknown surface cue that Pth11 senses remains open. Further examination of a possible interaction between Pth11 and fatty acid chains via molecular dynamics simulations are therefore suggested.

5. 3. 2. Accessibility of the binding cleft

The structural analysis of the *Ct*Pth11 CFEM domain revealed two possible ligand binding sites. However, it must be considered that Pth11 does not only consist of the CFEM domain, but also has a transmembrane region and a cytoplasmic C-terminal region. The orientation of the CFEM domain on the transmembrane region may not allow binding of a ligand at the potential binding sites due to limited accessibility. This problem was analyzed by prediction of residue-residue interactions, using *GREMLIN* (http://gremlin.bakerlab.org). *GREMLIN* conducts a covariance prediction; the input for which is a multiple sequence alignment. For the positions that vary in different proteins following assumption is made: when amino acid X varies, then amino acid Y interacting with X will also vary; the amino acids “co-vary”. These residues are usually found to interact with each other[183].

The covariance prediction for the N-terminal CFEM domain of *Ct*Pth11 and the transmembrane part was done using both sequences separately as input. The CFEM domain (stretching from V24 – S105) was aligned with 1155 sequences, the GPCR region (L109 – R380) with 6529 sequences. When joined, they were aligned with 344 sequences. The cytoplasmic C-terminal domain of *Ct*Pth11 was not included in the prediction. The contact map generated by *GREMLIN* is shown in Appendix IX, as is the full list of residues predicted to interact with each other. In general, many residues with a distance of three to four amino acids between each other were predicted to be adjacent, indicating the presence of α-helices. This demonstrates the reliability of the prediction, as it reflects the structure of both parts of CtPth11 – i. e. the CFEM domain and the transmembrane region. The predicted residue-residue interactions were used to generate a model of *Ct*Pth11 using *MODELLER*[118], which is shown in Figure 38. The model structure reveals that both potential bindings sites within the CFEM domain are accessible. Following residue pairs were predicted to be neighbors with a high probability: K86 – H259, K86 – F176, T90 – F173, and N93 – F173. K86, T90 and N93 are part of the most C-terminal α-helix of the CFEM domain. F173 and F176 are located between the second and the third α-helix of the transmembrane region; H259 is part of a longer loop between the forth and the fifth transmembrane helix. All three residues from the transmembrane region of *Ct*Pth11 are located on the extracellular side of the transmembrane region. Thus, the predicted interactions are indeed possible; they are shown in the model structure in Figure 38 B. It should be considered that loop regions and side-chain conformations cannot be modeled precisely, leading to unexpectedly long distances between interacting residues in the model structure.

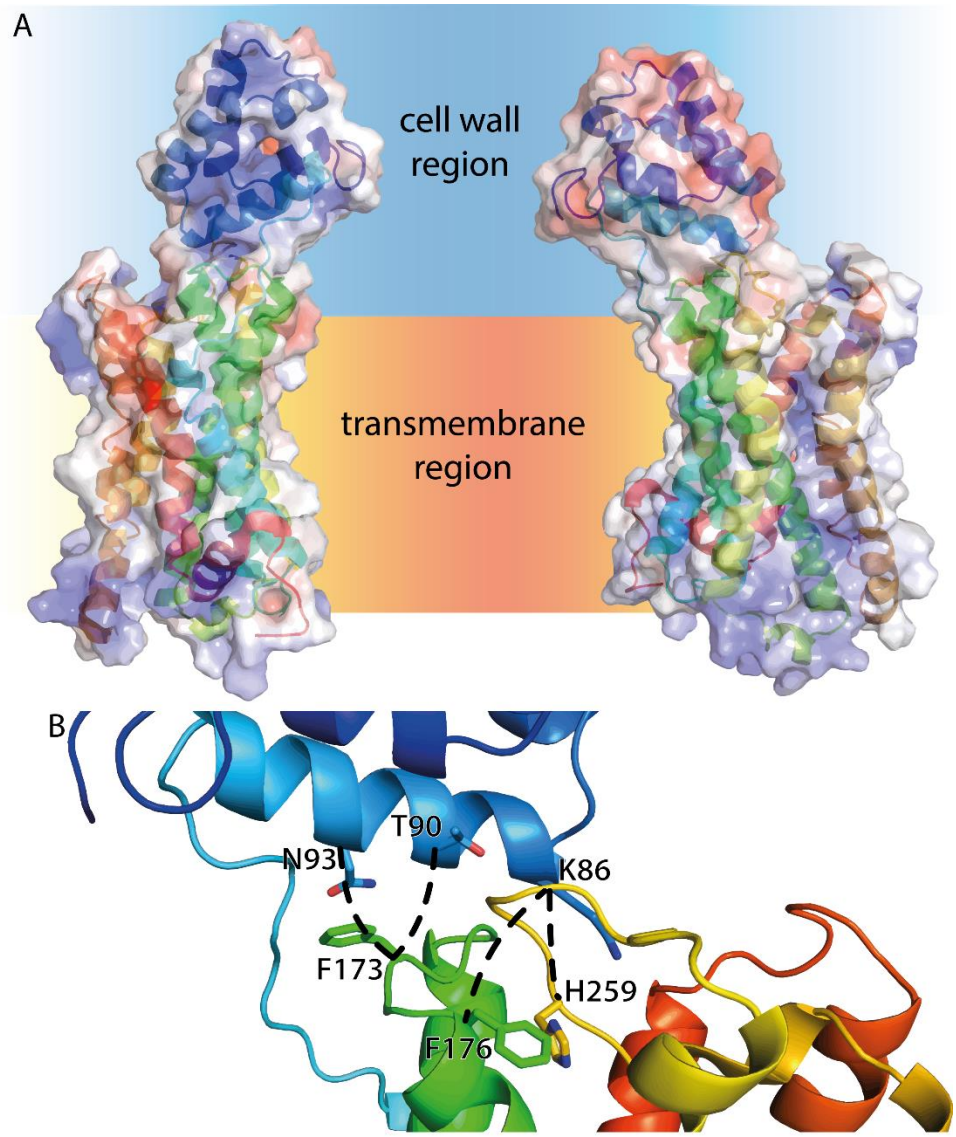


Figure 38: Model of the *Ct*Pth11 CFEM domain and the transmembrane region

A) The model of the *Ct*Pth11 CFEM domain and transmembrane region is shown in two orientations. The protein is depicted in cartoon and surface representation; the electrostatic potential of the surface was visualized using the APBS Electrostatics plugin in *PyMOL*. The model exhibits that both potential ligand binding sites are accessible. B) The region harboring residues predicted to be neighbors is shown. Interacting residues are shown as sticks and connected by dashed lines.

5. 3. 3. Fragment screening against the CtPth11 CFEM domain

The "resolution revolution" enabled the collection of data with higher and higher resolutions using single particle cryo-EM, a method that does not rely on formation of protein crystals[80,184,185]. This has changed the current and future perspectives on the applications of X-ray crystallography, which is also reflected by the more recent developments in the field. Besides the development of XFELs that allow the acquisition of time-resolved crystallography data, the speed of data acquisition at synchrotron beamlines and the applications running automated data analysis have extensively improved. X-ray crystallography is thus perfectly

suited for structure-based drug-discovery, which has become a commonly used method[185]. In this work, a fragment screen against the *Ct*Pth11 CFEM domain was conducted, serving two purposes: first, further information on the potential natural ligand of Pth11 should be gained. Second, some potential inhibitors, which may be of use for the development of antifungal agents for agriculture, may be identified. Fragments contained in the Frag Xtal Screen (*Jena Bioscience*) were used.

The automatic data analysis pipeline *DIMPLE*[106] was used for the evaluation of the fragment screen datasets. 118 records were handled using *DIMPLE*, 21 of those could not be handled by the pipeline. Since the diffraction quality of many crystals was severely compromised by the soaking process, *DIMPLE* can nevertheless be considered a reliable method for rapid structure solution of multiple datasets. The pipeline also identifies so called "unmodelled blobs" in the electron density maps – i. e. regions of electron density that do not contain any structure model. Bound fragments were observed in four datasets, but *DIMPLE* was able to identify unmodelled blobs in only one of them. Manual evaluation of the electron density maps of the datasets is therefore considered as necessary. The fragments that were bound the *Ct*Pth11 CFEM domain are shown in Figure 27.

The electron densities for fragment 3, fragment 34 and fragment 62 were unambiguous; fragment 94 could not be modelled into the electron density in any meaningful way. All fragments are placed in the larger cavity of the CFEM domain (see Figure 39). Only few interactions are formed between the compounds and residues from the CFEM domain: C66 and N72 interact with fragment 3; N72, T76 and T95 interact with fragment 34; fragment 62 does not seem to interact with any residue. In general, the hydrophobic compounds are located in the hydrophobic cavity. This is in agreement with the suggestion that Pth11 might sense hydrophobicity on the plant surface[44].

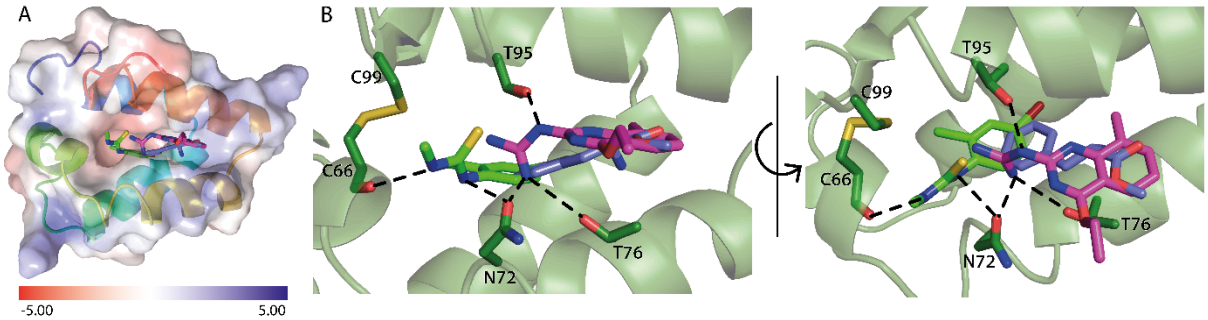


Figure 39: Placement of the bound fragments within the *Ct*Pth11 CFEM domain's cavity

A) The cartoon and surface representation of the CtPth11 CFEM domain shows that the fragments are all bound in the same cavity. The APBS Electrostatics plugin for *PyMOL* was used to generate the surface representation. The cavity, in which the ligands are placed, is hydrophobic, some positively charged residues are placed at its entrance. B) An overlay of all three bound fragments in two different orientations. The CFEM domain is shown in cartoon representation, the residues interacting with the ligands are depicted as sticks. Dashed lines indicate the interactions between protein and fragment.

5. 4. Perspectives on structural proteomics of the fungal cell wall

The topics covered in this thesis raise excellent opportunities for further research. By establishing the cell wall proteome of the thermophilic fungus *C. thermophilum*, several targets for biochemical and structural characterization could be identified. The usefulness of proteins originating from *C. thermophilum* for *in vitro* studies of fungal cell wall proteins was recently described by Vogt *et al.*[10]. It has also been demonstrated in this work by characterizing the CFEM domain of the GPCR Pth11. The increased stability of *C. thermophilum* proteins compared to their mesophilic counterparts, which was observed in both cases, might also be transferable to other targets of interest for characterization. These include Ecm33, which is regarded one of the most abundant cell wall proteins and implemented in cell wall integrity[131,186], as well as Gel1/2 and Kre9, which are both involved in cell wall biosynthesis[2,142,143]. The analysis of the *C. thermophilum* cell wall proteome also posed questions regarding the distinction between GPI-PMPs and GPI-CWPs, as many proteins that were expected to be located at the plasma membrane could be identified in cell wall isolates. Characterization of Cdc1 may provide further insight into cell wall sorting, turning it into another target for future biochemical and structural studies.

The structures of Awp1A and Awp3A, which were determined in this work, represent a new class of *C. glabrata* adhesins. In contrast to the Epa and Pwp adhesin families, which both have a PA14 domain, the cluster VI adhesins Awp1 and Awp3 were shown to contain a right-handed parallel β-helix. By generation of a SSN, the presence of this structural motif could also be revealed in the cluster V adhesins Awp2 and Awp4. *C. glabrata* contains a large repertoire of adhesins with partially overlapping functions and extensive differences between various strains or isolates. The diversity can evolve rapidly due to the high plasticity of the organism's genome, a characteristic often observed in pathogens[23,27]. Accordingly, the characterization of the various adhesins in fungal pathogens is a future objective. A reliable classification of adhesin families in combination with the characterization of individual members enables the prediction of the other proteins contained in the respective families. The foundations for this have been established by the characterization of various Epa proteins[26,29,30] and in this work. In addition, the SSN also revealed the similarity of Awp1/3 to protein families from other fungal organisms, thereby allowing the prediction of structures of members of the Iff family of adhesins from the pathogen *C. albicans* or of the bacterial cell surface proteins that are included in this network.

Also the analysis of the structure of the *Ct*Pth11 CFEM domain offers new possibilities to gain further insight on the protein. The structure reveals a hydrophobic tunnel through the molecule, with a bottleneck diameter of 1.2 Å. However, the diameter of the tunnel may be enlarged by displacement of side chain of F48, which may allow binding of a fatty acid chain. The analysis of this possible interaction using molecular dynamics simulations is therefore suggested. The accessibility of the potential ligand binding sites, respectively the tunnel, was predicted using the *GREMLIN* server, which conducts a sequence covariance analysis. The predicted residue-residue interactions were used to generate a model of the CFEM domain,

placed on the transmembrane region of Pth11 (see Figure 38). In cooperation with Prof. Dr. Neil Brown, the presence of the predicted interactions will be verified in *F. graminearum* Pth11 (FGRRES_16221)[47]. The corresponding interactions in FGRRES_16221 were determined using a model of the *Fg*Pth11 CFEM domain and transmembrane region, based on the *Ct*Pth11 model. Following residues are thought to interact with each other in FGRRES_16221: K76 – H249, K76 – F168, T80 – I165/F168, L79 – I165. These will be mutated to alanine residues in the mutation studies. Using the fragment screen, four compounds were identified to bind to the *Ct*Pth11 CFEM domain. These are fragment 3, 34, 62, and 94 from the Frag Xtal Screen (*Jena Bioscience*). Affinities of the CFEM domain and the compounds should be tested in future experiments. In addition, *in vivo* studies might be conducted to determine, if the fragments act as inhibitors for Pth11, a principal contributor to invasive fungal growth in plants.

6. Literature

1. Gow NAR, Latge J, Munro CA. The Fungal Cell Wall : Structure , Biosynthesis , and Function. *Microbiol Spectr*. 2017;5(3):1-25. doi:10.1128/microbiolspec.FUNK-0035-2016.Correspondence

2. Orlean P. Architecture and biosynthesis of the Saccharomyces cerevisiae cell wall. *Genetics*. 2012;192(3):775-818. doi:10.1534/genetics.112.144485

3. Klis FM, Boorsma A, De Groot PWJ. Cell wall construction in Saccharomyces cerevisiae. *Yeast*. 2006;23(3):185-202. doi:10.1002/yea.1349

4. Georgopapadakou NH, Tkacz JS. The fungal cell wall as a drug target. *Trends Microbiol*. 1995;3(3):98-104. doi:10.1016/S0966-842X(00)88890-3

5. Free SJ. *Fungal Cell Wall Organization and Biosynthesis*. Vol 81. 1st ed. Elsevier Inc.; 2013. doi:10.1016/B978-0-12-407677-8.00002-6

6. Cassone A. Development of vaccines for Candida albicans: fighting a skilled transformer. *Nat Rev Microbiol*. 2013;11(12):884-891. doi:10.1038/nrmicro3156

7. Erwig LP, Gow NAR. Interactions of fungal pathogens with phagocytes. *Nat Rev Microbiol*. 2016;14(3):163-176. doi:10.1038/nrmicro.2015.21

8. Ecker M, Deutzmann R, Lehle L, Mrsa V, Tanner W. Pir proteins of Saccharomyces cerevisiae are attached to β-1,3-glucan by a new protein-carbohydrate linkage. *J Biol Chem*. 2006. doi:10.1074/jbc.M600314200

9. Orlean P, Menon AK. Thematic review series: Lipid Posttranslational Modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids . *J Lipid Res*. 2007. doi:10.1194/jlr.r700002-jlr200

10. Vogt MS, Schmitz GF, Varón Silva D, Mösch HU, Essen LO. Structural base for the transfer of GPI-anchored glycoproteins into fungal cell walls. *Proc Natl Acad Sci U S A*. 2020. doi:10.1073/pnas.2010661117

11. de Groot PWJ, Hellingwerf KJ, Klis FM. Genome-wide identification of fungal GPI proteins. *Yeast*. 2003;20(9):781-796. doi:10.1002/yea.1007

12. Eisenhaber B, Schneider G, Wildpaner M, Eisenhaber F. A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for Aspergillus nidulans, Candida albicans Neurospora crassa, Saccharomyces cerevisiae and Schizosaccharomyces pombe. *J Mol Biol*. 2004. doi:10.1016/j.jmb.2004.01.025

13. Caro LHP, Tettelin H, Vossen JH, Ram AFJ, Van Den Ende H, Klis FM. In silicio identification of glycosyl-phosphatidylinositol-anchored plansma-membrane and cell wall proteins of Saccharomyces cerevisiae. *Yeast*. 1997. doi:10.1002/(SICI)1097-0061(199712)13:15<1477::AID-YEA184>3.0.CO;2-L

14. Perutz MF, Raidt H. Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature*. 1975. doi:10.1038/255256a0

15. Lawyer FC, Stoffel S, Saiki RK, et al. High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *Genome Res*. 1993. doi:10.1101/gr.2.4.275

16. de Oliveira TB, Gomes E, Rodrigues A. Thermophilic fungi in the new age of fungal taxonomy. *Extremophiles*. 2015;19(1):31-37. doi:10.1007/s00792-014-0707-0

17. Cava F, Hidalgo A, Berenguer J. Thermus thermophilus as biological model. *Extremophiles*. 2009. doi:10.1007/s00792-009-0226-6

18. Bock T, Chen WH, Ori A, et al. An integrated approach for genome annotation of the eukaryotic thermophile Chaetomium thermophilum. *Nucleic Acids Res*. 2014. doi:10.1093/nar/gku1147

19. Amlacher S, Sarges P, Flemming D, et al. Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell*. 2011. doi:10.1016/j.cell.2011.06.039

20. Baßler J, Ahmed YL, Kallas M, et al. Interaction network of the ribosome assembly machinery from a eukaryotic thermophile. *Protein Sci*. 2017;26(2):327-342. doi:10.1002/pro.3085

21. Ishida T, Fushinobu S, Kawai R, Kitaoka M, Igarashi K, Samejima M. Crystal structure of glycoside hydrolase family 55-β-1,3-glucanase from the basidiomycete Phanerochaete chrysosporium. *J Biol Chem*. 2009. doi:10.1074/jbc.M808122200

22. de Groot PWJ, Bader O, de Boer AD, Weig M, Chauhan N. Adhesins in human fungal pathogens: Glue with plenty of stick. *Eukaryot Cell*. 2013;12(4):470-481. doi:10.1128/EC.00364-12

23. Kaur R, Domergue R, Zupancic ML, Cormack BP. A yeast by any other name: Candida glabrata and its interaction with the host. *Curr Opin Microbiol*. 2005;8(4):378-384. doi:10.1016/j.mib.2005.06.012

24. De Groot PWJ, Kraneveld EA, Qing YY, et al. The cell wall of the human pathogen Candida glabrata: Differential incorporation of novel adhesin-like wall proteins. *Eukaryot Cell*. 2008;7(11):1951-1964. doi:10.1128/EC.00284-08

25. Kraneveld EA, de Soet JJ, Deng DM, et al. Identification and Differential Gene Expression of Adhesin-Like Wall Proteins in Candida glabrata Biofilms. *Mycopathologia*. 2011;172(6):415-427. doi:10.1007/s11046-011-9446-2

26. Diderrich R, Kock M, Maestre-Reyna M, et al. Structural Hot Spots Determine Functional Diversity of the Candida glabrata Epithelial Adhesin Family. *J Biol Chem*. 2015;290(32):19597-19613. doi:10.1074/jbc.M115.655654

27. Xu Z, Green B, Benoit N, Schatz M, Wheelan S, Cormack B. De novo genome assembly of Candida glabrata reveals cell wall protein complement and structure of dispersed tandem repeat arrays. *Mol Microbiol*. 2020. doi:10.1111/mmi.14488

28. Gómez-Molero E, de Boer AD, Dekker HL, et al. Proteomic analysis of hyperadhesive candida glabrata clinical isolates reveals a core wall proteome and differential incorporation of adhesins. *FEMS Yeast Res*. 2015;15(8):1-10.

doi:10.1093/femsyr/fov098

29. Maestre-reyna M, Diderrich R, Stefan M, Eulenburg G, Kalugin V, Brückner S. Structural basis for promiscuity and speci fi city during Candida glabrata invasion of host epithelia. 2012:1-6. doi:10.1073/pnas.1207653109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1207653109

30. Hoffmann D, Diderrich R, Reithofer V, et al. Functional reprogramming of Candida glabrata epithelial adhesins: the role of conserved and variable structural motifs in ligand binding. *J Biol Chem*. 2020. doi:10.1074/jbc.RA120.013968

31. Timmermans B, Peñas AD Las, Castaño I, Van Dijck P. Adhesins in candida glabrata. *J Fungi*. 2018;4(2):1-16. doi:10.3390/jof4020060

32. Desai C, Mavrianos J, Chauhan N. Candida glabrata Pwp7p and Aed1p are required for adherence to human endothelial cells. *FEMS Yeast Res*. 2011;11(7):595-601. doi:10.1111/j.1567-1364.2011.00743.x

33. Shimoi H, Sakamoto K, Okuda M, Atthi R, Iwashita K, Ito K. The AWA1 gene is required for the foam-forming phenotype and cell surface hydrophobicity of sake yeast. *Appl Environ Microbiol*. 2002. doi:10.1128/AEM.68.4.2018-2025.2002

34. Waters EJ, Pellerin P, Brillouet JM. A Saccharomyces mannoprotein that protects wine from protein haze. *Carbohydr Polym*. 1994. doi:10.1016/0144-8617(94)90101-5

35. Brown SL, Stockdale VJ, Pettolino F, et al. Reducing haziness in white wine by overexpression of Saccharomyces cerevisiae genes YOL155c and YDR055w. *Appl Microbiol Biotechnol*. 2007. doi:10.1007/s00253-006-0606-0

36. Kulkarni RD, Kelkar HS, Dean RA. An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. *Trends Biochem Sci*. 2003;28(3):118-121. doi:10.1016/S0968-0004(03)00025-2

37. Nasser L, Weissman Z, Pinsky M, Amartely H, Dvir H, Kornitzer D. Structural basis of haem-iron acquisition by fungal pathogens. *Nat Microbiol*. 2016;1(11):16156. doi:10.1038/nmicrobiol.2016.156

38. Sabnam N, Roy Barman S. WISH, a novel CFEM GPCR is indispensable for surface sensing, asexual and pathogenic differentiation in rice blast fungus. *Fungal Genet Biol*. 2017;105(May):37-51. doi:10.1016/j.fgb.2017.05.006

39. Moukadiri I, Armero J, Abad A, Sentandreu R, Zueco J. Identification of a mannoprotein present in the inner layer of the cell wall of Saccharomyces cerevisiae. *J Bacteriol*. 1997;179(7):2154-2162. doi:10.1128/jb.179.7.2154-2162.1997

40. Kou Y, Tan YH, Ramanujam R, Naqvi NI. Structure-function analyses of the Pth11 receptor reveal an important role for CFEM motif and redox regulation in rice blast. *New Phytol*. 2017;214(1):330-342. doi:10.1111/nph.14347

41. Kalugin V. Struktur- und Funktionsanalysen von pilzlichen Zellwandproteinen der SUN- und CFEM-Proteinfamilien. 2017. doi:https://doi.org/10.17192/z2017.0509

42. Choi W, Dean RA. The adenylate cyclase gene MAC1 of Magnaporthe grisea controls appressorium formation and other aspects of growth and development. *Plant Cell*.

1997;9(11):1973-1983. doi:10.1105/tpc.9.11.1973

43. Deng J, Dean RA. Characterization of Adenylate Cyclase Interacting Protein ACI1 in the Rice Blast Fungus, Magnaporthe oryzae. *Open Mycol J*. 2008. doi:10.2174/1874437000802010074

44. DeZwaan TM, Carroll AM, Valent B, Sweigard JA. Magnaporthe grisea Pth11p is a novel plasma membrane protein that mediates appressorium differentiation in response to inductive substrate cues. *Plant Cell*. 1999;11(10):2013-2030. doi:10.1105/tpc.11.10.2013

45. Zhang H, Zheng X, Zhang Z. The Magnaporthe grisea species complex and plant pathogenesis. *Mol Plant Pathol*. 2016. doi:10.1111/mpp.12342

46. Yan X, Talbot NJ. Investigating the cell biology of plant infection by the rice blast fungus Magnaporthe oryzae. *Curr Opin Microbiol*. 2016. doi:10.1016/j.mib.2016.10.001

47. Dilks T, Halsey K, De Vos RP, Hammond-Kosack KE, Brown NA. Non-canonical fungal G-protein coupled receptors promote Fusarium head blight on wheat. *PLoS Pathog*. 2019. doi:10.1371/journal.ppat.1007666

48. Kellner N, Schwarz J, Sturm M, et al. Developing genetic tools to exploit Chaetomium thermophilum for biochemical analyses of eukaryotic macromolecular assemblies. *Sci Rep*. 2016. doi:10.1038/srep20937

49. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*. 1986. doi:10.1016/0022-2836(86)90385-2

50. DelProposto J, Majmudar CY, Smith JL, Brown WC. Mocr: A novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expr Purif*. 2009. doi:10.1016/j.pep.2008.08.011

51. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr*. 2011;67(4):235-242. doi:10.1107/S0907444910045749

52. Adams PD, Afonine P V., Bunkóczi G, et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr*. 2010;66(2):213-221. doi:10.1107/S0907444909052925

53. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr*. 2010;66(4):486-501. doi:10.1107/S0907444910007493

54. Kabsch W. *XDS*. *Acta Crystallogr Sect D Biol Crystallogr*. 2010. doi:10.1107/S0907444909047337

55. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc*. 2008;3(7):1171-1179. doi:10.1038/nprot.2008.91

56. Cline MS, Smoot M, Cerami E, et al. Integration of biological networks and gene

expression data using cytoscape. *Nat Protoc*. 2007;2(10):2366-2382. doi:10.1038/nprot.2007.324

57. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: Architecture and applications. *BMC Bioinformatics*. 2009. doi:10.1186/1471-2105-10-421

58. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7(1):539. doi:10.1038/msb.2011.75

59. Gasteiger E, Hoogland C, Gattiker A, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: *The Proteomics Protocols Handbook*. ; 2005. doi:10.1385/1-59259-890-0:571

60. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med*. 2001;40(4):346-358. doi:10.1055/s-0038-1634431

61. Nielsen H, Tsirigos KD, Brunak S, von Heijne G. A Brief History of Protein Sorting Prediction. *Protein J*. 2019. doi:10.1007/s10930-019-09838-3

62. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019. doi:10.1038/s41587-019-0036-z

63. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol*. 2001. doi:10.1006/jmbi.2000.4315

64. Hernáez ML, Ximénez-Embún P, Martínez-Gomariz M, Gutiérrez-Blázquez MD, Nombela C, Gil C. Identification of Candida albicans exposed surface proteins in vivo by a rapid proteomic approach. *J Proteomics*. 2010. doi:10.1016/j.jprot.2010.02.008

65. Karkowska-Kuleta J, Kozik A. Cell wall proteome of pathogenic fungi. *Acta Biochim Pol*. 2015. doi:10.18388/abp.2015_1032

66. De Groot PWJ, De Boer AD, Cunningham J, et al. Proteomic analysis of Candida albicans cell walls reveals covalently bound carbohydrate-active enzymes and adhesins. *Eukaryot Cell*. 2004. doi:10.1128/EC.3.4.955-965.2004

67. Franken LE, Grünewald K, Boekema EJ, Stuart MCA. A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. *Small*. 2020. doi:10.1002/smll.201906198

68. K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn HE. Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. 1986:11.

69. Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-Van Dillen PME, Van Der Noordaa J. Rapid and simple method for purification of nucleic acids. *J Clin Microbiol*. 1990;28(3):495-503. doi:10.1128/jcm.28.3.495-503.1990

70. Blokesch M. Natural competence for transformation. *Curr Biol*. 2016;26(21):R1126-R1130. doi:10.1016/j.cub.2016.08.058

71. Hanahan D, Harbor CS. Studies o n T r a n s f o r m a t i o n o f Escherichia coli with P l a

s m i d s Department of Biochemistry and Molec ~ dar Biology. 1983;0:557-580.

72. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology Of The Cell, 4th Edition*.; 2002.

73. Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res*. 1979;7(6):1513-1523. https://linkinghub.elsevier.com/retrieve/pii/S1874391913002522.

74. Bachman J. Site-Directed Mutagenesis. *Methods Enzymol*. 2013;529:241-248. doi:10.1016/B978-0-12-418687-3.00019-7

75. Aslanidis C, de Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res*. 1990;18(20):6069-6074. doi:10.1093/nar/18.20.6069

76. Block H, Maertens B, Spriestersbach A, et al. Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC). A Review. *Methods Enzymol*. 2009;463(C):439-473. doi:10.1016/S0076-6879(09)63027-5

77. Ó'Fágáin C, Cummins PM, O'Connor BF. Gel-Filtration Chromatography. *Protein Chromatogr*. 2017;1485:423. doi:10.1007/978-1-4939-6412-3

78. Olson BJSC. Assays for determination of protein concentration. *Curr Protoc Pharmacol*. 2016;2016:A.3A.1-A.3A.32. doi:10.1002/cpph.3

79. Heimburg-Molinaro J, Song X, Smith DF, Cummings RD. Preparation and analysis of glycan microarrays. *Curr Protoc Protein Sci*. 2011. doi:10.1002/0471140864.ps1210s64

80. Kühlbrandt W. The resolution revolution. *Science (80-)*. 2014;343:1443-1444. doi:10.1126/science.360.6386.280-k

81. Martin-Garcia JM, Conrad CE, Coe J, Roy-Chowdhury S, Fromme P. Serial femtosecond crystallography: A revolution in structural biology. *Arch Biochem Biophys*. 2016;602:32-47. doi:10.1016/j.abb.2016.03.036

82. McPherson A, Gavira JA. Introduction to protein crystallization. *Acta Crystallogr Sect FStructural Biol Commun*. 2014;70(1):2-20. doi:10.1107/S2053230X13033141

83. Chayen NE, Saridakis E. Protein crystallization: From purified protein to diffraction-quality crystal. *Nat Methods*. 2008;5(2):147-153. doi:10.1038/nmeth.f.203

84. Chayen NE. Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta Crystallogr Sect D Biol Crystallogr*. 1998. doi:10.1107/S0907444997005374

85. Gorrec F. The MORPHEUS II protein crystallization screen. *Acta Crystallogr Sect Struct Biol Commun*. 2015;71:831-837. doi:10.1107/S2053230X1500967X

86. Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three dimensional structure. *EMBO J*. 1990;9(5):1665-1672. doi:10.1002/j.1460-2075.1990.tb08287.x

87. Pike ACW, Garman EF, Krojer T, Von Delft F, Carpenter EP. An overview of heavy-atom derivatization of protein crystals. *Acta Crystallogr Sect D Struct Biol*. 2016.

doi:10.1107/S2059798316000401

88. Baker M. Fragment-based lead discovery grows up. *Nat Rev Drug Discov*. 2013;12(1):5-7. doi:10.1038/nrd3926

89. Dauter Z. Collection of X-ray diffraction data from macromolecular crystals. *Methods Mol Biol*. 2017;1607:165-184. doi:10.1007/978-1-4939-7000-1

90. Gabadinho J, Beteva A, Guijarro M, et al. MxCuBE: A synchrotron beamline control environment customized for macromolecular crystallography experiments. *J Synchrotron Radiat*. 2010;17(5):700-707. doi:10.1107/S0909049510020005

91. Wojdyla JA, Kaminski JW, Panepucci E, et al. DA+ data acquisition and analysis software at the Swiss Light Source macromolecular crystallography beamlines. *J Synchrotron Radiat*. 2018;25(1):293-303. doi:10.1107/S1600577517014503

92. Drenth J, Mesters J. *Principles of Protein X-Ray Crystallography: Third Edition*.; 2007. doi:10.1007/0-387-33746-6

93. Basu S, Finke A, Vera L, Wang M, Olieric V. Making routine native SAD a reality: lessons from beamline X06DA at the Swiss Light Source. *Acta Crystallogr Sect D Struct Biol*. 2019. doi:10.1107/S2059798319003103

94. Evans PR. An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr Sect D Biol Crystallogr*. 2011;67(4):282-292. doi:10.1107/S090744491003982X

95. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr Sect D Biol Crystallogr*. 2013;69(7):1204-1214. doi:10.1107/S0907444913000061

96. Taylor GL. Introduction to phasing. *Acta Crystallogr Sect D Biol Crystallogr*. 2010;66(4):325-338. doi:10.1107/S0907444910006694

97. Rose JP, Wang BC. SAD phasing: History, current impact and future opportunities. *Arch Biochem Biophys*. 2016. doi:10.1016/j.abb.2016.03.018

98. Veelders M, Essen LO. Complex Gadolinium-Oxo Clusters Formed along Concave Protein Surfaces. *ChemBioChem*. 2012. doi:10.1002/cbic.201200441

99. Girard É, Stelter M, Vicat J, Kahn R. A new class of lanthanide complexes to obtain high-phasing-power heavy-atom derivatives for macromolecular crystallography. In: *Acta Crystallographica - Section D Biological Crystallography*. ; 2003. doi:10.1107/S0907444903020511

100. Patterson AL. A fourier series method for the determination of the components of interatomic distances in crystals. *Phys Rev*. 1934. doi:10.1103/PhysRev.46.372

101. Skubák P, Pannu NS. Automatic protein structure solution from weak X-ray data. *Nat Commun*. 2013;4. doi:10.1038/ncomms3777

102. Rose JP, Wang BC, Weiss MS. Native SAD is maturing. *IUCrJ*. 2015. doi:10.1107/S2052252515008337

103. Basu S, Olieric V, Leonarski F, et al. Long-wavelength native-SAD phasing:

Opportunities and challenges. *IUCrJ*. 2019. doi:10.1107/S2052252519002756

104. Evans P, McCoy A. An introduction to molecular replacement. In: *Acta Crystallographica Section D: Biological Crystallography*. ; 2007. doi:10.1107/S0907444907051554

105. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*. 2007;40(4):658-674. doi:10.1107/S0021889807021206

106. Wojdyr M, Keegan R, Winter G, Ashton A. DIMPLE - a pipeline for the rapid generation of difference maps from protein crystals with putatively bound ligands . *Acta Crystallogr Sect A Found Crystallogr*. 2013. doi:10.1107/s0108767313097419

107. Vagin AA, Steiner RA, Lebedev AA, et al. REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr Sect D Biol Crystallogr*. 2004;60(12 I):2184-2195. doi:10.1107/S0907444904023510

108. Varki A, Cummings RD, Esko JD, et al. *Essentials of Glycobiology, Third Edition*.; 2017.

109. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: An open-source platform for biological-image analysis. *Nat Methods*. 2012. doi:10.1038/nmeth.2019

110. Zallot R, Oberg N, Gerlt JA. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry*. 2019. doi:10.1021/acs.biochem.9b00735

111. Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. Erratum to: SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell Fact*. 2016;15(1):124. doi:10.1186/s12934-016-0512-9

112. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW. iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr Sect D Biol Crystallogr*. 2011;67(4):271-281. doi:10.1107/S0907444910048675

113. Yoder MD, Jurnak F. The refined three-dimensional structure of pectate lyase C from Erwinia chrysanthemi at 2.2 Angstrom resolution. Implications for an enzymatic mechanism. *Plant Physiol*. 1995. doi:10.1104/pp.107.2.349

114. Jenkins J, Mayans O, Pickersgill R. Structure and evolution of parallel β-helix proteins. *J Struct Biol*. 1998;122(1-2):236-246. doi:10.1006/jsbi.1998.3985

115. Yoder MD, Lietzke SE, Jurnak F. Unusual structural features in the parallel β-helix in pectate lyases. *Structure*. 1993. doi:10.1016/0969-2126(93)90013-7

116. Grøftehauge MK, Hajizadeh NR, Swann MJ, Pohl E. Protein-ligand interactions investigated by thermal shift assays (TSA) and dual polarization interferometry (DPI). *Acta Crystallogr Sect D Biol Crystallogr*. 2015. doi:10.1107/S1399004714016617

117. Geissner A, Seeberger PH. Glycan Arrays: From Basic Biochemical Research to Bioanalytical and Biomedical Applications. *Annu Rev Anal Chem*. 2016;9(1):223-247. doi:10.1146/annurev-anchem-071015-041641

118. Šali A. MODELLER A Program for Protein Structure Modeling. *Comp protein Model by*

Satisf Spat restraints. 1993. doi:10.1006/jmbi.1993.1626

119. Sheldrick GM. Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr Sect D Biol Crystallogr*. 2010. doi:10.1107/S0907444909038360

120. Skubák P, Pannu NS. Automatic protein structure solution from weak X-ray data. *Nat Commun*. 2013. doi:10.1038/ncomms3777

121. Deller MC, Kong L, Rupp B. Protein stability: A crystallographer's perspective. *Acta Crystallogr Sect Struct Biol Commun*. 2016. doi:10.1107/S2053230X15024619

122. Frieman MB, Cormack BP. The ω-site sequence of glycosylphosphatidylinositol-anchored proteins in Saccharomyces cerevisiae can determine distribution between the membrane and the cell wall. *Mol Microbiol*. 2003. doi:10.1046/j.1365-2958.2003.03722.x

123. Ouyang H, Chen X, Lü Y, et al. One single basic amino acid at the ω-1 or ω-2 site is a signal that Retains glycosylphosphatidylinositol-anchored protein in the plasma membrane of Aspergillus fumigatus. *Eukaryot Cell*. 2013. doi:10.1128/EC.00351-12

124. Frieman MB, Cormack BP. Multiple sequence signals determine the distribution of glycosylphosphatidylinositol proteins between the plasma membrane and cell wall in Saccharomyces cerevisiae. *Microbiology*. 2004. doi:10.1099/mic.0.27420-0

125. Vazquez HM, Vionnet C, Roubaty C, Conzelmann A. Cdc1 removes the ethanolamine phosphate of the first mannose of GPI anchors and thereby facilitates the integration of GPI proteins into the yeast cell wall. *Mol Biol Cell*. 2014. doi:10.1091/mbc.E14-06-1033

126. Mouyna I, Fontaine T, Vai M, et al. Glycosylphosphatidylinositol-anchored glucanosyltransferases play an active role in the biosynthesis of the fungal cell wall. *J Biol Chem*. 2000. doi:10.1074/jbc.275.20.14882

127. Terashima H, Hamada K, Kitada K. The localization change of Ybr078w / Ecm33 , a yeast GPI-associated protein , from the plasma membrane to the cell wall , a ¡ ecting the cellular function. *FEMS Microbiol Lett*. 2003;218:175-180.

128. Gil-Bona A, Amador-García A, Gil C, Monteoliva L. The external face of Candida albicans: A proteomic view of the cell surface and the extracellular environment. *J Proteomics*. 2018;180(May 2017):70-79. doi:10.1016/j.jprot.2017.12.002

129. Pittet M, Conzelmann A. Biosynthesis and function of GPI proteins in the yeast Saccharomyces cerevisiae. *Biochim Biophys Acta - Mol Cell Biol Lipids*. 2007. doi:10.1016/j.bbalip.2006.05.015

130. Banerjee P, Wehle M, Lipowsky R, Santer M. A molecular dynamics model for glycosylphosphatidyl-inositol anchors: "flop down" or "lollipop"? *Phys Chem Chem Phys*. 2018;20(46):29314-29324. doi:10.1039/c8cp04059a

131. Pardo M, Monteoliva L, Vázquez P, et al. PST1 and ECM33 encode two yeast cell surface GPI proteins important for cell wall integrity. *Microbiology*. 2004;150(12):4157-4170. doi:10.1099/mic.0.26924-0

132. Popolo L, Gilardelli D, Bonfante P, Vai M. Increase in chitin as an essential response to defects in assembly of cell wall polymers in the ggp1Δ mutant of Saccharomyces cerevisiae. *J Bacteriol*. 1997;179(2):463-469. doi:10.1128/jb.179.2.463-469.1997

133. Chaffin WL. Candida albicans Cell Wall Proteins. *Microbiol Mol Biol Rev*. 2008. doi:10.1128/mmbr.00032-07

134. Gurgel IL da S, Jorge KT de OS, Malacco NLS de O, et al. The Aspergillus fumigatus Mucin MsbA Regulates the Cell Wall Integrity Pathway and Controls Recognition of the Fungus by the Immune System . *mSphere*. 2019. doi:10.1128/msphere.00350-19

135. Perlin DS. Current perspectives on echinocandin class drugs. *Future Microbiol*. 2011. doi:10.2217/fmb.11.19

136. Ben-Ami R. Treatment of Invasive Candidiasis: A Narrative Review. *J Fungi*. 2018. doi:10.3390/jof4030097

137. Wiederhold NP. Echinocandin Resistance in Candida Species: a Review of Recent Developments. *Curr Infect Dis Rep*. 2016;18(12). doi:10.1007/s11908-016-0549-2

138. Tiwari U, Das S, Tandon M, Ramachandran VG, Saha R. Vaccines for fungal infections. *Natl Med J India*. 2015;28(1):14-19.

139. Tso GHW, Reales-Calderon JA, Pavelka N. The Elusive Anti-Candida Vaccine: Lessons From the Past and Opportunities for the Future. *Front Immunol*. 2018;9(April):897. doi:10.3389/fimmu.2018.00897

140. Fang W, Sanz AB, Bartual SG, et al. Mechanisms of redundancy and specificity of the Aspergillus fumigatus Crh transglycosylases. *Nat Commun*. 2019. doi:10.1038/s41467-019-09674-0

141. van Leeuwe TM, Wattjes J, Niehues A, et al. A seven-membered cell wall related transglycosylase gene family in Aspergillus niger is relevant for cell wall integrity in cell wall mutants with reduced α-glucan or galactomannan. *Cell Surf*. 2020. doi:10.1016/j.tcsw.2020.100039

142. Mouyna I, Morelle W, Vai M, et al. Deletion of GEL2 encoding for a β(1-3)glucanosyltransferase affects morphogenesis and virulence in Aspergillus fumigatus. *Mol Microbiol*. 2005. doi:10.1111/j.1365-2958.2005.04654.x

143. Li R, Liu Z, Dong W, et al. The antifungal peptide CGA-N12 inhibits cell wall synthesis of Candida tropicalis by interacting with KRE9. *Biochem J*. 2020. doi:10.1042/BCJ20190678

144. Gil-Bona A, Reales-Calderon JA, Parra-Giraldo CM, Martinez-Lopez R, Monteoliva L, Gil C. The cell wall protein Ecm33 of Candida albicans is involved in chronological life span, morphogenesis, cell wall regeneration, stress tolerance, and host-cell interaction. *Front Microbiol*. 2016;7(FEB):1-14. doi:10.3389/fmicb.2016.00064

145. Romano J, Nimrod G, Ben-Tal N, et al. Disruption of the Aspergillus fumigatus ECM33 homologue results in rapid conidial germination, antifungal resistance and hypervirulence. *Microbiology*. 2006;152(7):1919-1928. doi:10.1099/mic.0.28936-0

146. Chabane S, Sarfati J, Ibrahim-Granet O, et al. Ecm33p Influences Conidial Cell Wall

Biosynthesis in Aspergillus fumigatus Glycosylphosphatidylinositol-Anchored Ecm33p Influences Conidial Cell Wall Biosynthesis in Aspergillus fumigatus. *Appl Environ Microbiol*. 2006;72(5):3259-3267. doi:10.1128/AEM.72.5.3259

147. Chen Y, Zhu J, Ying SH, Feng MG. The GPI-anchored protein Ecm33 is vital for conidiation, cell wall integrity, and multi-stress tolerance of two filamentous entomopathogens but not for virulence. *Appl Microbiol Biotechnol*. 2014;98(12):5517-5529. doi:10.1007/s00253-014-5577-y

148. Wyrebek M, Bidochka MJ. Variability in the Insect and Plant Adhesins, Mad1 and Mad2, within the Fungal Genus Metarhizium Suggest Plant Adaptation as an Evolutionary Force. *PLoS One*. 2013. doi:10.1371/journal.pone.0059357

149. Wang C, St. Leger RJ. The MAD1 adhesin of Metarhizium anisopliae links adhesion with blastospore production and virulence to insects, and the MAD2 adhesin enables attachment to plants. *Eukaryot Cell*. 2007. doi:10.1128/EC.00409-06

150. Bono JL, Jaber B, Fisher MA, et al. Genetic diversity and transcriptional analysis of the bys1 gene from Blastomyces dermatitidis. *Mycopathologia*. 2001. doi:10.1023/A:1013121812329

151. Maheshwari R, Bharadwaj G, Bhat MK. Thermophilic Fungi: Their Physiology and Enzymes. *Microbiol Mol Biol Rev*. 2000;64(3):461-488. doi:10.1128/mmbr.64.3.461-488.2000

152. Ielasi FS, Decanniere K, Willaert RG. The epithelial adhesin 1 (Epa1p) from the human-pathogenic yeast Candida glabrata: Structural and functional study of the carbohydrate-binding domain. *Acta Crystallogr Sect D Biol Crystallogr*. 2012;68(3):210-217. doi:10.1107/S0907444911054898

153. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr Sect D Biol Crystallogr*. 2004. doi:10.1107/S0907444904026460

154. Baelen S, Dewitte F, Clantin B, Villeret V. Structure of the secretion domain of HxuA from Haemophilus influenzae. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2013;69(12):1322-1327. doi:10.1107/S174430911302962X

155. Zheng Y, Huang CH, Liu W, et al. Crystal structure and substrate-binding mode of a novel pectate lyase from alkaliphilic Bacillus sp. N16-5. *Biochem Biophys Res Commun*. 2012;420(2):269-274. doi:10.1016/j.bbrc.2012.02.148

156. Thomas LM, Doan CN, Oliver RL, Yoder MD. Structure of pectate lyase a: Comparison to other isoforms. *Acta Crystallogr Sect D Biol Crystallogr*. 2002;58(6 II):1008-1015. doi:10.1107/S0907444902005851

157. Itoh T, Nakagawa E, Yoda M, Nakaichi A, Hibi T, Kimoto H. Structural and biochemical characterisation of a novel alginate lyase from Paenibacillus sp. str. FPU-7. *Sci Rep*. 2019. doi:10.1038/s41598-019-51006-1

158. Close DW, D'angelo S, Bradbury ARM. A new family of β-helix proteins with similarities to the polysaccharide lyases. *Acta Crystallogr Sect D Biol Crystallogr*. 2014;70(10):2583-2592. doi:10.1107/S1399004714015934

159. Liston SD, McMahon SA, Le Bas A, Suits MDL, Naismith JH, Whitfield C. Periplasmic depolymerase provides insight into ABC transporter-dependent secretion of bacterial capsular polysaccharides. *Proc Natl Acad Sci U S A*. 2018. doi:10.1073/pnas.1801336115

160. Sequeira S, Kavanaugh D, MacKenzie DA, et al. Structural basis for the role of serine-rich repeat proteins from Lactobacillus reuteri in gut microbe-host interactions. *Proc Natl Acad Sci U S A*. 2018. doi:10.1073/pnas.1715016115

161. Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem J*. 2010;432(3):437-444. doi:10.1042/BJ20101185

162. Martin RB, Richardson FS. Lanthanides as probes for calcium in biological systems. *Q Rev Biophys*. 1979;12(2):181-209. doi:10.1017/S0033583500002754

163. Gupta R, Jung E, Brunak S. *NetNGlyc: Prediction of N-Glycosylation Sites in Human Proteins.*; 2004.

164. Goto M. Protein O-glycosylation in fungi: Diverse structures and multiple functions. *Biosci Biotechnol Biochem*. 2007. doi:10.1271/bbb.70080

165. Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S. NetOglyc: Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J*. 1998. doi:10.1023/A:1006960004440

166. Gupta R, Jung E, Gooley AA, Williams KL, Brunak S, Hansen J. Scanning the available Dictyostelium discoideum proteome for O-linked GlcNAc glycosylation sites using neural networks. *Glycobiology*. 1999. doi:10.1093/glycob/9.10.1009

167. González M, Brito N, González C. High abundance of Serine/Threonine-rich regions predicted to be hyper-O-glycosylated in the secretory proteins coded by eight fungal genomes. *BMC Microbiol*. 2012. doi:10.1186/1471-2180-12-213

168. Boisramé A, Cornu A, da Costa G, Richard ML. Unexpected Role for a Serine/Threonine-Rich Domain in the Candida albicans Iff Protein Family ∇. *Eukaryot Cell*. 2011. doi:10.1128/EC.05044-11

169. Lazareva-Ulitsky B, Diemer K, Thomas PD. On the quality of tree-based protein classification. *Bioinformatics*. 2005. doi:10.1093/bioinformatics/bti244

170. Gerlt JA, Bouvier JT, Davidson DB, et al. Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta - Proteins Proteomics*. 2015. doi:10.1016/j.bbapap.2015.04.015

171. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*. 1997. doi:10.1002/elps.1150181505

172. Ma BQ, Zhang DS, Gao S, Jin TZ, Yan CH. The formation of Gd4O4 cubane cluster controlled by L-valine. *New J Chem*. 2000. doi:10.1039/a910181k

173. Fox BA, Yee VC, Pedersen LC, et al. Identification of the calcium binding site and a novel ytterbium site in blood coagulation factor XIII by x-ray crystallography. *J Biol Chem*. 1999. doi:10.1074/jbc.274.8.4917

174. Saxena AK, Singh K, Su HP, et al. The essential mosquito-stage P25 and P28 proteins from Plasmodium form tile-like triangular prisms. *Nat Struct Mol Biol*. 2006. doi:10.1038/nsmb1024

175. Gonçalves S, Esteves AM, Santos H, Borges N, Matias PM. Three-dimensional structure of mannosyl-3-phosphoglycerate phosphatase from thermus thermophilus HB27: A new member of the haloalcanoic acid dehalogenase superfamily. *Biochemistry*. 2011. doi:10.1021/bi201171h

176. Aime S, Caravan P. Biodistribution of gadolinium-based contrast agents, including gadolinium deposition. In: *Journal of Magnetic Resonance Imaging*. ; 2009. doi:10.1002/jmri.21969

177. Ramalho J, Semelka RC, Ramalho M, Nunes RH, AlObaidy M, Castillo M. Gadolinium-based contrast agent accumulation and toxicity: An update. *Am J Neuroradiol*. 2016. doi:10.3174/ajnr.A4615

178. Tomasello G, Armenia I, Molla G. The Protein Imager: A full-featured online molecular viewer interface with server-side HQ-rendering capabilities. *Bioinformatics*. 2020. doi:10.1093/bioinformatics/btaa009

179. Pravda L, Sehnal D, Toušek D, et al. MOLEonline: A web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic Acids Res*. 2018. doi:10.1093/nar/gky309

180. Cid H, Bunster M, Canales M, Gazitúa F. Hydrophobicity and structural classes in proteins. *Protein Eng Des Sel*. 1992. doi:10.1093/protein/5.5.373

181. Rowland RS, Taylor R. Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der Waals Radii. *J Phys Chem*. 1996. doi:10.1021/jp953141+

182. Zeisler-Diehl VV, Barthlott W, Schreiber L. Plant Cuticular Waxes: Composition, Function, and Interactions with Microorganisms. In: *Hydrocarbons, Oils and Lipids: Diversity, Origin, Chemistry and Fate*. ; 2018. doi:10.1007/978-3-319-54529-5_7-1

183. Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife*. 2014. doi:10.7554/eLife.02030

184. Hurtley SM. Continuing the resolution revolution. *Science (80-)*. 2018. doi:10.1126/science.360.6386.280-k

185. Grimes JM, Hall DR, Ashton AW, et al. Where is crystallography going? *Acta Crystallogr Sect D Struct Biol*. 2018. doi:10.1107/S2059798317016709

186. Martinez-Lopez R, Monteoliva L, Diez-Orejas R, Nombela C, Gil C. The GPI-anchored protein CaEcm33p is required for cell wall integrity, morphogenesis and virulence in Candida albicans. *Microbiology*. 2004;150(10):3341-3354. doi:10.1099/mic.0.27320-0

7. Acknowledgements - Danksagung

An erster Stelle möchte ich hier Prof. Dr. Lars-Oliver Essen danken, der mir die Möglichkeit geboten hat, diese Arbeit anzufertigen. Vielen Dank für das Angebot, an dieser interessanten Themenstellung zu arbeiten, die wissenschaftliche Betreuung und die Freiräume, die mir bei der Planung, Gestaltung und Ausführung meiner Experimente gegeben wurden. Außerdem möchte ich mich auch für Ihre Menschlichkeit bedanken!

Prof. Dr. Hans-Ulrich Mösch danke ich für die Übernahme des Zweitgutachtens und der Mitgliedschaft in meinem TAC im Rahmen der Graduate School des IMPRS. Auch für den wissenschaftlichen Input im Rahmen diverser Meetings und die Kooperation beim Epa-Projekt bin ich sehr dankbar. An dieser Stelle auch vielen Dank an Daniel Hoffmann für die gelungene Kooperation!

Für die Mitgliedschaft in meinem TAC möchte ich mich auch bei Dr. Tobias Erb bedanken. Auch Sie haben diese Arbeit mit Ihrem wissenschaftlichen Input bereichert.

Für die Übernahme des Vorsitzes der Prüfungskommission danke ich Prof. Dr. Ulrich Tallarek.

I would like to thank my collaborator Dr. Piet de Groot for the opportunity to work on Awp's. Thank you for the two weeks at the Universidad de Castilla-La Mancha in Albacete and for sharing your expertise regarding the isolation of fungal cell walls and the determination of the *C. thermophilum* cell wall proteome.

Prof. Dr. Ed Hurt und Sabine Griesel danke ich für die Zusendung von *C. thermophilum* Sporen und die Bereitstellung der Protokolle zum Anziehen des Pilzes. Bei Prof. Dr. Gerhard Klebe und Prof. Dr. Andreas Heine bedanke ich mich für die Bereitstellung des Fragment Screens. Für die Hilfe bei der Aufnahme der Datensets für native SAD Phasierung der *Ct*Pth11 Kristalle möchte ich mich bei Dr. Vincen Olieric vom SLS bedanken. Außerdem danke ich Ralf Pöschke für das Pipettieren der Kristallisationsscreens.

Ich möchte mich bei meinem Masterstudenten Simon Völpel für seine Arbeit am CFEM Projekt bedanken. Vielen Dank für dein Vertrauen bei der Betreuung deiner Arbeit, ich wünsche dir viel Erfolg auf deinem weiteren wissenschaftlichen Weg!

Auch bei meiner Masterstudentin Christin Schulz bedanke ich mich für ihren Beitrag beim CFEM Projekt. Außerdem danke ich dir für die vielen guten Gespräche, für seelischen Beistand und eine wunderbare schlaflose Nacht am ESRF. *Welcome to the Hotel California!*

Dr. Patrick Pausch danke ich für die Bereitstellung des wirklich allerletzten Aliquots an *C. thermophilum* cDNA, das – angeblich – unter Einsatz seines Lebens geborgen wurde. Vielen Dank auch für die zahlreichen Roadtrips ans ESRF – es war immer sehr unterhaltsam!

Für ihre tatkräftige Unterstützung bei der vorliegenden Arbeit möchte ich auch all meinen Vertiefungsstudenten danken!

Bei Dr. Manuel Maestre-Reyna möchte ich mich dafür bedanken, dass ich in die Experimente am SwissFEL eingebunden wurde. Es war eine unvergleichliche Woche in der Schweiz!

Auch meine lieben Kollegen haben die letzten Jahren sehr bereichert und das nicht nur auf wissenschaftlicher Ebene! Vielen Dank an Petra für ihren Beistand in allen Phasen dieser Doktorarbeit, du bist so ein toller Mensch! Vielen Dank auch an Dr. Ankan Banerjee, Dr. Vitali Kalugin, Elisabeth Ignatz, Dr. Sophie Franz-Badur, Thomas Marcellino, Dr. Marian Vogt, Dr. Sebastian Hepp, Maximilian Biermeier, Johannes Scheffer, Jonathan Trauth, Filipp Bezold, Laura Werel, Martin Saft, Hans-Joachim Emmerich, Sophia Hasenjäger, Bastian Pook, Lukas Korf, Silvana Griesel und Christin Geil. Auch bei Julia Witsch, Dr. Roberta Spadaccini und Dr. Christof Taxis möchte ich mich bedanken. Für viele gute Gespräche danke ich auch der „Mensa-Crew“: Dr. Stefan Brückner, Dr. Annika Brych und Dr. Stephan Kiontke. Anke-Dorothee Werner möchte ich für unterhaltsame Beamtimes und die Diskussionen über Fragment Screen Troubles danken.

Besonderer Dank geht natürlich auch nach Österreich: Meiner Familie danke ich für den Beistand in allen Lebenslagen. Ihr wart immer für mich da, sei es auch „nur“ telefonisch. Dr. Julia Sommer bin ich sehr dankbar für remote Beistand zu allen Zeiten!

Mein Dank geht auch an alle Freunde, die mich in diesem Lebensabschnitt begleitet haben. Selbst wenn ihr vielleicht nicht die gesamte Zeit über Teil meines Lebens wart, seid ihr doch bei mir gewesen und werdet mir in Erinnerung bleiben!

8. Appendices

8. 1. Appendix I: Classification of *C. glabrata* adhesins

Cluster I	CAGL0E06644g|**Epa1**
	CAGL0E06666g|**Epa2**
	CAGL0E06688g|**Epa3**
	CAGL0C00110g|**Epa6**
	CAGL0C05643g|**Epa7**
	CAGL0C00847g|**Epa8**
	CAGL0A01366g|**Epa9**
	CAGL0A01284g|**Epa10**
	CAGL0L13299g|**Epa11**
	CAGL0M00132g|**Epa12**
	CAGL0L13332g|**Epa13**
	CAGL0L13552g|**Epa14a**
	CAGL0M14300g|**Epa14b**
	CAGL0J11968g|**Epa15**
	CAGL0F00077g|**Epa16**
	CAGL0A00099g|**Epa19**
	CAGL0E00275g|**Epa20**
	CAGL0D06743g|**Epa21**
	CAGL0K00170g|**Epa22**
	CAGL0I00220g|**Epa23**
Cluster II	CAGL0I10147|**Pwp1**
	CAGL0I10246g|**Pwp2**
	CAGL0I10200g|**Pwp3**
	CAGL0I10362g|**Pwp4**
	CAGL0I10340g|**Pwp5**
	CAGL0M14069g|**Pwp6**
	CAGL0I10098g|**Pwp7**
Cluster III	CAGL0C00253g
	CAGL0E00165g
	CAGL0E01661g
	CAGL0L10092g
	CAGL0K13002g|**Aed2**
	CAGL0K13024g|**Awp5/Aed1**
	CAGL0E00231g
	CAGL0A04851g
	CAGL0H10626g|**Awp13**
	CAGL0G00099g
	CAGL0L00157g
	CAGL0I00209g
	CAGL0J00132g
	CAGL0A04873g|**Awp14**[24]
Cluster IV	CAGL0G10175g|**Awp6**
	CAGL0C00209g|**Awp7**
Cluster V	CAGL0E06600g
	CAGL0I07293g
	CAGL0B00154g
	CAGL0K00110g|**Awp2**
	CAGL0H00209g
	CAGL0I11000g
	CAGL0J12067g
	CAGL0F09251g
	CAGL0L00227g
	CAGL0B05061g
	CAGL0F00099g
	CAGL0D00143g
	CAGL0M00121g|**Awp4**
	CAGL0B00110g|**Awp8**[28]
	CAGL0B05093g|**Awp9**[28]
	CAGL0F00110g|**Awp10**[28]
	CAGL0M00110g|**Awp11**[28]
Cluster VI	CAGL0J02508g|**Awp1**
	CAGL0J11902g|**Awp3a**
	CAGL0J11935g|**Awp3b**
	CAGL0J01774g
	CAGL0J01727g
	CAGL0J01800g
	CAGL0J02552g
	CAGL0J02530g
Cluster VII	CAGL0G10219g|**Awp12**
	CAGL0C00825g
	CAGL0C01133g
	CAGL0C00803g
	CAGL0C00858g
	CAGL0C00968g
Unclassified	CAGL0G04125g
	CAGL0J05159g
	CAGL0L09911g
	CAGL0G05896g
	CAGL0C03575g
	CAGL0D06226g
	CAGL0K10164g
	CAGL0M03773g
	CAGL0E00187g
	CAGL0J11462g
	CAGL0L06424g
	CAGL0M11726g

8. 2. Appendix II: List of fragment screen datasets

Fragment Nr	Concentration [mM]	Dataset name	Soaking time	Estimated resolution [A]
J2	50	VR_138	23 h	1.9
J3	50	VR_139	23 h	1.8
J4	50	VR_140	23 h	2.5
J4	50	VR_141	23 h	2.3
J5	50	VR_142	23 h	2.0
J1	50	VR_143	ca 10 sec	2.6
J1	50	VR_144	ca 10 sec	2.5
J6	100	VR_145	ca 2 min	2.3
J9	50	VR_146	ca 2 min	2.5
J9	50	VR_147	ca 4 min	2.4
J7	50	VR_148	26 h	no diffraction
J8	100	VR_149	26 h	2.8
J8	100	VR_150	26 h	2.5
J10	50	VR_151	26 h	2.3
J13	50	VR_152	26 h	2.4
J51	100	VR_158	19 h	no diffraction
J61	100	VR_159	19 h	no diffraction
J63	50	VR_160	19 h	1.8
J37	100	VR_161	ca 50 min	3.8
J40	100	VR_162	1 h	no diffraction
J41	50	VR_163	1 h	3.0
J47	100	VR_164	30 min	2.1
J48	100	VR_165	20 min	no diffraction
J60	100	VR_166	10 min	no diffraction
J49	100	VR_167	ca 20 sec	2.2
J52	100	VR_168	3 min	2.1
J58	100	VR_169	5 min	2.7
J60	100	VR_170	7 min	2.4
J62	100	VR_171	6 min	2.3
J65	100	VR_172	3 h	2.5
J68	50	VR_173	3 h	2.0
J70	50	VR_174	3 h	2.4
J71	100	VR_175	3 h	no diffraction
J72	100	VR_176	3 h	2.0
J33	100	VR_177	10 min	3.5
J35	100	VR_178	ca 30 sec	2.0
J32	100	VR_179	10 min	no diffraction
J27	100	VR_180	100 min	2.2
J28	100	VR_181	100 min	2.0
J29	100	VR_182	100 min	2.4
J39	100	VR_183	90 min	2.0
J32	100	VR_184	90 min	no diffraction
J36	100	VR_185	10 min	3.2

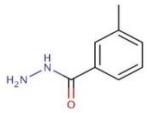
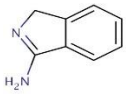
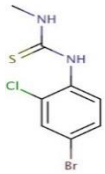
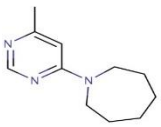
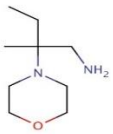
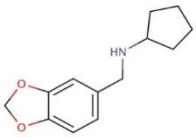
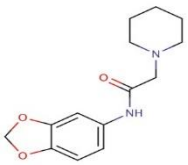
J39	100	VR_186	15 min	no diffraction
J46	100	VR_187	25 min	2.2
J43	100	VR_188	ca 30 min	2.6
J43	100	VR_189	ca 30 min	2.3
J42	100	VR_190	ca 30 min	2.6
J42	100	VR_191	ca 30 min	4.0
J34	50	VR_192	3 h	2.2
J73	100	VR_193	3 h	2.2
J74	50	VR_194	3 h	2.7
J75	50	VR_195	3 h	2.5
J76	50	VR_196	3 h	2.9
J77	50	VR_197	3 h	2.4
J78	50	VR_198	3 h	2.0
J80	50	VR_199	3 h	2.3
J80	50	VR_200	3 h	2.4
J81	50	VR_201	3 h	2.8
J16	100	VR_202	ca 5 min	2.0
J17	100	VR_203	ca 5 min	no diffraction
J18	100	VR_204	ca 30 min	3.5
J18	100	VR_205	ca 50 min	no diffraction
J26	100	VR_206	ca 1 min	2.5
J24	100	VR_207	1 h	no diffraction
J25	50	VR_208	1 h	no diffraction
J20	50	VR_209	3 h	2.1
J21	50	VR_210	3 h	3.2
J23	100	VR_211	3 h	no diffraction
J27	100	VR_212	3 h	no diffraction
J27	100	VR_213	3 h	no diffraction
J28	100	VR_214	3 h	no diffraction
J31	50	VR_215	3 h	low resolution
J21	50	VR_216	26 h	2.5
J20	50	VR_217	26 h	2.1
J64	50	VR_218	19.5 h	2.3
J63	50	VR_218	19,5 h	2.3
J34	50	VR_219	24 h	2.1
J41	50	VR_220	24 h	2.6
J67	50	VR_221	ca 15 sec	3.3
J66	100	VR_221	ca 15 sec	2.7
J69	100	VR_222	ca 30 sec	2.6
J83	100	VR_223	ca 30 sec	not processed
J84	100	VR_224	ca 10 sec	3.4
J94	100	VR_225	ca 2 min	2.0
J94	100	VR_226	ca 1 min	1.9
J66	100	VR_227	ca 2 h	2.5
J85	100	VR_228	ca 20 min	not processed
J91	100	VR_229	ca 15 min	no diffraction
J59	100	VR_230	3 h	2.3

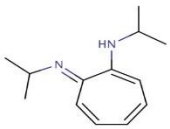
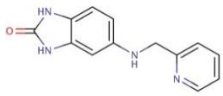
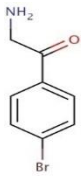
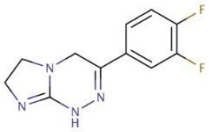
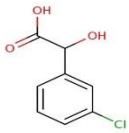
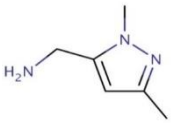
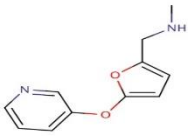
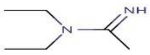
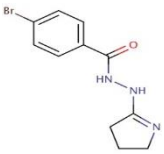
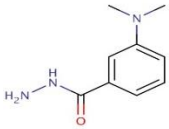
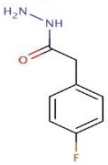
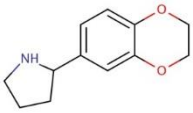
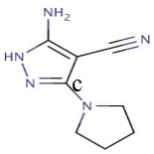
J61	100	VR_231	3 h	2.5
J63	50	VR_232	24 h	2.2
J64	50	VR_233	24 h	2.4
J86	50	VR_234	3 h	3.6
J93	100	VR_235	3 h	2.8
J68	50	VR_236	24 h	no diffraction
J72	100	VR_237	24 h	2.4
J73	100	VR_238	24 h	no diffraction
J74	50	VR_239	24 h	2.5
J75	50	VR_240	24 h	no diffraction
J76	50	VR_241	24 h	no diffraction
J76	50	VR_242	24 h	not processed
J77	50	VR_243	24 h	2.9
J78	50	VR_244	24 h	no diffraction
J80	50	VR_245	24 h	2.8
J80	50	VR_246	24 h	no diffraction
J84	100	VR_247	ca 1 min	2.0
J85	100	VR_248	15 min	2.9
J89	100	VR_249	12 min	3.1
J87	100	VR_250	20 min	2.8
J91	100	VR_251	14 min	3.0
J11	100*	VR_252	ca 30 sec	2.6
J11	100*	VR_253	ca 30 sec	2.3
J12	100*	VR_254	ca 30 sec	2.8
J12	100*	VR_255	1 min	2.9
J22	100*	VR_256	23 min	not processed
J14	100*	VR_257	ca 30 sec	not processed
J14	100*	VR_258	ca 30 sec	2.5
J14	100*	VR_259	ca 10 sec	2.5
J15	100*	VR_260	ca 30 sec	no diffraction
J15	100*	VR_261	ca 15 sec	2.3
J44	100*	VR_262	3 min	no diffraction
J44	100*	VR_263	ca 10 sec	2.2
J45	100*	VR_264	ca 1 min	2.3
J45	100*	VR_265	ca 15 sec	no diffraction
J79	100*	VR_266	ca 2 min	4.0
J79	100*	VR_267	ca 2 min	
J67	100*	VR_268	ca 6 min	2.5
J67	100*	VR_269	ca 6 min	no diffraction
J92	50*	VR_270	ca 30 sec	2.9
J90	50*	VR_271	ca 10 sec	2.2
J90	50*	VR_272	ca 10 sec	2.5
J94	100	VR_273	ca 1 min	3.1
J88	100	VR_274	90 min	2.6
J38	50	VR_275	3 h	2.2
J38	50	VR_276	4 h	2.9
J50	50*	VR_277	2 1/2 h	2.4

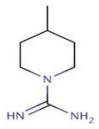
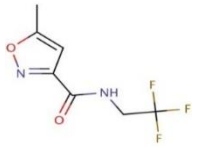
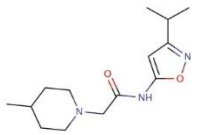
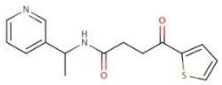
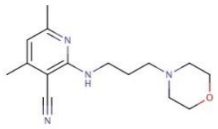
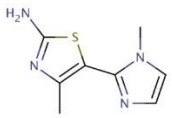
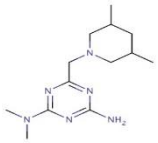
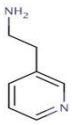
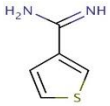
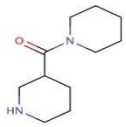
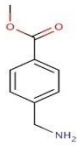
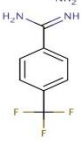
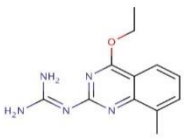
J50	50*	VR_278	2 1/2 h	Phaser error
J83	100	VR_285	30 sec	3.0
J7	50	VR_286	20 min	no diffraction
J22	100*	VR_287	10 min	3.7
J32	100	VR_288	4 min	2.5
J48	100	VR_289	1 min	no diffraction
J23	100	VR_290	5 min	3.5
J71	100	VR_291	1 h	no diffraction
J71	100	VR_292	1 h	no diffraction
J31	50	VR_293	1 h	no diffraction
J24	100	VR_294	20 min	2.4
J48	100	VR_295	1 min	2.7
J18	100	VR_296	5 min	no diffraction
J37	100	VR_299	15 min	2.7
J51	100	VR_300	1 h	bad diffraction

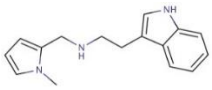
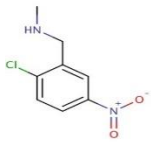
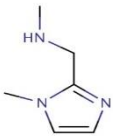
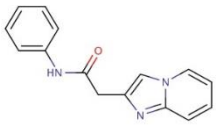
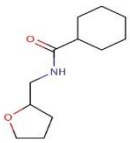
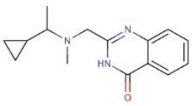
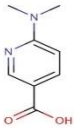
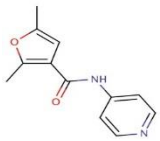
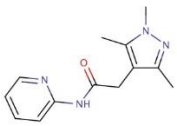
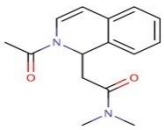
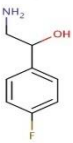
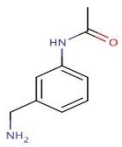
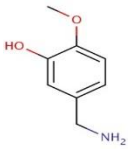
* Fragment powder remaining undissolved was centrifuged and the supernatant was used for soaking.

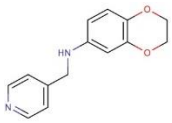
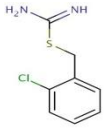
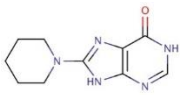
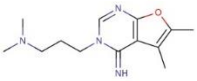
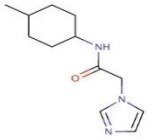
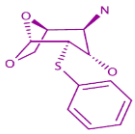
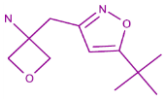
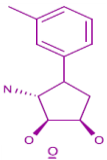
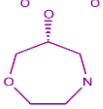
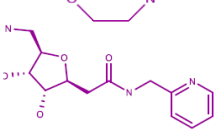
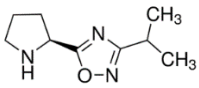
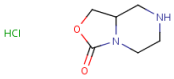
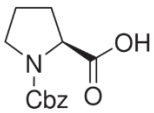
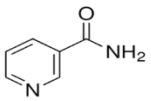
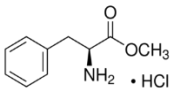
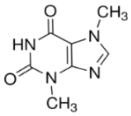
Fragment number refers to the Frag Xtal Screen from *Jena Bioscience*.

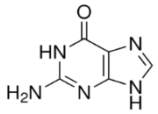
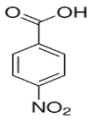
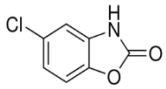
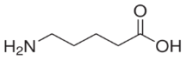
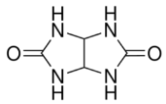
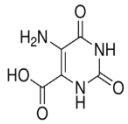
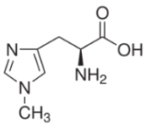
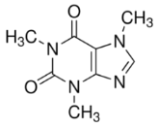
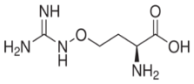
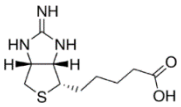
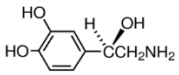
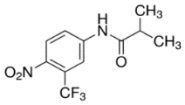
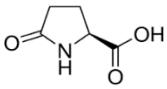
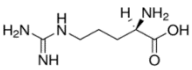
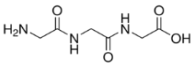
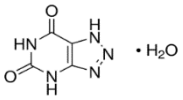
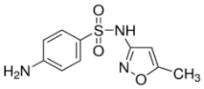
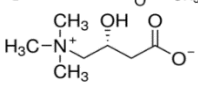
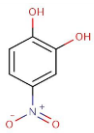
Fragment ID	Fragment	SMILES
1		CC1=CC(=CC=C1)C(=O)NN
2		C1C2=CC=CC=C2C(=N1)N
3		CNC(=S)NC1=C(C=C(C=C1)Br)Cl
4		CC1=NC=CC(=N1)N2CCCCCC2
5		CCC(C)(CN)N1CCOCC1
6		C1CCC(C1)NCC2=CC3=C(C=C2)OCO3
7		O=C(CN1CCCCC1)Nc1ccc2OCOc2c1

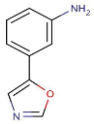
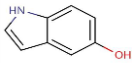
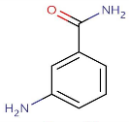
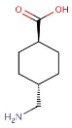
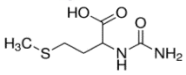
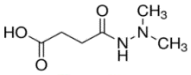
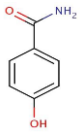
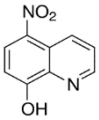
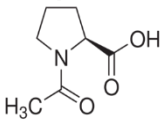
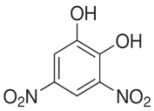
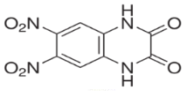
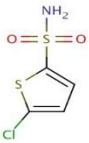
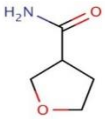
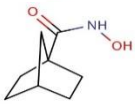
8		CC(C)Nc1ccccc\c1=N/C(C)C
9		O=c1[nH]c2ccc(NCc3ccccn3)cc2[nH]1
10		NCC(=O)c1ccc(Br)cc1
11		Fc1ccc(cc1F)C1=NNC2=NCCN2C1 |t:9,12|
12		OC(C(O)=O)c1cccc(Cl)c1
13		Cc1cc(CN)n(C)n1
14		CNCc1ccc(Oc2cccnc2)o1
15		CCN(CC)C(=N)C
16		Brc1ccc(cc1)C(=O)NNC1=NCCC1 |t:12|
17		CN(C)c1cccc(c1)C(=O)NN
18		NNC(=O)Cc1ccc(Br)cc1
19		C1CNC(C1)c1ccc2OCCOc2c1
20		Nc1[nH]nc(N2CCCC2)c1C#N

21		NC(=N)N1CCCCC1
22		CC1ON=C(C(NCC(F)(F)F)=O)C=1
23		CC(C1=NOC(NC(CN2CCC(C)CC2)=O)=C1)C
24		CC(NC(=O)CCC(=O)c1cccs1)c1cccnc1
25		Cc1cc(C)c(C#N)c(NCCCN2CCOCC2)n1
26		Cc1nc(N)sc1-c1nccn1C
27		CC1CC(C)CN(Cc2nc(N)nc(n2)N(C)C)C1
28		NCCc1cccnc1
29		NC(=N)c1ccsc1
30		NCC1OC(C(F)(F)F)CC1
31		O=C(C1CCCNC1)N1CCCCC1
32		COC(=O)c1ccc(CN)cc1
33		NC(=N)c1ccc(cc1)C(F)(F)F
34		CCOc1nc(NC(N)=N)nc2c(C)cccc12

35		Cn1cccc1CNCCc1c[nH]c2ccccc12
36		CNCC1=CC(=CC(=C1)Cl)[N+](=O)[O-]
37		CNCc1nccn1C
38		O=C(Cc1cn2ccccc2n1)Nc1ccccc1
39		O=C(NCC1CCCO1)C1CCCCC1
40		CC(N(CC1NC(=O)C2=C(C=CC=C2)N=1)C)C1CC1
41		CN(C)c1ccc(cn1)C(O)=O
42		Cc1cc(C(=O)Nc2ccncc2)c(C)o1
43		Cc1nn(C)c(C)c1CC(=O)Nc1ccccn1
44		CN(C(CC1C2=C(C=CC=C2)C=CN1C(C)=O)=O)C
45		NCC(O)c1ccc(F)cc1
46		CC(=O)Nc1cccc(CN)c1
47		COc1ccc(CN)cc1O

48		C(Nc1ccc2OCCOc2c1)c1ccncc1
49		NC(=N)SCc1ccccc1Cl
50		O=c1[nH]cnc2[nH]c(nc12)N1CCCCC1
51		CN(C)CCCn1cnc2oc(C)c(C)c2c1=N
52		CC1CCC(CC1)NC(=O)Cn1ccnc1
53		
54		CC(C)(C)c1cc(CC2(N)COC2)no1
55		Cc1cccc(c1)C1C[C@@H](O)[C@@H](O)[C@@H]1N
56		O[C@@H]1CNCCOC1
57		
58		CC(C)c1noc(n1)C1CCCN1
59		O=C1OCC2CNCCN12
60		
61		NC(=O)c1cccnc1
62		COC(=O)C(CC1=CC=CC=C1)N.Cl
63		Cn1cnc2n(C)c(=O)[nH]c(=O)c12

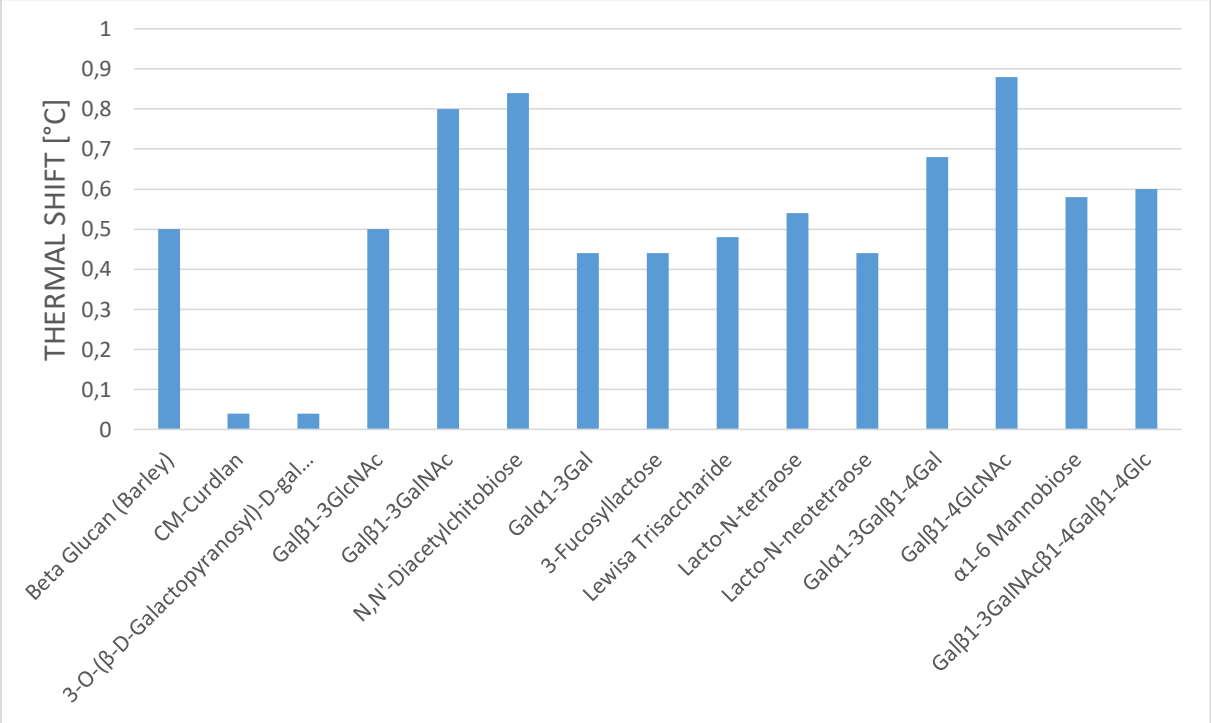
64		Nc1nc(O)c2[nH]cnc2n1
65		C1=CC(=CC=C1C(=O)O)[N+](=O)[O-]
66		Oc1nc2cc(Cl)ccc2o1
67		NCCCCC(O)=O
68		O=C1NC2NC(=O)NC2N1
69		C1(C(O)=O)NC(=O)NC(=O)C1N
70		Cn1cnc(C[C@H](N)C(O)=O)c1
71		C12N=CN(C)C=1C(N(C)C(=O)N2C)=O
72		[C@H](N)(CCONC(=N)N)C(=O)O
73		C1C2C(C(S1)CCCCC(=O)O)N=C(N2)N
74		C(N)C(C1=CC=C(C)C(C)=C1)O
75		
76		OC(=O)C1CCC(=O)N1
77		NC(CCCNC(N)=N)C(O)=O
78		NCC(=O)NCC(=O)NCC(O)=O
79		Oc1nc(O)c2nn[nH]c2n1
80		Cc1cc(NS(=O)(=O)c2ccc(N)cc2)no1
81		
82		Oc1ccc(cc1O)[N+]([O-])=O

83		Nc1cccc(c1)-c1cnco1
84		Oc1ccc2[nH]ccc2c1
85		NC(=O)c1cccc(N)c1
86		C1CC(CCC1CN)C(=O)O
87		CSCCC(NC(N)=O)C(O)=O
88		CN(C)NC(=O)CCC(O)=O
89		NC(=O)c1ccc(O)cc1
90		Oc1ccc([N+]([O-])=O)c2cccnc12
91		CC(=O)N1CCC[C@H]1C(O)=O |r|
92		Oc1cc(cc(c1O)[N+]([O-])=O)[N+]([O-])=O
93		Oc1nc2cc(c(cc2nc1O)[N+]([O-])=O)[N+]([O-])=O
94		NS(=O)(=O)c1ccc(Cl)s1
95		NC(=O)C1CCOC1
96		ONC(=O)C12CCC(CC1)C2

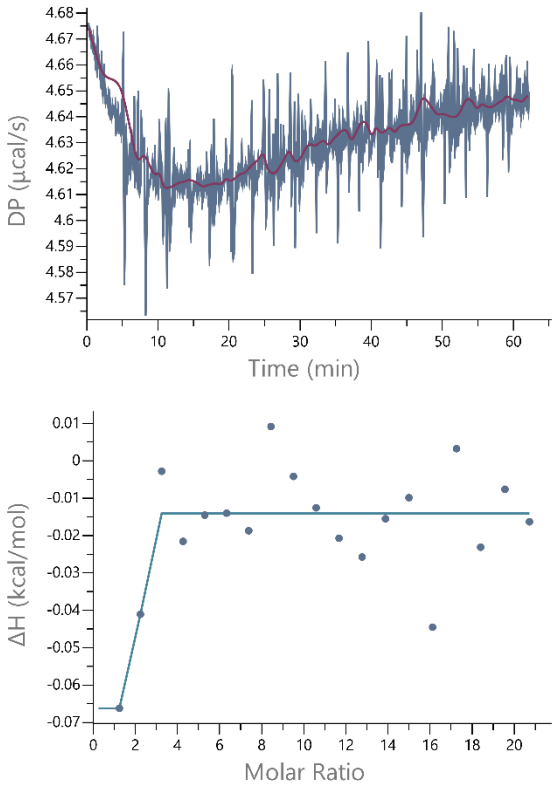
8. 3. Appendix III: *DIMPLE* script

```
#!/bin/bash -f
#
# usage: SLS_to_pandda.sh pdbin rfree-in out_dir
#
#
if [[ "$1" != "" && -f $1 ]]; then
   pdb_ref=$1
   echo "### Assign PDB reference structure to "$pdb_ref
else
   echo "Please give reference pdb structure !"
   echo "usage: SLS_to_pandda.sh pdbin rfree-in out_dir" && exit
fi
#
if [[ "$2" != "" && -f $2 ]]; then
   rfree_ref=$2
   echo "### Assign FreeR_flag reference mtz file to "$rfree_ref
else
   echo "No R-free flag reference mtz file given ! This file has to have a colum FreeR_flag."
   echo "usage: SLS_to_pandda.sh pdbin rfree-in out_dir" && exit
fi
#
if [[ "$3" != "" ]]; then
   outdir=$1
else
   outdir=aimless_dirs
fi
#
echo "### Set output directory to "$outdir
#
# The next line finds all successfully generated XDS_ASCII.HKL
#      in the gopy subdirs as generated by SLS pipeline
#
FILES=`find . -type f -wholename "*/*/gopy/XDS_ASCII.HKL"`
#FILES=`find . -type f -wholename "*/*/manual_XDS/XDS_ASCII.HKL"`
#
[ -e $outdir ]         && /bin/rm -rf $outdir
#
##
mkdir $outdir
#
#
for xdsfile in $FILES; do
   xdspath=`dirname $xdsfile`
   dataset_prefix=`echo $xdspath | sed 's/\.\/\([A-Z,a-z,0-9,\_,\-]*\).*/\1/'`
   outputs_prefix=$outdir/$dataset_prefix
   #
   echo "Dataset "$dataset_prefix" found: data under "$xdspath
   mkdir ${outputs_prefix}
   srun pointless -copy    XDSIN $xdsfile HKLOUT ${outputs_prefix}/XDS_ASCII.mtz\
                                 | tee ${outputs_prefix}/${dataset_prefix}.pointless.log \
            && aimless --no-input HKLIN ${outputs_prefix}/XDS_ASCII.mtz HKLOUT
${outputs_prefix}/${dataset_prefix}.aimless.mtz \
                           | tee ${outputs_prefix}/${dataset_prefix}.aimless.log \
            && dimple  --hklout ${dataset_prefix}.dimple.mtz --xyzout ${dataset_prefix}.dimple.pdb -R $rfree_ref
${outputs_prefix}/${dataset_prefix}.aimless.mtz $pdb_ref ${outputs_prefix} \
                           | tee ${outputs_prefix}/${dataset_prefix}.dimple.log >&
${outputs_prefix}/${dataset_prefix}.SLS_to_pandda.log &
done
#
Exit
```

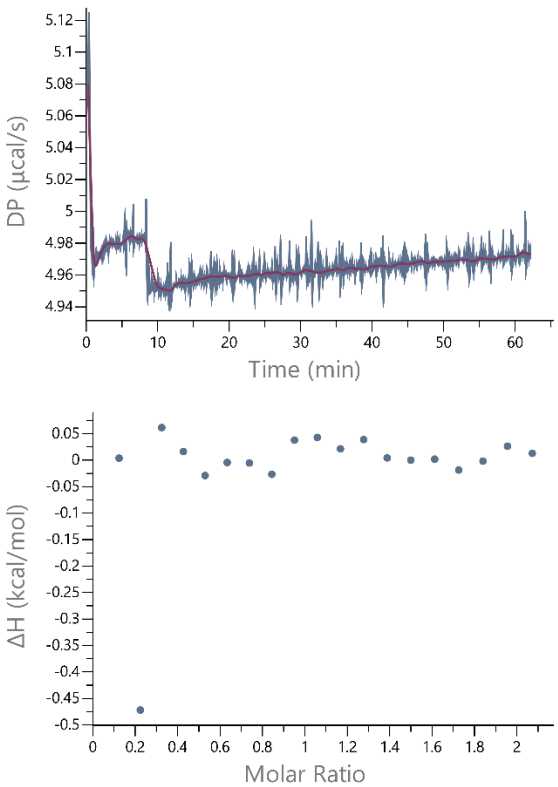
8. 4. Appendix IV: TSA – Awp1A



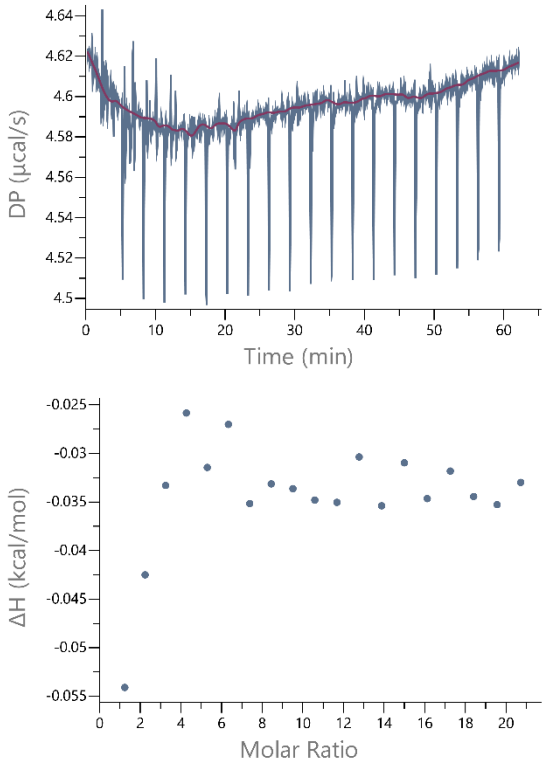
8. 5. Appendix V: ITC measurements of Awp3A and α-1,6-mannobiose



100 µM Awp3A
1 mM α-1,6-mannobiose



100 µM Awp3A
10 mM α-1,6-mannobiose



50 µM Awp3A
10 mM α-1,6-mannobiose

8. 6. Appendix VI: Glycan Array results

8. 6. 1. Awp1A (5 µg/mL) – Anti-His-488 (5 µg/mL)

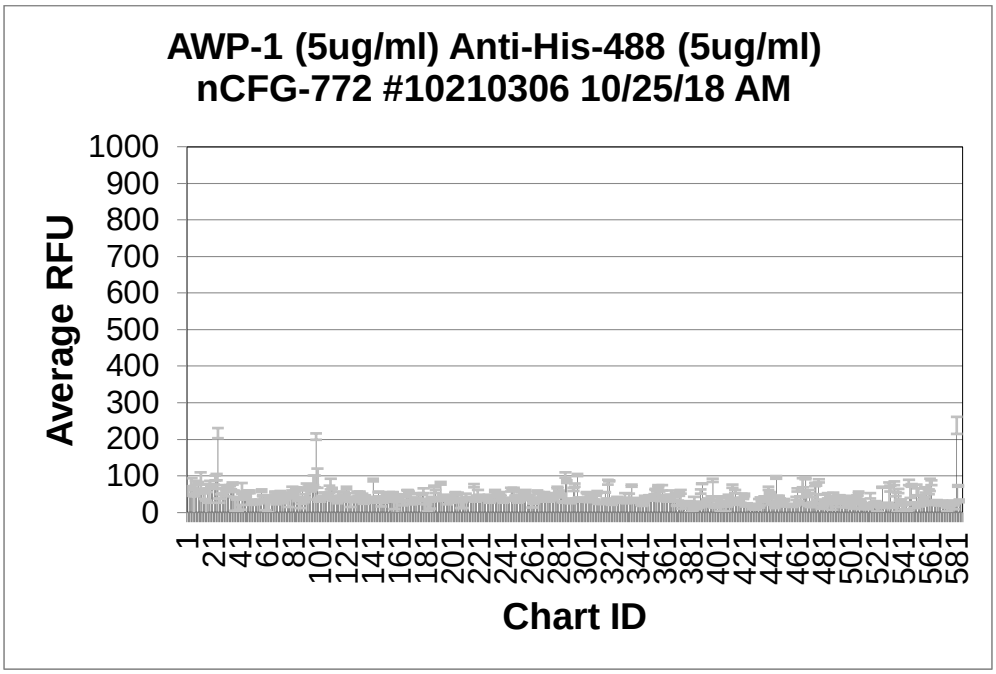


Chart ID	Sample (conc.) Secondary (conc.) Barcode# Slide # Request # Date Initials	Average RFU	StDev	%CV
1	Gala-Sp8	58	6	11
2	Glca-Sp8	48	2	5
3	Mana-Sp8	66	6	9
4	GalNAca-Sp8	81	13	16
5	GalNAca-Sp15	65	2	3
6	Fuca-Sp8	16	28	175
7	Fuca-Sp9	80	6	8
8	Rhaa-Sp8	58	5	9
9	Neu5Aca-Sp8	81	3	3
10	Neu5Aca-Sp11	54	2	4
11	Neu5Acb-Sp8	83	26	32
12	Galb-Sp8	55	3	5
13	Glcb-Sp8	59	6	10
14	Manb-Sp8	51	14	27
15	GalNAcb-Sp8	11	21	194
16	GlcNAcb-Sp0	66	10	15
17	GlcNAcb-Sp8	46	20	43
18	GlcN(Gc)b-Sp8	82	5	6
19	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAca-Sp8	70	2	2
20	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAc-Sp14	71	5	7
21	GlcNAcb1-6(GlcNAcb1-4)(GlcNAcb1-3)GlcNAc-Sp8	70	6	8
22	6S(3S)Galb1-4(6S)GlcNAcb-Sp0	94	6	6
23	6S(3S)Galb1-4GlcNAcb-Sp0	96	9	9
24	(3S)Galb1-4(Fuca1-3)(6S)Glc-Sp0	217	14	6
25	(3S)Galb1-4Glcb-Sp8	32	6	20
26	(3S)Galb1-4(6S)Glcb-Sp0	43	7	16
27	(3S)Galb1-4(6S)Glcb-Sp8	58	6	11
28	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp8	59	7	12
29	(3S)Galb1-3GalNAca-Sp8	70	4	6

30	(3S)Galb1-3GlcNAcb-Sp0	52	9	17
31	(3S)Galb1-3GlcNAcb-Sp8	66	5	7
32	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp0	63	3	4
33	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp8	73	2	2
34	(3S)Galb1-4(6S)GlcNAcb-Sp0	62	1	2
35	(3S)Galb1-4(6S)GlcNAcb-Sp8	79	3	4
36	(3S)Galb1-4GlcNAcb-Sp0	58	2	3
37	(3S)Galb1-4GlcNAcb-Sp8	22	18	81
38	(3S)Galb-Sp8	38	4	10
39	(6S)(4S)Galb1-4GlcNAcb-Sp0	19	18	98
40	(4S)Galb1-4GlcNAcb-Sp8	47	11	24
41	(6P)Mana-Sp8	14	6	45
42	(6S)Galb1-4Glcb-Sp0	66	15	23
43	(6S)Galb1-4Glcb-Sp8	38	2	5
44	(6S)Galb1-4GlcNAcb-Sp8	38	1	3
45	(6S)Galb1-4(6S)Glcb-Sp8	44	3	7
46	Neu5Aca2-3(6S)Galb1-4GlcNAcb-Sp8	56	2	3
47	(6S)GlcNAcb-Sp8	56	4	6
48	Neu5,9Ac_2a-Sp8	57	4	7
49	Neu5,9Ac2a2-6Galb1-4GlcNAcb-Sp8	28	7	25
50	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	27	1	3
51	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	28	2	6
52	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	31	4	13
53	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	26	4	16
54	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	32	1	3
55	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	27	3	10
56	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man-a1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	33	2	6
57	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	60	4	6
58	Fuca1-2Galb1-3GalNAcb1-3Gala-Sp9	46	2	4
59	Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp9	35	2	7
60	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb-Sp8	20	13	67
61	Fuca1-2Galb1-3GalNAca-Sp8	35	1	2
62	Fuca1-2Galb1-3GalNAca-Sp14	16	15	98
63	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	42	4	9
64	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp9	31	1	3
65	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp8	36	7	19
66	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10	31	3	10
67	Fuca1-2Galb1-3GlcNAcb-Sp0	56	2	3
68	Fuca1-2Galb1-3GlcNAcb-Sp8	32	10	31
69	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	53	3	6
70	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	50	3	7
71	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp0	53	5	10
72	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp8	39	1	1
73	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	28	2	9
74	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	36	3	8
75	Fuca1-2Galb1-4GlcNAcb-Sp0	41	3	8
76	Fuca1-2Galb1-4GlcNAcb-Sp8	49	4	8
77	Fuca1-2Galb1-4Glcb-Sp0	28	13	47
78	Fuca1-2Galb-Sp8	53	1	2
79	Fuca1-3GlcNAcb-Sp8	44	6	13
80	Fuca1-4GlcNAcb-Sp8	65	5	8
81	Fucb1-3GlcNAcb-Sp8	49	3	6
82	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	57	1	2
83	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	66	3	4
84	(3S)Galb1-4(Fuca1-3)Glcb-Sp0	37	5	15
85	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	37	3	8

86	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	24	13	53
87	GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	27	4	16
88	GlcNAcb1-3Galb1-3GalNAca-Sp8	63	3	4
89	GalNAca1-3(Fuca1-2)Galb-Sp8	32	5	14
90	GalNAca1-3(Fuca1-2)Galb-Sp18	46	4	9
91	GalNAca1-3GalNAcb-Sp8	73	6	8
92	GalNAca1-3Galb-Sp8	53	18	33
93	GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	72	4	5
94	GalNAcb1-3GalNAca-Sp8	63	5	9
95	GalNAcb1-3(Fuca1-2)Galb-Sp8	64	3	4
96	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	88	13	15
97	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0	85	16	18
98	GalNAcb1-4GlcNAcb-Sp0	207	9	4
99	GalNAcb1-4GlcNAcb-Sp8	93	27	29
100	Gala1-2Galb-Sp8	34	5	13
101	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	32	3	9
102	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp8	40	5	12
103	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	41	3	8
104	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp8	54	1	2
105	Gala1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	40	2	4
106	Gala1-3(Fuca1-2)Galb1-4Glcb-Sp0	42	3	7
107	Gala1-3(Fuca1-2)Galb-Sp8	44	4	9
108	Gala1-3(Fuca1-2)Galb-Sp18	63	9	15
109	Gala1-4(Gala1-3)Galb1-4GlcNAcb-Sp8	78	15	19
110	Gala1-3GalNAca-Sp8	65	2	3
111	Gala1-3GalNAca-Sp16	37	4	12
112	Gala1-3GalNAcb-Sp8	33	1	4
113	Gala1-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	32	2	5
114	Gala1-3Galb1-3GlcNAcb-Sp0	29	4	14
115	Gala1-3Galb1-4GlcNAcb-Sp8	51	5	10
116	Gala1-3Galb1-4Glcb-Sp0	37	3	9
117	Gala1-3Galb1-4Glc-Sp10	40	6	14
118	Gala1-3Galb-Sp8	44	3	6
119	Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	56	2	4
120	Gala1-4Galb1-4GlcNAcb-Sp0	36	3	9
121	Gala1-4Galb1-4GlcNAcb-Sp8	66	4	6
122	Gala1-4Galb1-4Glcb-Sp0	41	3	7
123	Gala1-4GlcNAcb-Sp8	41	12	29
124	Gala1-6Glcb-Sp8	24	8	35
125	Galb1-2Galb-Sp8	42	1	3
126	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	38	2	4
127	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	32	5	15
128	Galb1-3(Fuca1-4)GlcNAc-Sp0	35	4	11
129	Galb1-3(Fuca1-4)GlcNAc-Sp8	48	9	18
130	Fuca1-4(Galb1-3)GlcNAcb-Sp8	41	5	11
131	Galb1-4GlcNAcb1-6GalNAca-Sp8	55	2	3
132	Galb1-4GlcNAcb1-6GalNAc-Sp14	45	3	7
133	GlcNAcb1-6(Galb1-3)GalNAca-Sp8	49	3	6
134	GlcNAcb1-6(Galb1-3)GalNAca-Sp14	31	8	25
135	Neu5Aca2-6(Galb1-3)GalNAca-Sp8	43	4	10
136	Neu5Aca2-6(Galb1-3)GalNAca-Sp14	29	3	9
137	Neu5Acb2-6(Galb1-3)GalNAca-Sp8	39	3	6
138	Neu5Aca2-6(Galb1-3)GlcNAcb1-4Galb1-4Glcb-Sp10	32	3	9
139	Galb1-3GalNAca-Sp8	34	8	24
140	Galb1-3GalNAca-Sp14	31	1	3
141	Galb1-3GalNAca-Sp16	88	4	4
142	Galb1-3GalNAcb-Sp8	37	2	7
143	Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	34	2	6
144	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	39	2	5
145	Galb1-3GalNAcb1-4Galb1-4Glcb-Sp8	56	1	2
146	Galb1-3Galb-Sp8	40	3	9

147	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	21	5	21
148	Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10	27	1	5
149	Galb1-3GlcNAcb-Sp0	38	3	8
150	Galb1-3GlcNAcb-Sp8	33	2	5
151	Galb1-4(Fuca1-3)GlcNAcb-Sp0	47	5	10
152	Galb1-4(Fuca1-3)GlcNAcb-Sp8	49	3	5
153	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	54	2	4
154	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	30	1	3
155	Galb1-4(6S)Glcb-Sp0	45	2	5
156	Galb1-4(6S)Glcb-Sp8	55	1	2
157	Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	21	12	57
158	Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	44	4	8
159	Galb1-4GlcNAcb1-3GalNAca-Sp8	19	20	107
160	Galb1-4GlcNAcb1-3GalNAc-Sp14	31	3	10
161	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	47	3	5
162	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	31	3	8
163	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	32	14	42
164	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	45	2	3
165	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp8	34	1	3
166	Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp8	45	2	5
167	Galb1-4GlcNAcb1-6(Galb1-3)GalNAc-Sp14	61	1	1
168	Galb1-4GlcNAcb-Sp0	55	3	5
169	Galb1-4GlcNAcb-Sp8	28	5	16
170	Galb1-4GlcNAcb-Sp23	27	4	16
171	Galb1-4Glcb-Sp0	32	3	8
172	Galb1-4Glcb-Sp8	26	2	9
173	GlcNAca1-3Galb1-4GlcNAcb-Sp8	39	3	7
174	GlcNAca1-6Galb1-4GlcNAcb-Sp8	36	1	4
175	GlcNAcb1-2Galb1-3GalNAca-Sp8	52	2	3
176	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp8	35	2	6
177	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp14	32	1	4
178	GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAcb-Sp8	49	1	2
179	GlcNAcb1-3GalNAca-Sp8	59	6	11
180	GlcNAcb1-3GalNAca-Sp14	14	17	124
181	GlcNAcb1-3Galb-Sp8	28	7	24
182	GlcNAcb1-3Galb1-4GlcNAcb-Sp0	17	15	86
183	GlcNAcb1-3Galb1-4GlcNAcb-Sp8	33	1	4
184	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	20	10	47
185	GlcNAcb1-3Galb1-4Glcb-Sp0	40	3	8
186	GlcNAcb1-4-MDPLys	36	2	5
187	GlcNAcb1-6(GlcNAcb1-4)GalNAca-Sp8	71	2	3
188	GlcNAcb1-4Galb1-4GlcNAcb-Sp8	58	4	8
189	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	32	1	4
190	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	33	1	2
191	GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8	38	1	2
192	GlcNAcb1-6GalNAca-Sp8	80	3	4
193	GlcNAcb1-6GalNAca-Sp14	36	1	1
194	GlcNAcb1-6Galb1-4GlcNAcb-Sp8	47	2	3
195	Glca1-4Glcb-Sp8	33	1	4
196	Glca1-4Glca-Sp8	42	5	13
197	Glca1-6Glca1-6Glcb-Sp8	29	7	23
198	Glcb1-4Glcb-Sp8	34	2	6
199	Glcb1-6Glcb-Sp8	29	1	3
200	G-ol-Sp8	31	6	18
201	GlcAa-Sp8	40	2	5
202	GlcAb-Sp8	40	6	16
203	GlcAb1-3Galb-Sp8	55	2	3
204	GlcAb1-6Galb-Sp8	50	1	1
205	KDNa2-3Galb1-3GlcNAcb-Sp0	53	2	3
206	KDNa2-3Galb1-4GlcNAcb-Sp0	36	2	5

207	Mana1-2Mana1-2Mana1-3Mana-Sp9	22	10	46
208	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana-Sp9	30	2	6
209	Mana1-2Mana1-3Mana-Sp9	24	11	48
210	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	37	2	5
211	Mana1-6(Mana1-3)Mana-Sp9	47	4	8
212	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9	36	1	3
213	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	35	2	6
214	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	40	1	1
215	Manb1-4GlcNAcb-Sp0	40	1	2
216	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	34	1	2
217	(3S)Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	73	5	7
218	Fuca1-2(6S)Galb1-4GlcNAcb-Sp0	36	6	17
219	Fuca1-2Galb1-4(6S)GlcNAcb-Sp8	40	5	13
220	Fuca1-2(6S)Galb1-4(6S)Glcb-Sp0	54	8	15
221	Neu5Aca2-3Galb1-3GalNAca-Sp8	46	2	4
222	Neu5Aca2-3Galb1-3GalNAca-Sp14	38	2	5
223	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	37	3	7
224	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	39	2	6
225	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0	35	1	4
226	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	41	1	3
227	Neu5Aca2-8Neu5Aca2-8Neu5Aca-Sp8	34	1	2
228	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	40	8	19
229	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp8	29	2	8
230	GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	33	2	6
231	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	34	2	6
232	Neu5Aca2-6(Neu5Aca2-3)GalNAca-Sp8	46	2	4
233	Neu5Aca2-3GalNAca-Sp8	59	3	4
234	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0	42	1	1
235	Neu5Aca2-3Galb1-3(6S)GlcNAc-Sp8	50	3	5
236	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp8	55	1	1
237	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	46	2	5
238	Neu5Aca2-3Galb1-4(Neu5Aca2-3Galb1-3)GlcNAcb-Sp8	37	1	3
239	Neu5Aca2-3Galb1-3(6S)GalNAca-Sp8	32	6	20
240	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp8	31	2	7
241	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	38	1	4
242	Neu5Aca2-3Galb-Sp8	36	2	5
243	Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	38	1	2
244	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	35	1	1
245	Fuca1-2(6S)Galb1-4Glcb-Sp0	63	5	8
246	Neu5Aca2-3Galb1-3GlcNAcb-Sp0	64	2	3
247	Neu5Aca2-3Galb1-4(6S)GlcNAcb-Sp8	65	2	3
248	Neu5Aca2-3Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	39	7	18
249	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	41	8	20
250	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	28	2	6
251	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	35	5	14
252	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8	34	1	2
253	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8	57	2	3
254	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	35	1	4
255	Neu5Aca2-3Galb1-4GlcNAcb-Sp0	51	2	4
256	Neu5Aca2-3Galb1-4GlcNAcb-Sp8	59	3	5
257	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	42	1	3
258	Fuca1-2Galb1-4(6S)Glcb-Sp0	46	1	3
259	Neu5Aca2-3Galb1-4Glcb-Sp0	42	6	13
260	Neu5Aca2-3Galb1-4Glcb-Sp8	38	2	6
261	Neu5Aca2-6GalNAca-Sp8	24	11	47
262	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0	26	2	8
263	Neu5Aca2-6Galb1-4(6S)GlcNAcb-Sp8	36	2	7
264	Neu5Aca2-6Galb1-4GlcNAcb-Sp0	32	3	9
265	Neu5Aca2-6Galb1-4GlcNAcb-Sp8	58	2	4

266	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	54	2	3
267	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	37	0	0
268	Neu5Aca2-6Galb1-4Glcb-Sp0	52	2	4
269	Neu5Aca2-6Galb1-4Glcb-Sp8	44	1	2
270	Neu5Aca2-6Galb-Sp8	54	1	3
271	Neu5Aca2-8Neu5Aca-Sp8	43	2	5
272	Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0	32	3	8
273	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	37	7	19
274	Neu5Acb2-6GalNAca-Sp8	32	1	4
275	Neu5Acb2-6Galb1-4GlcNAcb-Sp8	41	6	14
276	Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	39	1	1
277	Neu5Gca2-3Galb1-3GlcNAcb-Sp0	38	3	8
278	Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	49	3	7
279	Neu5Gca2-3Galb1-4GlcNAcb-Sp0	44	1	2
280	Neu5Gca2-3Galb1-4Glcb-Sp0	68	2	3
281	Neu5Gca2-6GalNAca-Sp0	58	3	4
282	Neu5Gca2-6Galb1-4GlcNAcb-Sp0	44	3	6
283	Neu5Gca-Sp8	48	3	7
284	Neu5Aca2-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	29	2	7
285	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	27	1	2
286	Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp0	102	7	7
287	Galb1-4(Fuca1-3)(6S)Glcb-Sp0	84	4	5
288	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	36	2	6
289	Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	32	4	12
290	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	27	2	6
291	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	31	1	3
292	4S(3S)Galb1-4GlcNAcb-Sp0	63	4	7
293	(6S)Galb1-4(6S)GlcNAcb-Sp0	75	4	5
294	(6P)Glcb-Sp10	36	3	9
295	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3)GalNAca-Sp14	102	3	3
296	Galb1-3Galb1-4GlcNAcb-Sp8	39	2	6
297	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	30	1	2
298	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	36	1	1
299	GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	32	1	4
300	Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0	34	3	7
301	Galb1-4GlcNAcb1-6Galb1-4GlcNAcb-Sp0	38	1	2
302	GalNAcb1-3Galb-Sp8	54	2	3
303	GlcAb1-3GlcNAcb-Sp8	51	2	4
304	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	29	1	3
305	GlcNAcb1-3Man-Sp10	42	2	4
306	GlcNAcb1-4GlcNAcb-Sp10	40	2	5
307	GlcNAcb1-4GlcNAcb-Sp12	34	2	4
308	MurNAcb1-4GlcNAcb-Sp10	34	2	5
309	Mana1-6Manb-Sp10	49	4	7
310	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb-Sp10	56	3	6
311	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	24	1	3
312	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	25	1	6
313	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	25	1	5
314	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	26	2	9
315	Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	25	2	6
316	Neu5Aca2-8Neu5Acb-Sp17	55	1	1
317	Neu5Aca2-8Neu5Aca2-8Neu5Acb-Sp8	35	3	10
318	Neu5Gcb2-6Galb1-4GlcNAc-Sp8	82	7	8
319	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	87	1	1
320	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	24	0	0

321	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	22	1	4
322	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	29	1	5
323	Neu5,9Ac2a2-3Galb1-3GlcNAcb-Sp0	32	1	2
324	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	33	1	2
325	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	39	4	9
326	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	30	1	3
327	Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	35	1	3
328	GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	27	1	3
329	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	27	1	5
330	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	30	2	7
331	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAc-Sp14	47	7	14
332	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	28	3	11
333	GlcNAca1-4Galb1-4GlcNAcb-Sp0	31	5	16
334	GlcNAca1-4Galb1-3GlcNAcb-Sp0	43	9	20
335	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	35	1	1
336	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	73	3	3
337	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	38	2	4
338	GlcNAca1-4Galb1-3GalNAc-Sp14	31	5	15
339	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	30	1	4
340	Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	29	2	8
341	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	27	0	0
342	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	26	1	2
343	Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	25	4	14
344	Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	20	1	7
345	Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	27	1	5
346	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	42	4	9
347	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	36	2	5
348	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	36	2	7
349	(6S)GlcNAcb1-3Galb1-4GlcNAcb-Sp0	45	3	6
350	KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0	46	1	2
351	KDNa2-6Galb1-4GlcNAc-Sp0	37	1	3
352	KDNa2-3Galb1-4Glc-Sp0	36	3	8
353	KDNa2-3Galb1-3GalNAca-Sp14	45	5	10
354	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	63	2	3
355	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	59	1	2
356	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAb-Sp20	72	3	4
357	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	53	5	9
358	Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	32	1	3
359	Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-6(Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	68	7	11
360	Neu5Aca2-6GlcNAcb1-4GlcNAc-Sp21	42	3	7
361	Neu5Aca2-6GlcNAcb1-4GlcNAcb1-4GlcNAc-Sp21	42	4	9
362	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	36	1	3
363	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	31	1	4
364	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	43	1	3
365	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	45	4	8
366	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	54	6	11
367	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	32	1	4

368	Galα1-3(Fucα1-2)Galβ1-3GlcNAcβ1-2Manα1-6(Galα1-3(Fucα1-2)Galβ1-3GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp20	40	4	9
369	Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	48	6	12
370	Neu5Aca2-3Galb1-4GlcNAcb1-3GalNAc-Sp14	19	1	6
371	Neu5Aca2-6Galb1-4GlcNAcb1-3GalNAc-Sp14	31	1	4
372	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	51	6	11
373	GalNAcb1-4GlcNAcb1-2Mana1-6(GalNAcb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	56	5	9
374	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0	16	6	36
375	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	21	1	5
376	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp0	22	3	11
377	Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	22	3	15
378	Galb1-4GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	23	7	32
379	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	23	1	4
380	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	10	10	93
381	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	22	1	4
382	GlcNAcb1-2Mana1-6(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	13	8	65
383	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	36	6	18
384	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	20	9	43
385	Galb1-3GlcNAcb1-3GalNAca-Sp14	18	9	50
386	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb1-3GalNAca-Sp14	25	1	4
387	GalNAca1-3(Fuca1-2)Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	17	3	19
388	Gala1-3Galb1-3GlcNAcb1-2Mana1-6(Gala1-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	57	6	10
389	Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	79	1	1
390	GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	19	3	14
391	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	20	1	4
392	Neu5Aca2-3Galb1-3GlcNAcb1-3GalNAca-Sp14	25	4	18
393	Fuca1-2Galb1-4GlcNAcb1-3GalNAca-Sp14	34	1	2
394	Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	36	3	8
395	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	20	3	14
396	Gala1-4Galb1-3GlcNAcb1-2Mana1-6(Gala1-4Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	40	5	12
397	Gala1-4Galb1-4GlcNAcb1-2Mana1-6(Gala1-4Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	88	4	4
398	Gala1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	18	4	22
399	Galb1-3GlcNAcb1-6Galb1-4GlcNAcb-Sp0	32	2	6
400	Galb1-3GlcNAca1-6Galb1-4GlcNAcb-Sp0	23	10	41
401	GalNAcb1-3Gala1-6Galb1-4Glcb-Sp8	28	14	48
402	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)Glcb-Sp21	21	2	10
403	Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	15	10	69
404	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	42	2	5
405	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	28	5	17
406	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	24	4	15
407	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	10	7	63
408	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	23	7	32
409	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	45	3	7
410	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	25	4	15
411	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	36	2	5
412	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	72	4	6
413	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	39	2	4
414	GlcNAcb1-2(GlcNAcb1-6)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	57	4	7
415	Fuca1-2Galb1-3GlcNAcb1-3GalNAc-Sp14	22	1	4
416	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	25	3	13
417	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	28	5	18

418	Gala1-3Galb1-3GlcNAcb1-3GalNAc-Sp14	25	2	9
419	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	45	6	14
420	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	39	2	4
421	Galb1-3GlcNAcb1-6(Galb1-3GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	48	4	7
422	Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	19	2	10
423	Fuca1-3GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	23	2	8
424	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	18	2	14
425	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	20	2	12
426	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	21	2	10
427	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	15	6	39
428	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	17	5	29
429	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	11	4	34
430	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	20	1	5
431	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	16	3	18
432	Galb1-4Galb-Sp10	25	12	48
433	Galb1-6Galb-Sp10	30	11	38
434	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb-Sp8	31	2	7
435	GalNAcb1-6GalNAcb-Sp8	30	2	7
436	(6S)Galb1-3GlcNAcb-Sp0	39	5	13
437	(6S)Galb1-3(6S)GlcNAc-Sp0	32	4	13
438	Fuca1-2Galb1-4 GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2(Fuca1-2Galb1-4GlcNAcb1-4)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	41	4	9
439	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-4(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	66	4	6
440	Galb1-4(Fuca1-3)GlcNAcb1-6GalNAc-Sp14	52	4	7
441	Galb1-4GlcNAcb1-2Mana-Sp0	43	5	12
442	Fuca1-2Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)GalNAc-Sp14	23	2	9
443	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	26	3	11
444	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	16	4	27
445	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	95	4	4
446	GalNAcb1-4Galb1-4Glcb-Sp0	38	7	19
447	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	43	3	7
448	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	34	4	11
449	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	23	2	7
450	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	31	3	11
451	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	34	3	9
452	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	19	3	17
453	Neu5Aca2-3Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	20	1	7
454	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	18	1	5
455	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	18	1	8

456	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	19	2	9
457	Neu5Aca2-6Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	17	2	14
458	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	20	3	15
459	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	21	2	10
460	Gala1-3(Fuca1-2)Galb1-3GalNAca-Sp8	38	4	11
461	Gala1-3(Fuca1-2)Galb1-3GalNAcb-Sp8	61	4	6
462	Glca1-6Glca1-6Glca1-6Glcb-Sp10	27	4	14
463	Glca1-4Glca1-4Glca1-4Glcb-Sp10	42	1	3
464	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	24	1	5
465	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	81	13	16
466	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	59	2	4
467	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	93	3	3
468	Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	46	4	8
469	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21	19	2	10
470	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana-Sp0	57	4	7
471	Neu5Aca2-3Galb1-4GlcNAcb1-6GalNAca-Sp14	18	3	20
472	Neu5Aca2-6Galb1-4GlcNAcb1-6GalNAca-Sp14	36	7	21
473	Neu5Aca2-6Galb1-4 GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3)GalNAca-Sp14	21	6	28
474	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	76	4	6
475	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	77	4	5
476	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	42	7	18
477	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	87	4	5
478	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	30	6	19
479	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	16	6	35
480	Galb1-3GlcNAcb1-6GalNAca-Sp14	17	6	35
481	Gala1-3Galb1-3GlcNAcb1-6GalNAca-Sp14	20	3	17
482	Galb1-3(Fuca1-4)GlcNAcb1-6GalNAca-Sp14	39	10	25
483	Neu5Aca2-3Galb1-3GlcNAcb1-6GalNAca-Sp14	29	1	5
484	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp0	45	7	16
485	Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-6(Neu5Aca2-3Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	36	2	5
486	Fuca1-2Galb1-4GlcNAcb1-6GalNAca-Sp14	39	2	6
487	Gala1-3Galb1-4GlcNAcb1-6GalNAca-Sp14	15	7	49
488	Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	49	6	11
489	Fuca1-2(6S)Galb1-3GlcNAcb-Sp0	27	4	14
490	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAca-Sp14	21	5	25
491	Fuca1-2Galb1-4GlcNAcb1-2Mana-Sp0	34	5	14
492	Fuca1-2Galb1-3(6S)GlcNAcb-Sp0	39	7	19
493	Fuca1-2(6S)Galb1-3(6S)GlcNAcb-Sp0	44	3	6
494	Neu5Aca2-6GalNAcb1-4(6S)GlcNAcb-Sp8	25	3	10
495	GalNAcb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	35	3	10
496	(3S)GalNAcb1-4(Fuca1-3)GlcNAcb-Sp8	33	5	16
497	Fuca1-2Galb1-3GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)GalNAca-Sp14	42	3	8
498	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAca-Sp14	20	6	32
499	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	19	1	5
500	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	18	2	14
501	Galb1-3GlcNAca1-3Galb1-4GlcNAcb-Sp8	29	4	13

502	Galb1-3(6S)GlcNAcb-Sp8	20	7	36
503	(6S)(4S)GalNAcb1-4GlcNAc-Sp8	32	7	22
504	(6S)GalNAcb1-4GlcNAc-Sp8	16	6	37
505	(3S)GalNAcb1-4(3S)GlcNAc-Sp8	38	4	10
506	GalNAcb1-4(6S)GlcNAc-Sp8	46	1	3
507	(3S)GalNAcb1-4GlcNAc-Sp8	55	2	3
508	(4S)GalNAcb-Sp10	35	1	3
509	Galb1-4(6P)GlcNAcb-Sp0	28	1	4
510	(6P)Galb1-4GlcNAcb-SP0	16	7	42
511	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAc-Sp14	18	3	18
512	Neu5Aca2-6Galb1-4GlcNAcb1-2Man-Sp0	19	6	32
513	Gala1-3Galb1-4GlcNAcb1-2Mana-Sp0	23	3	13
514	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana-Sp0	16	8	48
515	GalNAca1-3(Fuca1-2)Galb1-4 GlcNAcb1-2Mana-Sp0	19	2	10
516	Galb1-3GlcNAcb1-2Mana-Sp0	46	7	16
517	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAc-Sp14	21	1	6
518	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana-Sp0	22	2	10
519	Gala1-3Galb1-3GlcNAcb1-2Mana-Sp0	25	3	12
520	GalNAcb1-4GlcNAcb1-2Mana-Sp0	30	1	4
521	Neu5Aca2-3Galb1-3GalNAcb1-4Galb1-4Glcb-Sp0	9	6	65
522	GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	8	7	90
523	Galb1-4GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	18	2	10
524	Galb1-4GlcNAcb1-2 Mana1-6(Galb1-4GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	15	3	17
525	Fuca1-4(Galb1-3)GlcNAcb1-2 Mana-Sp0	69	2	3
526	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	20	1	7
527	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAc-Sp0	18	2	10
528	GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	23	1	4
529	Gala1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	25	1	5
530	Galb1-3GalNAcb1-3Gal-Sp21	76	6	8
531	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	65	8	13
532	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	22	1	2
533	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-6(Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12	15	4	24
534	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	80	5	6
535	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	49	6	12
536	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	11	8	80
537	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	59	6	11
538	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp25	33	2	7
539	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAc-Sp0	30	5	18
540	Neu5Aca2-8Neu5Gca2-3Galb1-4GlcNAc-Sp0	11	8	69
541	Neu5Gca2-8Neu5Aca2-3Galb1-4GlcNAc-Sp0	25	1	2
542	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAc-Sp0	15	6	40
543	Neu5Gca2-8Neu5Gca2-6Galb1-4GlcNAc-Sp0	26	2	9
544	Neu5Aca2-8Neu5Aca2-3Galb1-4GlcNAc-Sp0	5	2	42
545	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Man a1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp24	82	7	9
546	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Mana1-4GlcNAcb1-4GlcNAc-Sp24	57	17	29
547	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp24	74	3	5
548	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-3)GalNAca-Sp14	25	2	8
549	GalNAcb1-3GlcNAcb-Sp0	17	6	33

550	GalNAcb1-4GlcNAcb1-3GalNAcb1-4GlcNAcb-Sp0	26	1	4
551	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	61	16	26
552	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	57	10	17
553	GlcNAb1-3Galb1-3GalNAc-Sp14	20	4	18
554	Galb1-3GlcNAcb1-6(Galb1-3)GalNAc-Sp14	24	3	15
555	(3S)GlcAb1-3Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	22	2	9
556	(3S)GlcAb1-3Galb1-4GlcNAcb1-2Mana-Sp0	33	3	10
557	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	55	13	23
558	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	60	6	11
559	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp21	29	2	8
560	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	56	7	12
561	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	76	17	22
562	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	82	7	8
563	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	24	3	14
564	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	29	2	7
565	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	25	3	14
566	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	30	4	14
567	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	25	4	17
568	GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	22	3	14
569	GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	23	1	5
570	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	31	1	4
571	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	30	1	4
572	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	23	5	23
573	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	20	1	3
574	Galb1-4GlcNAcb1-3Galb1-3GalNAca-Sp14	8	6	70
575	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	25	5	21
576	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	30	2	6
577	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	22	1	4
578	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	30	2	6
579	GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	15	5	34
580	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	26	1	5
581	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	238	23	10
582	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	73	3	3
583	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	34	2	4
584	GlcNAcb1-3Fuca-Sp21	30	1	2
585	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp21	28	1	2

8. 6. 2. Awp1A (50 µg/mL) – Anti-His-488 (50 µg/mL)

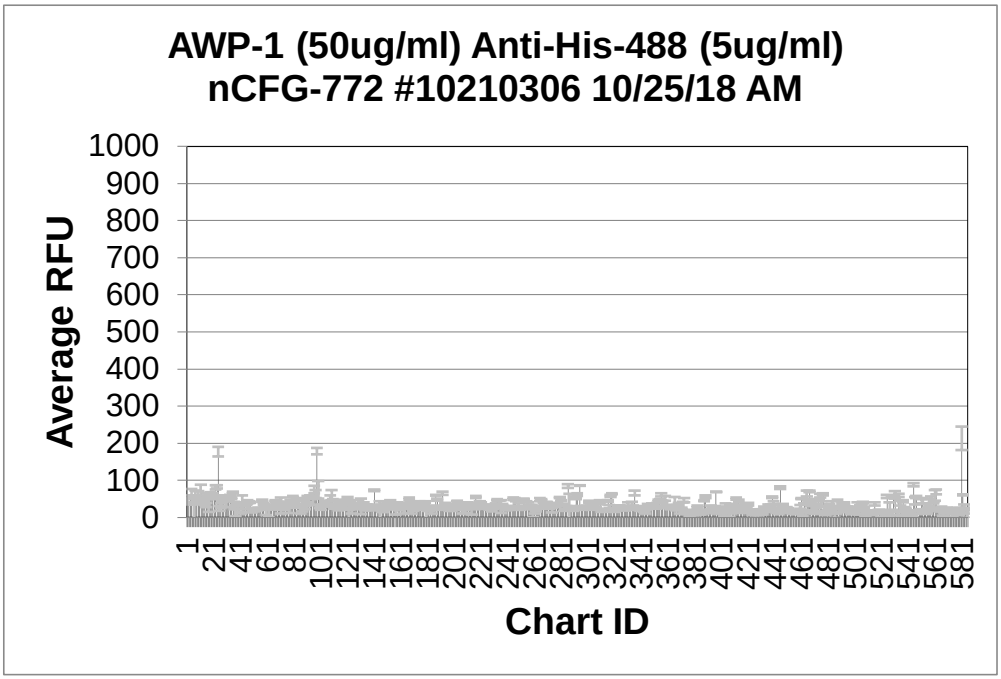


Chart ID	Sample (conc.) Secondary (conc.) Barcode# Slide # Request # Date Initials	Average RFU	StDev	%CV
1	Gala-Sp8	44	6	14
2	Glca-Sp8	39	4	10
3	Mana-Sp8	55	4	7
4	GalNAca-Sp8	64	13	21
5	GalNAca-Sp15	56	3	5
6	Fuca-Sp8	13	22	174
7	Fuca-Sp9	64	9	14
8	Rhaa-Sp8	47	3	5
9	Neu5Aca-Sp8	62	9	15
10	Neu5Aca-Sp11	43	3	7
11	Neu5Acb-Sp8	66	22	32
12	Galb-Sp8	43	3	6
13	Glcb-Sp8	51	8	16
14	Manb-Sp8	44	11	24
15	GalNAcb-Sp8	34	6	17
16	GlcNAcb-Sp0	57	8	14
17	GlcNAcb-Sp8	34	19	56
18	GlcN(Gc)b-Sp8	66	3	5
19	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAca-Sp8	53	4	8
20	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAc-Sp14	62	3	5
21	GlcNAcb1-6(GlcNAcb1-4)(GlcNAcb1-3)GlcNAc-Sp8	56	4	7
22	6S(3S)Galb1-4(6S)GlcNAcb-Sp0	81	6	7
23	6S(3S)Galb1-4GlcNAcb-Sp0	83	4	5
24	(3S)Galb1-4(Fuca1-3)(6S)Glc-Sp0	177	13	7
25	(3S)Galb1-4Glcb-Sp8	24	6	25
26	(3S)Galb1-4(6S)Glcb-Sp0	31	7	22
27	(3S)Galb1-4(6S)Glcb-Sp8	46	3	7
28	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp8	50	3	6
29	(3S)Galb1-3GalNAca-Sp8	56	4	6
30	(3S)Galb1-3GlcNAcb-Sp0	36	8	23
31	(3S)Galb1-3GlcNAcb-Sp8	53	5	10
32	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp0	50	2	3

33	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp8	59	1	2
34	(3S)Galb1-4(6S)GlcNAcb-Sp0	51	2	4
35	(3S)Galb1-4(6S)GlcNAcb-Sp8	67	2	3
36	(3S)Galb1-4GlcNAcb-Sp0	33	10	32
37	(3S)Galb1-4GlcNAcb-Sp8	19	15	81
38	(3S)Galb-Sp8	33	4	11
39	(6S)(4S)Galb1-4GlcNAcb-Sp0	29	6	22
40	(4S)Galb1-4GlcNAcb-Sp8	35	5	14
41	(6P)Mana-Sp8	13	3	21
42	(6S)Galb1-4Glcb-Sp0	50	9	18
43	(6S)Galb1-4Glcb-Sp8	28	6	21
44	(6S)Galb1-4GlcNAcb-Sp8	33	2	5
45	(6S)Galb1-4(6S)Glcb-Sp8	35	1	4
46	Neu5Aca2-3(6S)Galb1-4GlcNAcb-Sp8	43	4	10
47	(6S)GlcNAcb-Sp8	42	2	5
48	Neu5,9Ac$_2$a-Sp8	41	2	5
49	Neu5,9Ac2a2-6Galb1-4GlcNAcb-Sp8	19	6	32
50	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	21	4	19
51	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	19	2	10
52	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	22	2	9
53	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	22	3	11
54	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	24	2	9
55	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	22	2	10
56	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man-a1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	24	1	5
57	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	45	2	5
58	Fuca1-2Galb1-3GalNAcb1-3Gala-Sp9	37	3	7
59	Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp9	27	2	8
60	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb-Sp8	15	8	58
61	Fuca1-2Galb1-3GalNAca-Sp8	29	2	7
62	Fuca1-2Galb1-3GalNAca-Sp14	14	13	92
63	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	34	2	4
64	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp9	24	3	11
65	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp8	28	5	19
66	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10	26	4	14
67	Fuca1-2Galb1-3GlcNAcb-Sp0	44	2	5
68	Fuca1-2Galb1-3GlcNAcb-Sp8	31	8	26
69	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	40	2	6
70	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	39	4	10
71	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp0	32	21	65
72	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp8	28	3	11
73	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	23	2	10
74	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	25	4	14
75	Fuca1-2Galb1-4GlcNAcb-Sp0	34	2	6
76	Fuca1-2Galb1-4GlcNAcb-Sp8	36	2	6
77	Fuca1-2Galb1-4Glcb-Sp0	21	9	41
78	Fuca1-2Galb-Sp8	43	4	8
79	Fuca1-3GlcNAcb-Sp8	32	5	16
80	Fuca1-4GlcNAcb-Sp8	56	3	4
81	Fucb1-3GlcNAcb-Sp8	37	2	5
82	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	46	1	1
83	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	49	4	7
84	(3S)Galb1-4(Fuca1-3)Glcb-Sp0	31	4	14
85	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	30	3	10
86	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	19	9	46
87	GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	21	5	26
88	GlcNAcb1-3Galb1-3GalNAca-Sp8	49	2	3

89	GalNAca1-3(Fuca1-2)Galb-Sp8	26	4	16
90	GalNAca1-3(Fuca1-2)Galb-Sp18	37	4	10
91	GalNAca1-3GalNAcb-Sp8	54	3	6
92	GalNAca1-3Galb-Sp8	43	10	23
93	GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	56	4	7
94	GalNAcb1-3GalNAca-Sp8	53	3	6
95	GalNAcb1-3(Fuca1-2)Galb-Sp8	46	13	27
96	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	75	10	13
97	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0	63	11	18
98	GalNAcb1-4GlcNAcb-Sp0	179	8	5
99	GalNAcb1-4GlcNAcb-Sp8	75	24	32
100	Gala1-2Galb-Sp8	29	4	15
101	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	26	5	21
102	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp8	33	2	7
103	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	33	1	2
104	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp8	42	3	8
105	Gala1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	32	3	9
106	Gala1-3(Fuca1-2)Galb1-4Glcb-Sp0	34	4	12
107	Gala1-3(Fuca1-2)Galb-Sp8	33	3	9
108	Gala1-3(Fuca1-2)Galb-Sp18	52	8	16
109	Gala1-4(Gala1-3)Galb1-4GlcNAcb-Sp8	60	14	23
110	Gala1-3GalNAca-Sp8	48	1	3
111	Gala1-3GalNAca-Sp16	29	5	18
112	Gala1-3GalNAcb-Sp8	28	1	2
113	Gala1-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	28	1	5
114	Gala1-3Galb1-3GlcNAcb-Sp0	24	4	16
115	Gala1-3Galb1-4GlcNAcb-Sp8	38	4	9
116	Gala1-3Galb1-4Glcb-Sp0	30	3	9
117	Gala1-3Galb1-4Glc-Sp10	31	7	22
118	Gala1-3Galb-Sp8	37	2	6
119	Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	45	3	6
120	Gala1-4Galb1-4GlcNAcb-Sp0	29	5	16
121	Gala1-4Galb1-4GlcNAcb-Sp8	52	3	6
122	Gala1-4Galb1-4Glcb-Sp0	7	14	212
123	Gala1-4GlcNAcb-Sp8	33	8	25
124	Gala1-6Glcb-Sp8	17	7	39
125	Galb1-2Galb-Sp8	32	6	17
126	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	30	2	5
127	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	25	3	11
128	Galb1-3(Fuca1-4)GlcNAc-Sp0	30	5	15
129	Galb1-3(Fuca1-4)GlcNAc-Sp8	39	8	19
130	Fuca1-4(Galb1-3)GlcNAcb-Sp8	33	5	15
131	Galb1-4GlcNAcb1-6GalNAca-Sp8	30	21	69
132	Galb1-4GlcNAcb1-6GalNAc-Sp14	36	2	5
133	GlcNAcb1-6(Galb1-3)GalNAca-Sp8	36	4	12
134	GlcNAcb1-6(Galb1-3)GalNAca-Sp14	25	6	23
135	Neu5Aca2-6(Galb1-3)GalNAca-Sp8	32	4	12
136	Neu5Aca2-6(Galb1-3)GalNAca-Sp14	23	2	7
137	Neu5Acb2-6(Galb1-3)GalNAca-Sp8	33	3	10
138	Neu5Aca2-6(Galb1-3)GlcNAcb1-4Galb1-4Glcb-Sp10	24	4	18
139	Galb1-3GalNAca-Sp8	28	3	12
140	Galb1-3GalNAca-Sp14	24	3	14
141	Galb1-3GalNAca-Sp16	73	2	2
142	Galb1-3GalNAcb-Sp8	30	2	5
143	Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	26	1	4
144	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	30	1	3
145	Galb1-3GalNAcb1-4Galb1-4Glcb-Sp8	21	23	113
146	Galb1-3Galb-Sp8	31	5	15
147	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	16	1	6
148	Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10	18	3	15
149	Galb1-3GlcNAcb-Sp0	34	3	8

150	Galb1-3GlcNAcb-Sp8	24	2	7
151	Galb1-4(Fuca1-3)GlcNAcb-Sp0	38	5	12
152	Galb1-4(Fuca1-3)GlcNAcb-Sp8	38	1	3
153	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	42	5	11
154	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	23	1	6
155	Galb1-4(6S)Glcb-Sp0	33	1	3
156	Galb1-4(6S)Glcb-Sp8	38	1	3
157	Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	20	12	63
158	Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	36	4	12
159	Galb1-4GlcNAcb1-3GalNAca-Sp8	27	5	21
160	Galb1-4GlcNAcb1-3GalNAc-Sp14	22	2	11
161	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	38	2	5
162	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	26	2	9
163	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	21	9	45
164	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	35	5	14
165	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp8	27	2	9
166	Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp8	36	1	2
167	Galb1-4GlcNAcb1-6(Galb1-3)GalNAc-Sp14	51	3	6
168	Galb1-4GlcNAcb-Sp0	40	3	7
169	Galb1-4GlcNAcb-Sp8	19	9	45
170	Galb1-4GlcNAcb-Sp23	20	3	15
171	Galb1-4Glcb-Sp0	27	5	18
172	Galb1-4Glcb-Sp8	23	2	7
173	GlcNAca1-3Galb1-4GlcNAcb-Sp8	28	2	7
174	GlcNAca1-6Galb1-4GlcNAcb-Sp8	27	3	11
175	GlcNAcb1-2Galb1-3GalNAca-Sp8	41	4	10
176	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp8	27	1	5
177	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp14	24	1	2
178	GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAcb-Sp8	37	2	7
179	GlcNAcb1-3GalNAca-Sp8	41	2	5
180	GlcNAcb1-3GalNAca-Sp14	12	13	115
181	GlcNAcb1-3Galb-Sp8	22	5	20
182	GlcNAcb1-3Galb1-4GlcNAcb-Sp0	24	6	25
183	GlcNAcb1-3Galb1-4GlcNAcb-Sp8	26	2	7
184	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	14	4	30
185	GlcNAcb1-3Galb1-4Glcb-Sp0	29	4	14
186	GlcNAcb1-4-MDPLys	26	2	9
187	GlcNAcb1-6(GlcNAcb1-4)GalNAca-Sp8	60	1	2
188	GlcNAcb1-4Galb1-4GlcNAcb-Sp8	48	3	6
189	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	26	1	5
190	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	26	1	4
191	GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8	27	1	5
192	GlcNAcb1-6GalNAca-Sp8	65	4	6
193	GlcNAcb1-6GalNAca-Sp14	24	3	11
194	GlcNAcb1-6Galb1-4GlcNAcb-Sp8	37	2	7
195	Glca1-4Glcb-Sp8	24	2	10
196	Glca1-4Glca-Sp8	33	2	6
197	Glca1-6Glca1-6Glcb-Sp8	23	6	26
198	Glcb1-4Glcb-Sp8	26	2	8
199	Glcb1-6Glcb-Sp8	22	4	18
200	G-ol-Sp8	28	6	20
201	GlcAa-Sp8	31	2	5
202	GlcAb-Sp8	30	5	17
203	GlcAb1-3Galb-Sp8	44	2	4
204	GlcAb1-6Galb-Sp8	36	3	7
205	KDNa2-3Galb1-3GlcNAcb-Sp0	40	1	3
206	KDNa2-3Galb1-4GlcNAcb-Sp0	25	1	4
207	Mana1-2Mana1-2Mana1-3Mana-Sp9	19	9	45
208	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana-Sp9	23	1	6
209	Mana1-2Mana1-3Mana-Sp9	18	6	34

210	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	29	1	5
211	Mana1-6(Mana1-3)Mana-Sp9	39	2	6
212	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9	28	2	5
213	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	29	2	6
214	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	22	13	61
215	Manb1-4GlcNAcb-Sp0	26	2	6
216	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	24	2	7
217	(3S)Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	55	3	5
218	Fuca1-2(6S)Galb1-4GlcNAcb-Sp0	30	5	18
219	Fuca1-2Galb1-4(6S)GlcNAcb-Sp8	32	5	14
220	Fuca1-2(6S)Galb1-4(6S)Glcb-Sp0	42	3	6
221	Neu5Aca2-3Galb1-3GalNAca-Sp8	38	2	5
222	Neu5Aca2-3Galb1-3GalNAca-Sp14	31	3	9
223	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	27	3	10
224	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	29	2	5
225	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0	26	1	5
226	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	32	2	5
227	Neu5Aca2-8Neu5Aca2-8Neu5Aca-Sp8	23	2	8
228	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	32	4	12
229	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp8	20	4	19
230	GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	26	2	7
231	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	26	1	3
232	Neu5Aca2-6(Neu5Aca2-3)GalNAca-Sp8	35	2	6
233	Neu5Aca2-3GalNAca-Sp8	47	3	5
234	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0	33	2	6
235	Neu5Aca2-3Galb1-3(6S)GlcNAc-Sp8	40	5	11
236	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp8	30	17	58
237	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	35	2	6
238	Neu5Aca2-3Galb1-4(Neu5Aca2-3Galb1-3)GlcNAcb-Sp8	29	1	5
239	Neu5Aca2-3Galb1-3(6S)GalNAca-Sp8	25	5	20
240	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp8	22	3	12
241	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	29	2	6
242	Neu5Aca2-3Galb-Sp8	27	2	6
243	Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	29	1	5
244	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	28	1	2
245	Fuca1-2(6S)Galb1-4Glcb-Sp0	51	2	5
246	Neu5Aca2-3Galb1-3GlcNAcb-Sp0	49	3	5
247	Neu5Aca2-3Galb1-4(6S)GlcNAcb-Sp8	46	3	5
248	Neu5Aca2-3Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	26	5	18
249	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	33	5	17
250	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	22	2	10
251	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	25	2	9
252	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8	26	2	7
253	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8	50	1	3
254	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	27	2	6
255	Neu5Aca2-3Galb1-4GlcNAcb-Sp0	42	2	4
256	Neu5Aca2-3Galb1-4GlcNAcb-Sp8	32	9	29
257	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	32	2	6
258	Fuca1-2Galb1-4(6S)Glcb-Sp0	31	3	9
259	Neu5Aca2-3Galb1-4Glcb-Sp0	32	2	6
260	Neu5Aca2-3Galb1-4Glcb-Sp8	14	13	99
261	Neu5Aca2-6GalNAca-Sp8	19	8	41
262	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0	19	3	17
263	Neu5Aca2-6Galb1-4(6S)GlcNAcb-Sp8	27	3	10
264	Neu5Aca2-6Galb1-4GlcNAcb-Sp0	26	2	7
265	Neu5Aca2-6Galb1-4GlcNAcb-Sp8	48	4	8
266	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	48	1	3
267	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	30	1	3

268	Neu5Aca2-6Galb1-4Glcb-Sp0	40	2	4
269	Neu5Aca2-6Galb1-4Glcb-Sp8	34	1	3
270	Neu5Aca2-6Galb-Sp8	41	1	2
271	Neu5Aca2-8Neu5Aca-Sp8	29	2	6
272	Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0	27	3	10
273	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	31	3	11
274	Neu5Acb2-6GalNAca-Sp8	21	7	32
275	Neu5Acb2-6Galb1-4GlcNAcb-Sp8	34	5	13
276	Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	31	1	4
277	Neu5Gca2-3Galb1-3GlcNAcb-Sp0	30	3	9
278	Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	37	2	6
279	Neu5Gca2-3Galb1-4GlcNAcb-Sp0	34	1	1
280	Neu5Gca2-3Galb1-4Glcb-Sp0	54	2	3
281	Neu5Gca2-6GalNAca-Sp0	45	1	3
282	Neu5Gca2-6Galb1-4GlcNAcb-Sp0	35	2	5
283	Neu5Gca-Sp8	26	9	35
284	Neu5Aca2-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	24	1	2
285	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	26	6	22
286	Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp0	84	5	6
287	Galb1-4(Fuca1-3)(6S)Glcb-Sp0	31	33	110
288	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	25	3	12
289	Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	27	4	14
290	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	20	2	9
291	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	24	1	6
292	4S(3S)Galb1-4GlcNAcb-Sp0	53	4	7
293	(6S)Galb1-4(6S)GlcNAcb-Sp0	62	4	6
294	(6P)Glcb-Sp10	26	1	4
295	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3)GalNAca-Sp14	87	1	1
296	Galb1-3Galb1-4GlcNAcb-Sp8	29	1	3
297	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	23	2	7
298	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	26	2	7
299	GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	26	2	9
300	Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0	28	5	16
301	Galb1-4GlcNAcb1-6Galb1-4GlcNAcb-Sp0	29	1	2
302	GalNAcb1-3Galb-Sp8	42	4	8
303	GlcAb1-3GlcNAcb-Sp8	40	2	4
304	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	21	1	4
305	GlcNAcb1-3Man-Sp10	32	1	4
306	GlcNAcb1-4GlcNAcb-Sp10	27	2	7
307	GlcNAcb1-4GlcNAcb-Sp12	28	1	2
308	MurNAcb1-4GlcNAcb-Sp10	23	1	4
309	Mana1-6Manb-Sp10	39	1	2
310	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb-Sp10	41	5	13
311	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	18	3	14
312	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	20	2	9
313	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	18	1	5
314	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	20	1	4
315	Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	19	1	7
316	Neu5Aca2-8Neu5Acb-Sp17	44	2	4
317	Neu5Aca2-8Neu5Aca2-8Neu5Acb-Sp8	29	4	15
318	Neu5Gcb2-6Galb1-4GlcNAc-Sp8	58	2	4
319	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	64	2	3
320	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	18	1	3
321	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	17	2	12

322	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	23	2	11
323	Neu5,9Ac2a2-3Galb1-3GlcNAcb-Sp0	22	2	10
324	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	28	1	4
325	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	31	1	2
326	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	23	1	2
327	Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	28	2	5
328	GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	21	1	5
329	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	22	2	11
330	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	24	2	8
331	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAc-Sp14	35	5	14
332	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	23	2	9
333	GlcNAca1-4Galb1-4GlcNAcb-Sp0	28	3	11
334	GlcNAca1-4Galb1-3GlcNAcb-Sp0	34	9	26
335	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	30	2	5
336	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	66	6	9
337	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	30	1	3
338	GlcNAca1-4Galb1-3GalNAc-Sp14	24	2	8
339	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	23	2	9
340	Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	22	3	12
341	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	19	1	3
342	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	19	1	3
343	Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	17	2	10
344	Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	15	4	26
345	Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	23	1	2
346	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	33	3	10
347	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	31	3	11
348	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	28	1	5
349	(6S)GlcNAcb1-3Galb1-4GlcNAcb-Sp0	36	2	6
350	KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0	33	1	3
351	KDNa2-6Galb1-4GlcNAc-Sp0	29	1	2
352	KDNa2-3Galb1-4Glc-Sp0	25	2	7
353	KDNa2-3Galb1-3GalNAca-Sp14	35	4	11
354	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	46	3	6
355	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	46	4	8
356	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAb-Sp20	61	4	7
357	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	42	3	7
358	Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	28	3	10
359	Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-6(Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	52	7	13
360	Neu5Aca2-6GlcNAcb1-4GlcNAc-Sp21	32	2	5
361	Neu5Aca2-6GlcNAcb1-4GlcNAcb1-4GlcNAc-Sp21	30	3	10
362	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	29	2	8
363	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	24	1	5
364	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	31	2	5
365	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	30	1	2
366	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	44	11	25
367	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	22	3	13
368	Galα1-3(Fucα1-2)Galβ1-3GlcNAcβ1-2Manα1-6(Galα1-3(Fucα1-2)Galβ1-3GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp20	31	2	5

369	Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	36	3	9
370	Neu5Aca2-3Galb1-4GlcNAcb1-3GalNAc-Sp14	15	4	27
371	Neu5Aca2-6Galb1-4GlcNAcb1-3GalNAc-Sp14	24	2	10
372	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	42	1	2
373	GalNAcb1-4GlcNAcb1-2Mana1-6(GalNAcb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	46	6	12
374	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0	14	4	32
375	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	15	3	17
376	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp0	19	1	3
377	Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	9	6	67
378	Galb1-4GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	17	4	22
379	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	17	2	13
380	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	6	6	116
381	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	18	1	3
382	GlcNAcb1-2Mana1-6(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	9	3	28
383	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	30	3	10
384	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	14	5	39
385	Galb1-3GlcNAcb1-3GalNAca-Sp14	14	5	39
386	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb1-3GalNAca-Sp14	20	2	9
387	GalNAca1-3(Fuca1-2)Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	15	5	30
388	Gala1-3Galb1-3GlcNAcb1-2Mana1-6(Gala1-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	46	1	3
389	Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	56	4	6
390	GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	14	2	14
391	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	17	2	10
392	Neu5Aca2-3Galb1-3GlcNAcb1-3GalNAca-Sp14	15	1	9
393	Fuca1-2Galb1-4GlcNAcb1-3GalNAca-Sp14	27	5	18
394	Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	26	4	14
395	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	15	3	22
396	Gala1-4Galb1-3GlcNAcb1-2Mana1-6(Gala1-4Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	29	1	5
397	Gala1-4Galb1-4GlcNAcb1-2Mana1-6(Gala1-4Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	69	1	2
398	Gala1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	14	3	19
399	Galb1-3GlcNAcb1-6Galb1-4GlcNAcb-Sp0	25	3	14
400	Galb1-3GlcNAca1-6Galb1-4GlcNAcb-Sp0	6	9	172
401	GalNAcb1-3Gala1-6Galb1-4Glcb-Sp8	21	9	44
402	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)Glcb-Sp21	18	1	6
403	Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	10	5	47
404	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	33	5	16
405	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	24	3	14
406	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	19	1	4
407	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	8	5	68
408	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	17	6	38
409	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	36	1	4
410	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	23	3	13
411	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	31	2	5
412	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	51	3	6
413	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	32	2	7
414	GlcNAcb1-2(GlcNAcb1-6)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	45	3	7
415	Fuca1-2Galb1-3GlcNAcb1-3GalNAc-Sp14	17	4	21
416	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	17	4	25
417	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	22	2	10
418	Gala1-3Galb1-3GlcNAcb1-3GalNAc-Sp14	19	4	22

419	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	32	3	10
420	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	31	3	9
421	Galb1-3GlcNAcb1-6(Galb1-3GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	36	4	10
422	Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	15	1	7
423	Fuca1-3GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	16	2	10
424	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	9	4	39
425	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	18	4	25
426	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	16	1	5
427	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	12	5	46
428	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	15	4	29
429	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	9	2	23
430	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	13	3	25
431	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	11	3	27
432	Galb1-4Galb-Sp10	15	6	41
433	Galb1-6Galb-Sp10	22	7	33
434	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb-Sp8	25	1	2
435	GalNAcb1-6GalNAcb-Sp8	19	3	15
436	(6S)Galb1-3GlcNAcb-Sp0	31	4	15
437	(6S)Galb1-3(6S)GlcNAc-Sp0	28	4	15
438	Fuca1-2Galb1-4 GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2(Fuca1-2Galb1-4GlcNAcb1-4)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	32	2	7
439	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-4(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	55	2	4
440	Galb1-4(Fuca1-3)GlcNAcb1-6GalNAc-Sp14	40	4	10
441	Galb1-4GlcNAcb1-2Mana-Sp0	32	5	16
442	Fuca1-2Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)GalNAc-Sp14	15	2	16
443	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	18	2	11
444	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	12	4	31
445	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	80	3	4
446	GalNAcb1-4Galb1-4Glcb-Sp0	29	5	19
447	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	34	3	8
448	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	26	1	3
449	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	17	1	3
450	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	25	3	11
451	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	23	2	8
452	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	17	3	17
453	Neu5Aca2-3Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	13	1	10
454	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	15	2	16
455	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	13	2	19

456	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	14	2	14
457	Neu5Aca2-6Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	12	3	21
458	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	17	3	16
459	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	15	2	14
460	Gala1-3(Fuca1-2)Galb1-3GalNAca-Sp8	28	5	19
461	Gala1-3(Fuca1-2)Galb1-3GalNAcb-Sp8	51	2	3
462	Glca1-6Glca1-6Glca1-6Glcb-Sp10	13	8	66
463	Glca1-4Glca1-4Glca1-4Glcb-Sp10	37	1	1
464	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	8	9	116
465	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	64	8	13
466	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	47	2	5
467	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	70	3	4
468	Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	35	4	12
469	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21	16	3	22
470	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana-Sp0	39	5	13
471	Neu5Aca2-3Galb1-4GlcNAcb1-6GalNAca-Sp14	12	2	15
472	Neu5Aca2-6Galb1-4GlcNAcb1-6GalNAca-Sp14	31	6	21
473	Neu5Aca2-6Galb1-4 GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3)GalNAca-Sp14	14	2	18
474	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	48	2	3
475	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	59	5	8
476	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	36	8	21
477	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	63	3	5
478	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	23	7	32
479	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	13	6	46
480	Galb1-3GlcNAcb1-6GalNAca-Sp14	12	3	28
481	Gala1-3Galb1-3GlcNAcb1-6GalNAca-Sp14	15	4	29
482	Galb1-3(Fuca1-4)GlcNAcb1-6GalNAca-Sp14	29	7	23
483	Neu5Aca2-3Galb1-3GlcNAcb1-6GalNAca-Sp14	24	3	13
484	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp0	27	16	60
485	Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-6(Neu5Aca2-3Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	29	1	2
486	Fuca1-2Galb1-4GlcNAcb1-6GalNAca-Sp14	15	15	104
487	Gala1-3Galb1-4GlcNAcb1-6GalNAca-Sp14	13	3	21
488	Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	39	9	23
489	Fuca1-2(6S)Galb1-3GlcNAcb-Sp0	20	5	24
490	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAca-Sp14	20	5	24
491	Fuca1-2Galb1-4GlcNAcb1-2Mana-Sp0	26	3	11
492	Fuca1-2Galb1-3(6S)GlcNAcb-Sp0	35	5	13
493	Fuca1-2(6S)Galb1-3(6S)GlcNAcb-Sp0	37	3	9
494	Neu5Aca2-6GalNAcb1-4(6S)GlcNAcb-Sp8	22	3	15
495	GalNAcb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	26	3	12
496	(3S)GalNAcb1-4(Fuca1-3)GlcNAcb-Sp8	27	1	5
497	Fuca1-2Galb1-3GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)GalNAca-Sp14	28	1	4
498	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAca-Sp14	17	8	48
499	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	14	2	12
500	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	14	2	17
501	Galb1-3GlcNAca1-3Galb1-4GlcNAcb-Sp8	25	5	20

502	Galb1-3(6S)GlcNAcb-Sp8	17	8	47
503	(6S)(4S)GalNAcb1-4GlcNAc-Sp8	31	10	33
504	(6S)GalNAcb1-4GlcNAc-Sp8	12	4	36
505	(3S)GalNAcb1-4(3S)GlcNAc-Sp8	32	1	2
506	GalNAcb1-4(6S)GlcNAc-Sp8	35	1	4
507	(3S)GalNAcb1-4GlcNAc-Sp8	41	2	5
508	(4S)GalNAcb-Sp10	28	2	6
509	Galb1-4(6P)GlcNAcb-Sp0	10	9	100
510	(6P)Galb1-4GlcNAcb-SP0	10	2	20
511	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAc-Sp14	14	2	12
512	Neu5Aca2-6Galb1-4GlcNAcb1-2Man-Sp0	13	3	21
513	Gala1-3Galb1-4GlcNAcb1-2Mana-Sp0	17	3	15
514	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana-Sp0	12	5	40
515	GalNAca1-3(Fuca1-2)Galb1-4 GlcNAcb1-2Mana-Sp0	13	2	16
516	Galb1-3GlcNAcb1-2Mana-Sp0	37	4	12
517	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAc-Sp14	15	1	7
518	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana-Sp0	14	3	21
519	Gala1-3Galb1-3GlcNAcb1-2Mana-Sp0	18	3	14
520	GalNAcb1-4GlcNAcb1-2Mana-Sp0	20	1	5
521	Neu5Aca2-3Galb1-3GalNAcb1-4Galb1-4Glcb-Sp0	7	4	67
522	GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	5	6	132
523	Galb1-4GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	14	3	21
524	Galb1-4GlcNAcb1-2 Mana1-6(Galb1-4GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	13	2	13
525	Fuca1-4(Galb1-3)GlcNAcb1-2 Mana-Sp0	57	4	7
526	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	14	3	24
527	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAc-Sp0	23	18	78
528	GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	18	1	7
529	Gala1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	20	1	7
530	Galb1-3GalNAcb1-3Gal-Sp21	58	2	4
531	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	61	10	16
532	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	16	1	5
533	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-6(Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12	12	5	45
534	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	60	5	8
535	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	39	6	15
536	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	7	2	27
537	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	44	3	8
538	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp25	24	4	17
539	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAc-Sp0	24	3	11
540	Neu5Aca2-8Neu5Gca2-3Galb1-4GlcNAc-Sp0	7	3	41
541	Neu5Gca2-8Neu5Aca2-3Galb1-4GlcNAc-Sp0	16	3	21
542	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAc-Sp0	12	4	34
543	Neu5Gca2-8Neu5Gca2-6Galb1-4GlcNAc-Sp0	18	2	11
544	Neu5Aca2-8Neu5Aca2-3Galb1-4GlcNAc-Sp0	5	1	10
545	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Man a1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp24	88	5	5
546	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Mana1-4GlcNAcb1-4GlcNAc-Sp24	45	10	23
547	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp24	56	3	5
548	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-3)GalNAca-Sp14	19	1	3
549	GalNAcb1-3GlcNAcb-Sp0	15	4	30

550	GalNAcb1-4GlcNAcb1-3GalNAcb1-4GlcNAcb-Sp0	20	2	10
551	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	43	7	15
552	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	46	7	16
553	GlcNAb1-3Galb1-3GalNAc-Sp14	14	4	27
554	Galb1-3GlcNAcb1-6(Galb1-3)GalNAc-Sp14	16	2	14
555	(3S)GlcAb1-3Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	14	3	21
556	(3S)GlcAb1-3Galb1-4GlcNAcb1-2Mana-Sp0	25	2	7
557	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	40	9	22
558	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	48	9	20
559	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp21	19	2	9
560	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	40	2	5
561	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	60	14	24
562	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	69	6	9
563	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	20	3	13
564	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	22	4	17
565	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	19	3	15
566	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	26	4	15
567	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	11	6	56
568	GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	14	2	15
569	GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	17	6	37
570	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	25	1	4
571	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	7	10	149
572	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	16	3	20
573	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	15	1	4
574	Galb1-4GlcNAcb1-3Galb1-3GalNAca-Sp14	6	4	65
575	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	18	4	22
576	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	22	1	4
577	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	14	3	18
578	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	25	2	8
579	GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	11	4	36
580	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	23	2	8
581	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	213	31	15
582	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	61	2	2
583	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	23	13	55
584	GlcNAcb1-3Fuca-Sp21	25	1	4
585	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp21	10	10	104

8. 6. 3. Awp3A (5 µg/mL) – Anti-His-488 (5 µg/mL)

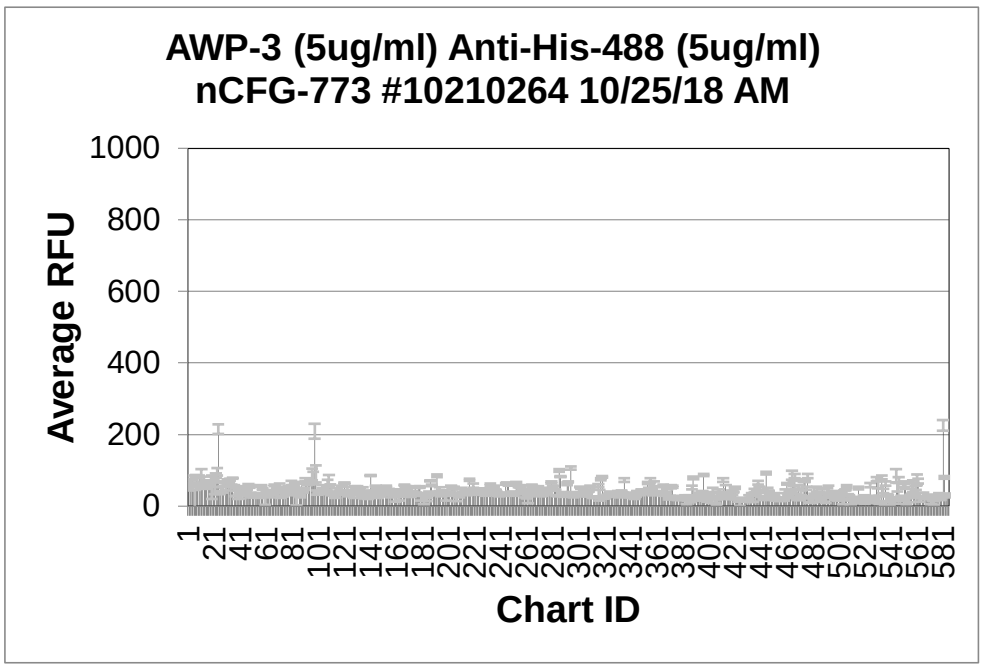


Chart ID	Sample (conc.) Secondary (conc.) Barcode# Slide # Request # Date Initials	Average RFU	StDev	%CV
1	Gala-Sp8	58	7	12
2	Glca-Sp8	49	5	11
3	Mana-Sp8	66	10	16
4	GalNAca-Sp8	78	7	9
5	GalNAca-Sp15	65	2	3
6	Fuca-Sp8	16	29	183
7	Fuca-Sp9	83	3	4
8	Rhaa-Sp8	60	3	5
9	Neu5Aca-Sp8	82	2	3
10	Neu5Aca-Sp11	55	5	8
11	Neu5Acb-Sp8	77	26	33
12	Galb-Sp8	61	4	6
13	Glcb-Sp8	70	12	18
14	Manb-Sp8	61	4	7
15	GalNAcb-Sp8	52	5	9
16	GlcNAcb-Sp0	65	7	10
17	GlcNAcb-Sp8	57	3	5
18	GlcN(Gc)b-Sp8	68	3	5
19	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAca-Sp8	52	32	62
20	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAc-Sp14	69	3	5
21	GlcNAcb1-6(GlcNAcb1-4)(GlcNAcb1-3)GlcNAc-Sp8	52	19	36
22	6S(3S)Galb1-4(6S)GlcNAcb-Sp0	86	6	7
23	6S(3S)Galb1-4GlcNAcb-Sp0	96	11	11
24	(3S)Galb1-4(Fuca1-3)(6S)Glc-Sp0	215	14	6
25	(3S)Galb1-4Glcb-Sp8	38	5	13
26	(3S)Galb1-4(6S)Glcb-Sp0	46	7	15
27	(3S)Galb1-4(6S)Glcb-Sp8	46	8	17
28	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp8	60	4	7
29	(3S)Galb1-3GalNAca-Sp8	69	0	0
30	(3S)Galb1-3GlcNAcb-Sp0	50	7	14
31	(3S)Galb1-3GlcNAcb-Sp8	69	5	8
32	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp0	65	2	4

33	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp8	74	3	4
34	(3S)Galb1-4(6S)GlcNAcb-Sp0	58	2	4
35	(3S)Galb1-4(6S)GlcNAcb-Sp8	78	2	2
36	(3S)Galb1-4GlcNAcb-Sp0	56	3	6
37	(3S)Galb1-4GlcNAcb-Sp8	38	11	29
38	(3S)Galb-Sp8	35	7	19
39	(6S)(4S)Galb1-4GlcNAcb-Sp0	31	12	38
40	(4S)Galb1-4GlcNAcb-Sp8	44	11	25
41	(6P)Mana-Sp8	28	6	22
42	(6S)Galb1-4Glcb-Sp0	53	2	5
43	(6S)Galb1-4Glcb-Sp8	38	1	4
44	(6S)Galb1-4GlcNAcb-Sp8	39	1	3
45	(6S)Galb1-4(6S)Glcb-Sp8	47	5	11
46	Neu5Aca2-3(6S)Galb1-4GlcNAcb-Sp8	49	4	7
47	(6S)GlcNAcb-Sp8	45	16	36
48	Neu5,9Ac$_2$a-Sp8	55	4	7
49	Neu5,9Ac2a2-6Galb1-4GlcNAcb-Sp8	36	1	4
50	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	28	3	11
51	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	26	3	11
52	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	34	1	3
53	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	27	3	9
54	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	31	3	8
55	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	27	4	14
56	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man-a1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	35	1	2
57	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	56	3	5
58	Fuca1-2Galb1-3GalNAcb1-3Gala-Sp9	48	2	5
59	Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp9	34	1	1
60	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb-Sp8	19	16	82
61	Fuca1-2Galb1-3GalNAca-Sp8	37	3	8
62	Fuca1-2Galb1-3GalNAca-Sp14	28	1	5
63	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	42	3	7
64	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp9	34	3	9
65	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp8	32	8	24
66	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10	37	2	5
67	Fuca1-2Galb1-3GlcNAcb-Sp0	58	3	4
68	Fuca1-2Galb1-3GlcNAcb-Sp8	41	2	6
69	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	48	2	4
70	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	50	4	8
71	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp0	57	5	9
72	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp8	32	10	31
73	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	28	1	3
74	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	32	4	13
75	Fuca1-2Galb1-4GlcNAcb-Sp0	45	3	6
76	Fuca1-2Galb1-4GlcNAcb-Sp8	43	9	21
77	Fuca1-2Galb1-4Glcb-Sp0	35	4	12
78	Fuca1-2Galb-Sp8	55	1	2
79	Fuca1-3GlcNAcb-Sp8	44	5	12
80	Fuca1-4GlcNAcb-Sp8	64	5	8
81	Fucb1-3GlcNAcb-Sp8	49	3	7
82	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	56	3	6
83	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	65	2	3
84	(3S)Galb1-4(Fuca1-3)Glcb-Sp0	20	18	91
85	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	40	5	13
86	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	29	1	4
87	GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	28	4	12
88	GlcNAcb1-3Galb1-3GalNAca-Sp8	65	2	2

89	GalNAca1-3(Fuca1-2)Galb-Sp8	40	2	5
90	GalNAca1-3(Fuca1-2)Galb-Sp18	48	3	7
91	GalNAca1-3GalNAcb-Sp8	73	5	7
92	GalNAca1-3Galb-Sp8	60	2	3
93	GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	65	4	5
94	GalNAcb1-3GalNAca-Sp8	63	3	4
95	GalNAcb1-3(Fuca1-2)Galb-Sp8	67	5	7
96	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	89	16	18
97	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0	80	16	19
98	GalNAcb1-4GlcNAcb-Sp0	209	21	10
99	GalNAcb1-4GlcNAcb-Sp8	90	24	27
100	Gala1-2Galb-Sp8	37	6	15
101	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	37	1	2
102	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp8	37	4	10
103	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	37	2	5
104	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp8	54	4	8
105	Gala1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	41	2	4
106	Gala1-3(Fuca1-2)Galb1-4Glcb-Sp0	43	1	3
107	Gala1-3(Fuca1-2)Galb-Sp8	44	2	5
108	Gala1-3(Fuca1-2)Galb-Sp18	64	9	14
109	Gala1-4(Gala1-3)Galb1-4GlcNAcb-Sp8	71	15	21
110	Gala1-3GalNAca-Sp8	59	1	2
111	Gala1-3GalNAca-Sp16	32	1	4
112	Gala1-3GalNAcb-Sp8	38	3	7
113	Gala1-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	32	3	8
114	Gala1-3Galb1-3GlcNAcb-Sp0	29	6	22
115	Gala1-3Galb1-4GlcNAcb-Sp8	41	9	23
116	Gala1-3Galb1-4Glcb-Sp0	36	3	8
117	Gala1-3Galb1-4Glc-Sp10	37	2	5
118	Gala1-3Galb-Sp8	44	2	4
119	Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	56	1	2
120	Gala1-4Galb1-4GlcNAcb-Sp0	35	2	5
121	Gala1-4Galb1-4GlcNAcb-Sp8	64	2	4
122	Gala1-4Galb1-4Glcb-Sp0	39	5	12
123	Gala1-4GlcNAcb-Sp8	49	5	9
124	Gala1-6Glcb-Sp8	30	3	11
125	Galb1-2Galb-Sp8	37	4	12
126	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	39	3	7
127	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	33	2	7
128	Galb1-3(Fuca1-4)GlcNAc-Sp0	34	11	34
129	Galb1-3(Fuca1-4)GlcNAc-Sp8	47	9	20
130	Fuca1-4(Galb1-3)GlcNAcb-Sp8	43	6	13
131	Galb1-4GlcNAcb1-6GalNAca-Sp8	53	2	3
132	Galb1-4GlcNAcb1-6GalNAc-Sp14	46	2	4
133	GlcNAcb1-6(Galb1-3)GalNAca-Sp8	45	6	14
134	GlcNAcb1-6(Galb1-3)GalNAca-Sp14	31	2	6
135	Neu5Aca2-6(Galb1-3)GalNAca-Sp8	46	6	14
136	Neu5Aca2-6(Galb1-3)GalNAca-Sp14	25	4	17
137	Neu5Acb2-6(Galb1-3)GalNAca-Sp8	41	2	5
138	Neu5Aca2-6(Galb1-3)GlcNAcb1-4Galb1-4Glcb-Sp10	26	2	8
139	Galb1-3GalNAca-Sp8	28	8	28
140	Galb1-3GalNAca-Sp14	29	2	6
141	Galb1-3GalNAca-Sp16	85	1	2
142	Galb1-3GalNAcb-Sp8	37	2	6
143	Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	35	1	3
144	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	39	3	6
145	Galb1-3GalNAcb1-4Galb1-4Glcb-Sp8	56	2	3
146	Galb1-3Galb-Sp8	40	8	20
147	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	24	2	9
148	Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10	28	1	3
149	Galb1-3GlcNAcb-Sp0	40	3	8

150	Galb1-3GlcNAcb-Sp8	35	3	8
151	Galb1-4(Fuca1-3)GlcNAcb-Sp0	46	7	14
152	Galb1-4(Fuca1-3)GlcNAcb-Sp8	49	2	4
153	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	55	2	3
154	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	33	2	5
155	Galb1-4(6S)Glcb-Sp0	45	4	8
156	Galb1-4(6S)Glcb-Sp8	46	1	3
157	Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	33	8	25
158	Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	43	4	9
159	Galb1-4GlcNAcb1-3GalNAca-Sp8	33	2	5
160	Galb1-4GlcNAcb1-3GalNAc-Sp14	22	9	43
161	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	45	2	5
162	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	27	1	5
163	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	29	14	48
164	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	42	3	8
165	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp8	35	2	6
166	Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp8	46	1	2
167	Galb1-4GlcNAcb1-6(Galb1-3)GalNAc-Sp14	58	1	2
168	Galb1-4GlcNAcb-Sp0	54	1	2
169	Galb1-4GlcNAcb-Sp8	41	5	11
170	Galb1-4GlcNAcb-Sp23	30	1	5
171	Galb1-4Glcb-Sp0	29	4	12
172	Galb1-4Glcb-Sp8	28	2	9
173	GlcNAca1-3Galb1-4GlcNAcb-Sp8	39	3	8
174	GlcNAca1-6Galb1-4GlcNAcb-Sp8	34	3	8
175	GlcNAcb1-2Galb1-3GalNAca-Sp8	53	2	4
176	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp8	36	2	6
177	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp14	32	2	8
178	GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAcb-Sp8	51	1	2
179	GlcNAcb1-3GalNAca-Sp8	53	2	4
180	GlcNAcb1-3GalNAca-Sp14	23	8	36
181	GlcNAcb1-3Galb-Sp8	34	3	8
182	GlcNAcb1-3Galb1-4GlcNAcb-Sp0	20	16	78
183	GlcNAcb1-3Galb1-4GlcNAcb-Sp8	31	2	7
184	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	19	5	24
185	GlcNAcb1-3Galb1-4Glcb-Sp0	31	3	9
186	GlcNAcb1-4-MDPLys	35	5	16
187	GlcNAcb1-6(GlcNAcb1-4)GalNAca-Sp8	70	2	3
188	GlcNAcb1-4Galb1-4GlcNAcb-Sp8	58	2	4
189	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	32	1	3
190	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	32	1	2
191	GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8	37	2	5
192	GlcNAcb1-6GalNAca-Sp8	84	4	5
193	GlcNAcb1-6GalNAca-Sp14	36	2	4
194	GlcNAcb1-6Galb1-4GlcNAcb-Sp8	47	5	11
195	Glca1-4Glcb-Sp8	29	2	5
196	Glca1-4Glca-Sp8	43	1	3
197	Glca1-6Glca1-6Glcb-Sp8	33	1	3
198	Glcb1-4Glcb-Sp8	35	3	9
199	Glcb1-6Glcb-Sp8	26	10	41
200	G-ol-Sp8	32	4	12
201	GlcAa-Sp8	36	3	9
202	GlcAb-Sp8	35	5	14
203	GlcAb1-3Galb-Sp8	55	2	3
204	GlcAb1-6Galb-Sp8	49	2	5
205	KDNa2-3Galb1-3GlcNAcb-Sp0	52	1	2
206	KDNa2-3Galb1-4GlcNAcb-Sp0	34	1	1
207	Mana1-2Mana1-2Mana1-3Mana-Sp9	22	8	37
208	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana-Sp9	33	2	5
209	Mana1-2Mana1-3Mana-Sp9	28	5	16

210	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	36	1	2
211	Mana1-6(Mana1-3)Mana-Sp9	49	2	5
212	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9	36	1	1
213	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	36	1	3
214	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	40	2	4
215	Manb1-4GlcNAcb-Sp0	34	5	14
216	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	34	1	2
217	(3S)Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	73	3	4
218	Fuca1-2(6S)Galb1-4GlcNAcb-Sp0	39	4	10
219	Fuca1-2Galb1-4(6S)GlcNAcb-Sp8	41	6	14
220	Fuca1-2(6S)Galb1-4(6S)Glcb-Sp0	57	6	10
221	Neu5Aca2-3Galb1-3GalNAca-Sp8	44	1	1
222	Neu5Aca2-3Galb1-3GalNAca-Sp14	35	3	8
223	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	32	1	2
224	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	39	1	4
225	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0	37	1	3
226	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	40	1	1
227	Neu5Aca2-8Neu5Aca2-8Neu5Aca-Sp8	34	1	4
228	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	48	1	3
229	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp8	30	1	3
230	GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	34	2	6
231	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	33	1	3
232	Neu5Aca2-6(Neu5Aca2-3)GalNAca-Sp8	43	2	5
233	Neu5Aca2-3GalNAca-Sp8	60	3	5
234	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0	41	1	3
235	Neu5Aca2-3Galb1-3(6S)GlcNAc-Sp8	48	1	1
236	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp8	55	2	4
237	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	45	1	2
238	Neu5Aca2-3Galb1-4(Neu5Aca2-3Galb1-3)GlcNAcb-Sp8	39	3	7
239	Neu5Aca2-3Galb1-3(6S)GalNAca-Sp8	33	4	13
240	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp8	30	2	6
241	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	33	2	7
242	Neu5Aca2-3Galb-Sp8	32	4	12
243	Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	35	2	6
244	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	35	1	1
245	Fuca1-2(6S)Galb1-4Glcb-Sp0	62	2	3
246	Neu5Aca2-3Galb1-3GlcNAcb-Sp0	64	2	3
247	Neu5Aca2-3Galb1-4(6S)GlcNAcb-Sp8	60	3	5
248	Neu5Aca2-3Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	36	1	4
249	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	36	5	15
250	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	35	3	8
251	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	33	3	11
252	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8	36	3	7
253	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8	64	3	5
254	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	34	1	2
255	Neu5Aca2-3Galb1-4GlcNAcb-Sp0	54	2	3
256	Neu5Aca2-3Galb1-4GlcNAcb-Sp8	55	4	7
257	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	42	2	6
258	Fuca1-2Galb1-4(6S)Glcb-Sp0	45	3	6
259	Neu5Aca2-3Galb1-4Glcb-Sp0	46	2	4
260	Neu5Aca2-3Galb1-4Glcb-Sp8	33	9	27
261	Neu5Aca2-6GalNAca-Sp8	28	6	22
262	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0	19	11	60
263	Neu5Aca2-6Galb1-4(6S)GlcNAcb-Sp8	32	5	16
264	Neu5Aca2-6Galb1-4GlcNAcb-Sp0	31	1	4
265	Neu5Aca2-6Galb1-4GlcNAcb-Sp8	57	2	4
266	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	54	2	3
267	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	38	2	5

268	Neu5Aca2-6Galb1-4Glcb-Sp0	50	2	4
269	Neu5Aca2-6Galb1-4Glcb-Sp8	44	2	5
270	Neu5Aca2-6Galb-Sp8	52	3	5
271	Neu5Aca2-8Neu5Aca-Sp8	37	2	4
272	Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0	31	4	12
273	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	39	6	15
274	Neu5Acb2-6GalNAca-Sp8	31	2	8
275	Neu5Acb2-6Galb1-4GlcNAcb-Sp8	48	3	7
276	Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	39	1	1
277	Neu5Gca2-3Galb1-3GlcNAcb-Sp0	37	4	10
278	Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	48	1	3
279	Neu5Gca2-3Galb1-4GlcNAcb-Sp0	45	1	1
280	Neu5Gca2-3Galb1-4Glcb-Sp0	66	3	4
281	Neu5Gca2-6GalNAca-Sp0	56	2	3
282	Neu5Gca2-6Galb1-4GlcNAcb-Sp0	45	2	4
283	Neu5Gca-Sp8	44	2	3
284	Neu5Aca2-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	30	1	3
285	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	29	2	6
286	Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp0	99	3	3
287	Galb1-4(Fuca1-3)(6S)Glcb-Sp0	82	1	2
288	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	36	3	8
289	Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	30	4	12
290	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	27	1	5
291	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	31	1	5
292	4S(3S)Galb1-4GlcNAcb-Sp0	63	3	5
293	(6S)Galb1-4(6S)GlcNAcb-Sp0	67	1	1
294	(6P)Glcb-Sp10	33	1	4
295	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3)GalNAca-Sp14	106	4	4
296	Galb1-3Galb1-4GlcNAcb-Sp8	36	4	11
297	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	27	3	10
298	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	34	2	7
299	GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	32	2	5
300	Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0	36	2	4
301	Galb1-4GlcNAcb1-6Galb1-4GlcNAcb-Sp0	36	1	4
302	GalNAcb1-3Galb-Sp8	54	1	1
303	GlcAb1-3GlcNAcb-Sp8	49	1	2
304	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	28	1	5
305	GlcNAcb1-3Man-Sp10	41	1	3
306	GlcNAcb1-4GlcNAcb-Sp10	40	1	2
307	GlcNAcb1-4GlcNAcb-Sp12	33	2	5
308	MurNAcb1-4GlcNAcb-Sp10	33	5	16
309	Mana1-6Manb-Sp10	44	3	7
310	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb-Sp10	55	2	4
311	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	26	1	3
312	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	25	2	10
313	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	24	2	9
314	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	26	1	4
315	Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	24	1	3
316	Neu5Aca2-8Neu5Acb-Sp17	57	4	7
317	Neu5Aca2-8Neu5Aca2-8Neu5Acb-Sp8	30	15	49
318	Neu5Gcb2-6Galb1-4GlcNAc-Sp8	74	3	4
319	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	82	2	3
320	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	22	1	2
321	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	23	1	6

322	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	27	1	5
323	Neu5,9Ac2a2-3Galb1-3GlcNAcb-Sp0	28	3	10
324	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	33	2	5
325	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	38	2	5
326	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	30	1	3
327	Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	35	1	4
328	GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	27	1	2
329	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	28	1	2
330	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	35	2	5
331	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAc-Sp14	41	2	5
332	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	27	2	7
333	GlcNAca1-4Galb1-4GlcNAcb-Sp0	34	6	18
334	GlcNAca1-4Galb1-3GlcNAcb-Sp0	43	1	3
335	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	34	3	9
336	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	72	5	7
337	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	35	2	6
338	GlcNAca1-4Galb1-3GalNAc-Sp14	30	1	5
339	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	27	1	2
340	Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	30	5	17
341	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	26	2	7
342	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	26	2	7
343	Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	24	2	6
344	Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	23	3	15
345	Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	28	1	5
346	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	40	2	4
347	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	35	1	4
348	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	33	2	5
349	(6S)GlcNAcb1-3Galb1-4GlcNAcb-Sp0	44	2	5
350	KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0	43	1	2
351	KDNa2-6Galb1-4GlcNAc-Sp0	38	2	5
352	KDNa2-3Galb1-4Glc-Sp0	37	3	7
353	KDNa2-3Galb1-3GalNAca-Sp14	45	2	5
354	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	65	3	4
355	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	48	3	6
356	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAb-Sp20	72	5	7
357	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	47	2	3
358	Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	31	1	2
359	Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-6(Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	62	7	11
360	Neu5Aca2-6GlcNAcb1-4GlcNAc-Sp21	39	2	5
361	Neu5Aca2-6GlcNAcb1-4GlcNAcb1-4GlcNAc-Sp21	33	1	2
362	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	35	2	4
363	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	34	1	2
364	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	40	1	3
365	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	41	1	1
366	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	56	3	6
367	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	31	2	6
368	Galα1-3(Fucα1-2)Galβ1-3GlcNAcβ1-2Manα1-6(Galα1-3(Fucα1-2)Galβ1-3GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp20	36	3	7

369	Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	49	3	6
370	Neu5Aca2-3Galb1-4GlcNAcb1-3GalNAc-Sp14	22	4	16
371	Neu5Aca2-6Galb1-4GlcNAcb1-3GalNAc-Sp14	26	3	11
372	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	54	4	8
373	GalNAcb1-4GlcNAcb1-2Mana1-6(GalNAcb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	54	3	5
374	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0	23	1	2
375	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	20	1	5
376	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp0	22	2	11
377	Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	19	3	15
378	Galb1-4GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	26	3	10
379	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	23	1	2
380	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	19	2	11
381	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	22	3	11
382	GlcNAcb1-2Mana1-6(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	21	2	11
383	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	26	4	16
384	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	13	8	65
385	Galb1-3GlcNAcb1-3GalNAca-Sp14	20	4	22
386	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb1-3GalNAca-Sp14	25	5	18
387	GalNAca1-3(Fuca1-2)Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	14	4	27
388	Gala1-3Galb1-3GlcNAcb1-2Mana1-6(Gala1-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	51	6	11
389	Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	78	4	5
390	GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	19	2	9
391	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	21	5	25
392	Neu5Aca2-3Galb1-3GlcNAcb1-3GalNAca-Sp14	21	1	2
393	Fuca1-2Galb1-4GlcNAcb1-3GalNAca-Sp14	31	6	21
394	Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	29	4	12
395	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	20	2	9
396	Gala1-4Galb1-3GlcNAcb1-2Mana1-6(Gala1-4Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	41	1	3
397	Gala1-4Galb1-4GlcNAcb1-2Mana1-6(Gala1-4Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	87	3	4
398	Gala1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	19	3	16
399	Galb1-3GlcNAcb1-6Galb1-4GlcNAcb-Sp0	31	4	12
400	Galb1-3GlcNAca1-6Galb1-4GlcNAcb-Sp0	17	5	31
401	GalNAcb1-3Gala1-6Galb1-4Glcb-Sp8	33	4	11
402	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)Glcb-Sp21	23	2	8
403	Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	15	6	41
404	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	45	6	14
405	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	30	10	33
406	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	16	6	42
407	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	11	8	71
408	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	24	4	18
409	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	41	5	12
410	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	25	1	5
411	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	36	3	9
412	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	72	5	7
413	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	32	6	19
414	GlcNAcb1-2(GlcNAcb1-6)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	58	0	0
415	Fuca1-2Galb1-3GlcNAcb1-3GalNAc-Sp14	25	6	22
416	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	25	1	4
417	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	29	3	10
418	Gala1-3Galb1-3GlcNAcb1-3GalNAc-Sp14	25	5	20

419	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	36	7	19
420	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	38	2	4
421	Galb1-3GlcNAcb1-6(Galb1-3GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	51	2	5
422	Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	20	1	7
423	Fuca1-3GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	22	1	6
424	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	17	4	26
425	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	10	10	105
426	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	21	1	5
427	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	16	4	22
428	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	20	3	13
429	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	17	3	16
430	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	16	4	23
431	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	19	3	14
432	Galb1-4Galb-Sp10	25	5	20
433	Galb1-6Galb-Sp10	20	10	49
434	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb-Sp8	32	4	11
435	GalNAcb1-6GalNAcb-Sp8	22	9	43
436	(6S)Galb1-3GlcNAcb-Sp0	43	5	12
437	(6S)Galb1-3(6S)GlcNAc-Sp0	36	3	8
438	Fuca1-2Galb1-4 GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2(Fuca1-2Galb1-4GlcNAcb1-4)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	42	2	6
439	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-4(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	65	6	9
440	Galb1-4(Fuca1-3)GlcNAcb1-6GalNAc-Sp14	50	5	10
441	Galb1-4GlcNAcb1-2Mana-Sp0	40	3	8
442	Fuca1-2Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)GalNAc-Sp14	27	1	2
443	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	23	2	11
444	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	20	3	16
445	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	92	3	3
446	GalNAcb1-4Galb1-4Glcb-Sp0	43	7	16
447	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	42	1	3
448	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	31	3	9
449	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	26	2	7
450	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	29	1	2
451	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	32	4	11
452	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	22	4	20
453	Neu5Aca2-3Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	18	2	10
454	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	20	1	6
455	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	18	3	15

456	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	18	5	26
457	Neu5Aca2-6Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	20	2	8
458	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	21	2	12
459	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	19	3	14
460	Gala1-3(Fuca1-2)Galb1-3GalNAca-Sp8	41	5	11
461	Gala1-3(Fuca1-2)Galb1-3GalNAcb-Sp8	62	3	4
462	Glca1-6Glca1-6Glca1-6Glcb-Sp10	26	4	14
463	Glca1-4Glca1-4Glca1-4Glcb-Sp10	41	1	3
464	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	24	1	6
465	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	86	13	15
466	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	63	7	11
467	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	84	6	7
468	Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	41	7	17
469	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21	19	2	9
470	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana-Sp0	53	5	10
471	Neu5Aca2-3Galb1-4GlcNAcb1-6GalNAca-Sp14	19	3	16
472	Neu5Aca2-6Galb1-4GlcNAcb1-6GalNAca-Sp14	37	7	18
473	Neu5Aca2-6Galb1-4 GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3)GalNAca-Sp14	25	1	5
474	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	71	3	5
475	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	70	3	4
476	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	49	4	9
477	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	84	6	7
478	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	18	9	52
479	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	18	8	43
480	Galb1-3GlcNAcb1-6GalNAca-Sp14	15	6	39
481	Gala1-3Galb1-3GlcNAcb1-6GalNAca-Sp14	17	6	35
482	Galb1-3(Fuca1-4)GlcNAcb1-6GalNAca-Sp14	44	8	19
483	Neu5Aca2-3Galb1-3GlcNAcb1-6GalNAca-Sp14	29	3	9
484	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp0	46	7	16
485	Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-6(Neu5Aca2-3Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	35	2	6
486	Fuca1-2Galb1-4GlcNAcb1-6GalNAca-Sp14	38	2	6
487	Gala1-3Galb1-4GlcNAcb1-6GalNAca-Sp14	16	3	17
488	Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	49	1	3
489	Fuca1-2(6S)Galb1-3GlcNAcb-Sp0	19	7	38
490	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAca-Sp14	28	6	20
491	Fuca1-2Galb1-4GlcNAcb1-2Mana-Sp0	22	13	61
492	Fuca1-2Galb1-3(6S)GlcNAcb-Sp0	41	8	18
493	Fuca1-2(6S)Galb1-3(6S)GlcNAcb-Sp0	47	10	22
494	Neu5Aca2-6GalNAcb1-4(6S)GlcNAcb-Sp8	28	2	7
495	GalNAcb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	36	3	9
496	(3S)GalNAcb1-4(Fuca1-3)GlcNAcb-Sp8	38	3	7
497	Fuca1-2Galb1-3GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)GalNAca-Sp14	38	3	9
498	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAca-Sp14	19	1	4
499	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	25	2	7
500	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	17	2	10
501	Galb1-3GlcNAca1-3Galb1-4GlcNAcb-Sp8	26	1	2

502	Galb1-3(6S)GlcNAcb-Sp8	27	10	37
503	(6S)(4S)GalNAcb1-4GlcNAc-Sp8	27	11	39
504	(6S)GalNAcb1-4GlcNAc-Sp8	13	8	61
505	(3S)GalNAcb1-4(3S)GlcNAc-Sp8	39	5	14
506	GalNAcb1-4(6S)GlcNAc-Sp8	45	2	4
507	(3S)GalNAcb1-4GlcNAc-Sp8	55	4	7
508	(4S)GalNAcb-Sp10	31	2	8
509	Galb1-4(6P)GlcNAcb-Sp0	29	2	6
510	(6P)Galb1-4GlcNAcb-SP0	11	6	54
511	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAc-Sp14	18	3	14
512	Neu5Aca2-6Galb1-4GlcNAcb1-2Man-Sp0	20	1	5
513	Gala1-3Galb1-4GlcNAcb1-2Mana-Sp0	22	4	16
514	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana-Sp0	14	6	41
515	GalNAca1-3(Fuca1-2)Galb1-4 GlcNAcb1-2Mana-Sp0	15	5	32
516	Galb1-3GlcNAcb1-2Mana-Sp0	50	4	7
517	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAc-Sp14	16	6	37
518	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana-Sp0	21	2	11
519	Gala1-3Galb1-3GlcNAcb1-2Mana-Sp0	23	1	4
520	GalNAcb1-4GlcNAcb1-2Mana-Sp0	29	1	4
521	Neu5Aca2-3Galb1-3GalNAcb1-4Galb1-4Glcb-Sp0	12	5	42
522	GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	16	5	30
523	Galb1-4GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	19	1	5
524	Galb1-4GlcNAcb1-2 Mana1-6(Galb1-4GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	14	4	30
525	Fuca1-4(Galb1-3)GlcNAcb1-2 Mana-Sp0	58	6	11
526	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	18	3	15
527	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAc-Sp0	16	4	23
528	GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	22	2	10
529	Gala1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	28	1	2
530	Galb1-3GalNAcb1-3Gal-Sp21	75	6	8
531	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	60	10	17
532	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	18	11	63
533	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-6(Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12	10	4	39
534	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	81	5	6
535	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	42	6	14
536	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	6	4	62
537	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	63	6	9
538	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp25	30	3	10
539	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAc-Sp0	30	2	5
540	Neu5Aca2-8Neu5Gca2-3Galb1-4GlcNAc-Sp0	8	5	60
541	Neu5Gca2-8Neu5Aca2-3Galb1-4GlcNAc-Sp0	22	1	4
542	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAc-Sp0	18	2	10
543	Neu5Gca2-8Neu5Gca2-6Galb1-4GlcNAc-Sp0	23	1	4
544	Neu5Aca2-8Neu5Aca2-3Galb1-4GlcNAc-Sp0	11	4	39
545	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Man a1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp24	85	18	21
546	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Mana1-4GlcNAcb1-4GlcNAc-Sp24	54	13	24
547	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp24	73	9	12
548	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-3)GalNAca-Sp14	23	2	8
549	GalNAcb1-3GlcNAcb-Sp0	21	8	38

550	GalNAcb1-4GlcNAcb1-3GalNAcb1-4GlcNAcb-Sp0	23	4	15
551	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	56	5	8
552	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	59	10	18
553	GlcNAb1-3Galb1-3GalNAc-Sp14	14	9	68
554	Galb1-3GlcNAcb1-6(Galb1-3)GalNAc-Sp14	24	1	6
555	(3S)GlcAb1-3Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	17	6	37
556	(3S)GlcAb1-3Galb1-4GlcNAcb1-2Mana-Sp0	29	3	10
557	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	50	16	32
558	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	56	17	30
559	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp21	30	2	5
560	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	51	6	12
561	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	74	14	20
562	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	70	7	9
563	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	24	12	52
564	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	25	2	7
565	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	23	2	9
566	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	31	6	20
567	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	23	3	11
568	GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	22	1	5
569	GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	23	1	2
570	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	30	1	3
571	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	29	1	5
572	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	13	7	58
573	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	18	3	16
574	Galb1-4GlcNAcb1-3Galb1-3GalNAca-Sp14	8	5	58
575	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	24	7	28
576	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	26	5	18
577	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	21	2	9
578	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	33	4	12
579	GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	18	1	3
580	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	26	2	7
581	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	226	14	6
582	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	80	4	5
583	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	34	1	3
584	GlcNAcb1-3Fuca-Sp21	33	1	2
585	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp21	26	2	6

8. 6. 4. Awp3A (50 µg/mL) – Anti-His-488 (50 µg/mL)

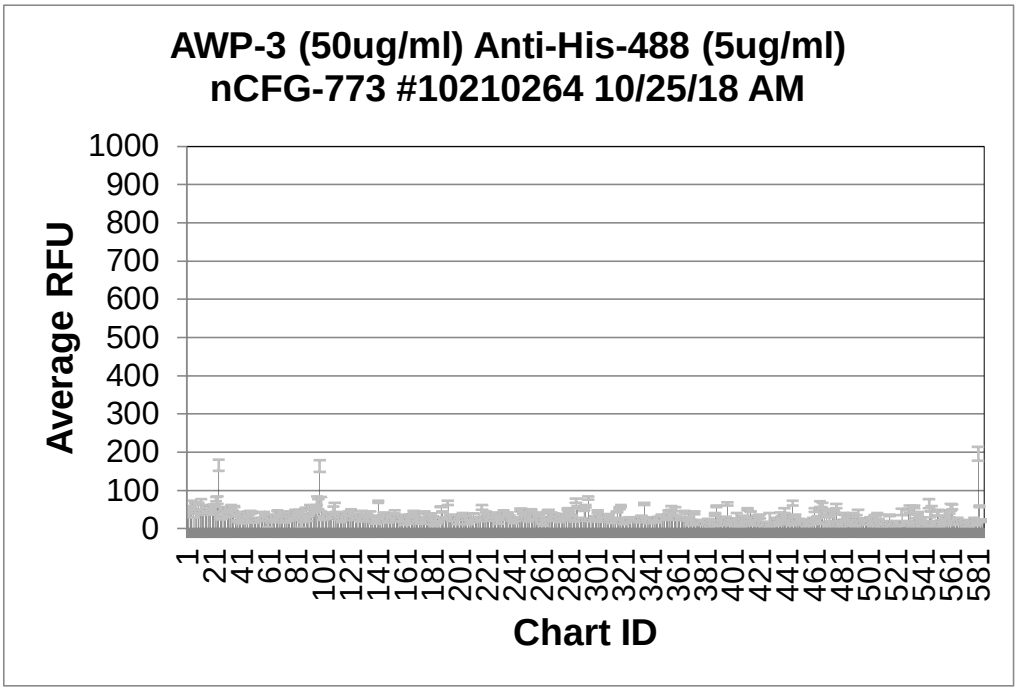


Chart ID	Sample (conc.) Secondary (conc.) Barcode# Slide # Request # Date Initials	Average RFU	StDev	%CV
1	Gala-Sp8	43	6	13
2	Glca-Sp8	34	5	15
3	Mana-Sp8	51	9	17
4	GalNAca-Sp8	62	11	18
5	GalNAca-Sp15	51	3	6
6	Fuca-Sp8	11	20	181
7	Fuca-Sp9	67	2	3
8	Rhaa-Sp8	47	2	5
9	Neu5Aca-Sp8	65	2	3
10	Neu5Aca-Sp11	41	2	5
11	Neu5Acb-Sp8	57	20	36
12	Galb-Sp8	47	4	8
13	Glcb-Sp8	53	9	17
14	Manb-Sp8	50	1	2
15	GalNAcb-Sp8	42	4	9
16	GlcNAcb-Sp0	52	9	18
17	GlcNAcb-Sp8	43	7	16
18	GlcN(Gc)b-Sp8	55	5	9
19	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAca-Sp8	48	11	23
20	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAc-Sp14	61	3	5
21	GlcNAcb1-6(GlcNAcb1-4)(GlcNAcb1-3)GlcNAc-Sp8	52	8	15
22	6S(3S)Galb1-4(6S)GlcNAcb-Sp0	77	6	8
23	6S(3S)Galb1-4GlcNAcb-Sp0	74	10	13
24	(3S)Galb1-4(Fuca1-3)(6S)Glc-Sp0	166	15	9
25	(3S)Galb1-4Glcb-Sp8	28	3	11
26	(3S)Galb1-4(6S)Glcb-Sp0	30	4	14
27	(3S)Galb1-4(6S)Glcb-Sp8	35	7	19
28	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp8	46	3	8
29	(3S)Galb1-3GalNAca-Sp8	54	3	5
30	(3S)Galb1-3GlcNAcb-Sp0	37	6	17
31	(3S)Galb1-3GlcNAcb-Sp8	53	2	4

32	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp0	53	3	5
33	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp8	58	4	6
34	(3S)Galb1-4(6S)GlcNAcb-Sp0	47	2	4
35	(3S)Galb1-4(6S)GlcNAcb-Sp8	58	2	3
36	(3S)Galb1-4GlcNAcb-Sp0	40	2	4
37	(3S)Galb1-4GlcNAcb-Sp8	28	8	30
38	(3S)Galb-Sp8	28	6	22
39	(6S)(4S)Galb1-4GlcNAcb-Sp0	24	9	37
40	(4S)Galb1-4GlcNAcb-Sp8	36	7	21
41	(6P)Mana-Sp8	18	4	24
42	(6S)Galb1-4Glcb-Sp0	43	1	3
43	(6S)Galb1-4Glcb-Sp8	30	1	3
44	(6S)Galb1-4GlcNAcb-Sp8	31	2	8
45	(6S)Galb1-4(6S)Glcb-Sp8	33	4	12
46	Neu5Aca2-3(6S)Galb1-4GlcNAcb-Sp8	41	5	11
47	(6S)GlcNAcb-Sp8	35	11	32
48	Neu5,9Ac$_2$a-Sp8	42	3	7
49	Neu5,9Ac2a2-6Galb1-4GlcNAcb-Sp8	24	4	19
50	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	21	4	21
51	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	21	0	0
52	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	25	1	4
53	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	23	3	14
54	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	23	3	11
55	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	23	1	6
56	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man-a1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	28	1	2
57	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	44	1	1
58	Fuca1-2Galb1-3GalNAcb1-3Gala-Sp9	36	2	5
59	Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp9	25	1	5
60	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb-Sp8	16	12	75
61	Fuca1-2Galb1-3GalNAca-Sp8	29	4	15
62	Fuca1-2Galb1-3GalNAca-Sp14	21	4	17
63	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	34	2	6
64	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp9	25	4	16
65	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp8	24	6	27
66	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10	29	2	6
67	Fuca1-2Galb1-3GlcNAcb-Sp0	47	2	4
68	Fuca1-2Galb1-3GlcNAcb-Sp8	32	3	10
69	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	41	1	3
70	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	36	4	11
71	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp0	44	1	3
72	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp8	22	6	28
73	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	21	2	8
74	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	23	3	14
75	Fuca1-2Galb1-4GlcNAcb-Sp0	32	4	12
76	Fuca1-2Galb1-4GlcNAcb-Sp8	31	6	18
77	Fuca1-2Galb1-4Glcb-Sp0	26	3	10
78	Fuca1-2Galb-Sp8	41	3	7
79	Fuca1-3GlcNAcb-Sp8	31	4	13
80	Fuca1-4GlcNAcb-Sp8	47	3	6
81	Fucb1-3GlcNAcb-Sp8	36	4	12
82	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	43	2	5
83	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	47	2	5
84	(3S)Galb1-4(Fuca1-3)Glcb-Sp0	13	11	84
85	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	35	3	7
86	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	22	5	22
87	GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	23	1	6

88	GlcNAcb1-3Galb1-3GalNAca-Sp8	52	3	5
89	GalNAca1-3(Fuca1-2)Galb-Sp8	25	5	20
90	GalNAca1-3(Fuca1-2)Galb-Sp18	35	1	3
91	GalNAca1-3GalNAcb-Sp8	58	4	8
92	GalNAca1-3Galb-Sp8	46	5	12
93	GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	58	2	3
94	GalNAcb1-3GalNAca-Sp8	46	3	6
95	GalNAcb1-3(Fuca1-2)Galb-Sp8	52	3	6
96	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	73	11	15
97	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0	64	13	20
98	GalNAcb1-4GlcNAcb-Sp0	164	15	9
99	GalNAcb1-4GlcNAcb-Sp8	64	19	29
100	Gala1-2Galb-Sp8	26	4	13
101	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	29	3	11
102	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp8	29	3	12
103	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	31	1	3
104	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp8	40	3	8
105	Gala1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	32	1	4
106	Gala1-3(Fuca1-2)Galb1-4Glcb-Sp0	35	2	6
107	Gala1-3(Fuca1-2)Galb-Sp8	32	3	10
108	Gala1-3(Fuca1-2)Galb-Sp18	49	8	16
109	Gala1-4(Gala1-3)Galb1-4GlcNAcb-Sp8	53	15	28
110	Gala1-3GalNAca-Sp8	42	2	5
111	Gala1-3GalNAca-Sp16	27	1	5
112	Gala1-3GalNAcb-Sp8	27	3	12
113	Gala1-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	27	2	7
114	Gala1-3Galb1-3GlcNAcb-Sp0	21	4	19
115	Gala1-3Galb1-4GlcNAcb-Sp8	29	6	20
116	Gala1-3Galb1-4Glcb-Sp0	28	3	11
117	Gala1-3Galb1-4Glc-Sp10	27	1	4
118	Gala1-3Galb-Sp8	37	3	7
119	Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	43	2	4
120	Gala1-4Galb1-4GlcNAcb-Sp0	29	2	8
121	Gala1-4Galb1-4GlcNAcb-Sp8	49	2	3
122	Gala1-4Galb1-4Glcb-Sp0	27	3	12
123	Gala1-4GlcNAcb-Sp8	34	6	17
124	Gala1-6Glcb-Sp8	25	2	7
125	Galb1-2Galb-Sp8	30	4	15
126	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	28	2	8
127	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	27	2	9
128	Galb1-3(Fuca1-4)GlcNAc-Sp0	25	7	27
129	Galb1-3(Fuca1-4)GlcNAc-Sp8	36	11	32
130	Fuca1-4(Galb1-3)GlcNAcb-Sp8	33	4	13
131	Galb1-4GlcNAcb1-6GalNAca-Sp8	38	3	7
132	Galb1-4GlcNAcb1-6GalNAc-Sp14	32	2	6
133	GlcNAcb1-6(Galb1-3)GalNAca-Sp8	31	7	21
134	GlcNAcb1-6(Galb1-3)GalNAca-Sp14	24	1	5
135	Neu5Aca2-6(Galb1-3)GalNAca-Sp8	37	9	24
136	Neu5Aca2-6(Galb1-3)GalNAca-Sp14	21	2	10
137	Neu5Acb2-6(Galb1-3)GalNAca-Sp8	32	3	8
138	Neu5Aca2-6(Galb1-3)GlcNAcb1-4Galb1-4Glcb-Sp10	18	5	25
139	Galb1-3GalNAca-Sp8	19	5	29
140	Galb1-3GalNAca-Sp14	23	2	10
141	Galb1-3GalNAca-Sp16	71	3	4
142	Galb1-3GalNAcb-Sp8	28	1	2
143	Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	25	1	4
144	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	29	1	4
145	Galb1-3GalNAcb1-4Galb1-4Glcb-Sp8	39	3	7
146	Galb1-3Galb-Sp8	28	4	13
147	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	19	2	9
148	Galb1-3GlcNAcb1-3Galb1-4Glcb-Sp10	19	3	15

149	Galb1-3GlcNAcb-Sp0	27	3	11
150	Galb1-3GlcNAcb-Sp8	27	2	8
151	Galb1-4(Fuca1-3)GlcNAcb-Sp0	33	6	17
152	Galb1-4(Fuca1-3)GlcNAcb-Sp8	38	2	4
153	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	44	5	11
154	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	21	1	5
155	Galb1-4(6S)Glcb-Sp0	32	1	4
156	Galb1-4(6S)Glcb-Sp8	33	2	7
157	Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	24	6	25
158	Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	33	3	10
159	Galb1-4GlcNAcb1-3GalNAca-Sp8	27	2	9
160	Galb1-4GlcNAcb1-3GalNAc-Sp14	17	7	40
161	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	35	1	4
162	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	20	2	12
163	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	20	8	38
164	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	35	2	6
165	Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp8	25	1	4
166	Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp8	35	2	6
167	Galb1-4GlcNAcb1-6(Galb1-3)GalNAc-Sp14	44	3	6
168	Galb1-4GlcNAcb-Sp0	39	2	5
169	Galb1-4GlcNAcb-Sp8	27	4	15
170	Galb1-4GlcNAcb-Sp23	21	2	8
171	Galb1-4Glcb-Sp0	19	5	23
172	Galb1-4Glcb-Sp8	20	6	33
173	GlcNAca1-3Galb1-4GlcNAcb-Sp8	29	1	5
174	GlcNAca1-6Galb1-4GlcNAcb-Sp8	26	1	4
175	GlcNAcb1-2Galb1-3GalNAca-Sp8	42	3	7
176	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp8	25	4	16
177	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp14	21	1	5
178	GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAcb-Sp8	38	1	3
179	GlcNAcb1-3GalNAca-Sp8	39	1	1
180	GlcNAcb1-3GalNAca-Sp14	19	7	37
181	GlcNAcb1-3Galb-Sp8	27	5	19
182	GlcNAcb1-3Galb1-4GlcNAcb-Sp0	16	10	65
183	GlcNAcb1-3Galb1-4GlcNAcb-Sp8	23	3	15
184	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	16	3	21
185	GlcNAcb1-3Galb1-4Glcb-Sp0	23	1	6
186	GlcNAcb1-4-MDPLys	27	4	13
187	GlcNAcb1-6(GlcNAcb1-4)GalNAca-Sp8	58	1	1
188	GlcNAcb1-4Galb1-4GlcNAcb-Sp8	44	2	4
189	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	23	1	4
190	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	25	1	4
191	GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8	28	2	6
192	GlcNAcb1-6GalNAca-Sp8	68	6	9
193	GlcNAcb1-6GalNAca-Sp14	26	1	4
194	GlcNAcb1-6Galb1-4GlcNAcb-Sp8	36	1	4
195	Glca1-4Glcb-Sp8	21	1	5
196	Glca1-4Glca-Sp8	31	3	10
197	Glca1-6Glca1-6Glcb-Sp8	25	1	4
198	Glcb1-4Glcb-Sp8	25	1	5
199	Glcb1-6Glcb-Sp8	20	8	41
200	G-ol-Sp8	28	3	11
201	GlcAa-Sp8	30	2	5
202	GlcAb-Sp8	26	3	12
203	GlcAb1-3Galb-Sp8	39	2	6
204	GlcAb1-6Galb-Sp8	33	2	7
205	KDNa2-3Galb1-3GlcNAcb-Sp0	36	1	4
206	KDNa2-3Galb1-4GlcNAcb-Sp0	25	1	2
207	Mana1-2Mana1-2Mana1-3Mana-Sp9	20	8	39
208	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana-Sp9	25	2	9

209	Mana1-2Mana1-3Mana-Sp9	22	4	20
210	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	28	3	9
211	Mana1-6(Mana1-3)Mana-Sp9	39	1	2
212	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9	28	1	2
213	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	28	1	3
214	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	27	2	6
215	Manb1-4GlcNAcb-Sp0	24	3	14
216	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	22	1	6
217	(3S)Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	56	6	11
218	Fuca1-2(6S)Galb1-4GlcNAcb-Sp0	27	1	5
219	Fuca1-2Galb1-4(6S)GlcNAcb-Sp8	30	3	12
220	Fuca1-2(6S)Galb1-4(6S)Glcb-Sp0	43	2	5
221	Neu5Aca2-3Galb1-3GalNAca-Sp8	35	2	6
222	Neu5Aca2-3Galb1-3GalNAca-Sp14	30	2	6
223	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	22	1	2
224	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	31	3	8
225	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0	25	2	9
226	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	29	1	5
227	Neu5Aca2-8Neu5Aca2-8Neu5Aca-Sp8	24	1	5
228	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	39	4	10
229	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp8	18	5	27
230	GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	25	2	10
231	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0	26	2	6
232	Neu5Aca2-6(Neu5Aca2-3)GalNAca-Sp8	33	2	5
233	Neu5Aca2-3GalNAca-Sp8	44	3	7
234	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0	30	1	2
235	Neu5Aca2-3Galb1-3(6S)GlcNAc-Sp8	40	1	2
236	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp8	40	3	8
237	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	33	1	4
238	Neu5Aca2-3Galb1-4(Neu5Aca2-3Galb1-3)GlcNAcb-Sp8	26	1	4
239	Neu5Aca2-3Galb1-3(6S)GalNAca-Sp8	24	3	13
240	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp8	20	4	23
241	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	24	2	7
242	Neu5Aca2-3Galb-Sp8	25	4	17
243	Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	29	1	2
244	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	28	1	2
245	Fuca1-2(6S)Galb1-4Glcb-Sp0	48	3	7
246	Neu5Aca2-3Galb1-3GlcNAcb-Sp0	51	2	3
247	Neu5Aca2-3Galb1-4(6S)GlcNAcb-Sp8	41	3	7
248	Neu5Aca2-3Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	27	2	7
249	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	30	5	18
250	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	25	1	2
251	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	23	5	21
252	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8	27	3	12
253	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8	49	2	4
254	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	26	2	6
255	Neu5Aca2-3Galb1-4GlcNAcb-Sp0	39	1	2
256	Neu5Aca2-3Galb1-4GlcNAcb-Sp8	40	3	7
257	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	31	1	3
258	Fuca1-2Galb1-4(6S)Glcb-Sp0	31	1	4
259	Neu5Aca2-3Galb1-4Glcb-Sp0	35	4	10
260	Neu5Aca2-3Galb1-4Glcb-Sp8	21	4	18
261	Neu5Aca2-6GalNAca-Sp8	19	6	33
262	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0	14	7	48
263	Neu5Aca2-6Galb1-4(6S)GlcNAcb-Sp8	24	4	15
264	Neu5Aca2-6Galb1-4GlcNAcb-Sp0	23	4	15
265	Neu5Aca2-6Galb1-4GlcNAcb-Sp8	44	2	4
266	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	46	3	6

267	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	29	2	6
268	Neu5Aca2-6Galb1-4Glcb-Sp0	41	1	3
269	Neu5Aca2-6Galb1-4Glcb-Sp8	31	2	6
270	Neu5Aca2-6Galb-Sp8	37	2	6
271	Neu5Aca2-8Neu5Aca-Sp8	26	1	5
272	Neu5Aca2-8Neu5Aca2-3Galb1-4Glcb-Sp0	22	1	6
273	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	30	7	24
274	Neu5Acb2-6GalNAca-Sp8	22	4	19
275	Neu5Acb2-6Galb1-4GlcNAcb-Sp8	38	1	2
276	Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	29	1	3
277	Neu5Gca2-3Galb1-3GlcNAcb-Sp0	27	4	16
278	Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	37	1	4
279	Neu5Gca2-3Galb1-4GlcNAcb-Sp0	35	1	4
280	Neu5Gca2-3Galb1-4Glcb-Sp0	51	2	5
281	Neu5Gca2-6GalNAca-Sp0	42	2	5
282	Neu5Gca2-6Galb1-4GlcNAcb-Sp0	32	2	6
283	Neu5Gca-Sp8	33	3	8
284	Neu5Aca2-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	22	1	2
285	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	20	1	5
286	Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp0	74	6	8
287	Galb1-4(Fuca1-3)(6S)Glcb-Sp0	55	4	7
288	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	27	2	6
289	Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	24	2	9
290	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	21	0	0
291	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	26	2	7
292	4S(3S)Galb1-4GlcNAcb-Sp0	51	4	7
293	(6S)Galb1-4(6S)GlcNAcb-Sp0	59	3	5
294	(6P)Glcb-Sp10	23	2	6
295	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3)GalNAca-Sp14	80	4	4
296	Galb1-3Galb1-4GlcNAcb-Sp8	24	2	9
297	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	20	1	4
298	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	23	1	4
299	GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	24	3	12
300	Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0	29	3	11
301	Galb1-4GlcNAcb1-6Galb1-4GlcNAcb-Sp0	30	2	6
302	GalNAcb1-3Galb-Sp8	44	3	8
303	GlcAb1-3GlcNAcb-Sp8	36	1	3
304	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	20	1	3
305	GlcNAcb1-3Man-Sp10	31	1	3
306	GlcNAcb1-4GlcNAcb-Sp10	30	1	2
307	GlcNAcb1-4GlcNAcb-Sp12	26	1	4
308	MurNAcb1-4GlcNAcb-Sp10	23	2	8
309	Mana1-6Manb-Sp10	34	2	6
310	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb-Sp10	36	2	6
311	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	21	2	7
312	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	18	3	15
313	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	19	1	4
314	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	22	1	5
315	Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	20	1	3
316	Neu5Aca2-8Neu5Acb-Sp17	44	2	4
317	Neu5Aca2-8Neu5Aca2-8Neu5Acb-Sp8	27	2	9
318	Neu5Gcb2-6Galb1-4GlcNAc-Sp8	54	2	4
319	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	59	3	5
320	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	17	1	7
321	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	16	4	24

322	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	18	2	11
323	Neu5,9Ac2a2-3Galb1-3GlcNAcb-Sp0	23	2	10
324	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	24	1	5
325	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	28	2	8
326	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	23	1	4
327	Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	23	1	5
328	GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	21	2	7
329	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	20	1	2
330	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	26	3	13
331	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAc-Sp14	29	1	3
332	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	19	2	8
333	GlcNAca1-4Galb1-4GlcNAcb-Sp0	22	5	22
334	GlcNAca1-4Galb1-3GlcNAcb-Sp0	31	2	7
335	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glcb-Sp0	25	2	6
336	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	65	3	4
337	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	28	2	8
338	GlcNAca1-4Galb1-3GalNAc-Sp14	22	2	9
339	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	22	1	6
340	Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	19	4	21
341	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	19	1	3
342	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	19	1	3
343	Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	20	2	10
344	Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	17	2	12
345	Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	24	3	11
346	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	29	2	5
347	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	28	2	9
348	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	26	1	4
349	(6S)GlcNAcb1-3Galb1-4GlcNAcb-Sp0	35	2	5
350	KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0	29	2	6
351	KDNa2-6Galb1-4GlcNAc-Sp0	28	1	5
352	KDNa2-3Galb1-4Glc-Sp0	25	2	7
353	KDNa2-3Galb1-3GalNAca-Sp14	31	2	5
354	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	45	3	6
355	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	36	3	7
356	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAb-Sp20	54	5	10
357	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	36	1	3
358	Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	26	1	2
359	Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-6(Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	51	6	12
360	Neu5Aca2-6GlcNAcb1-4GlcNAc-Sp21	31	2	6
361	Neu5Aca2-6GlcNAcb1-4GlcNAcb1-4GlcNAc-Sp21	24	1	2
362	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	27	1	4
363	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	24	1	6
364	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	28	1	2
365	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	28	1	3
366	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	43	3	7
367	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	23	1	5
368	Galα1-3(Fucα1-2)Galβ1-3GlcNAcβ1-2Manα1-6(Galα1-3(Fucα1-2)Galβ1-3GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp20	24	3	11

369	Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	34	5	13
370	Neu5Aca2-3Galb1-4GlcNAcb1-3GalNAc-Sp14	11	2	21
371	Neu5Aca2-6Galb1-4GlcNAcb1-3GalNAc-Sp14	19	4	24
372	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	41	3	6
373	GalNAcb1-4GlcNAcb1-2Mana1-6(GalNAcb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	41	4	9
374	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0	16	2	9
375	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	12	1	7
376	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp0	16	1	8
377	Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	15	4	27
378	Galb1-4GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	22	2	9
379	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	20	2	8
380	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	15	2	11
381	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	17	2	14
382	GlcNAcb1-2Mana1-6(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	14	1	4
383	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	20	5	26
384	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	7	3	37
385	Galb1-3GlcNAcb1-3GalNAca-Sp14	17	6	34
386	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb1-3GalNAca-Sp14	17	5	31
387	GalNAca1-3(Fuca1-2)Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	12	4	32
388	Gala1-3Galb1-3GlcNAcb1-2Mana1-6(Gala1-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	38	3	9
389	Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	59	2	3
390	GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	12	4	39
391	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	15	3	21
392	Neu5Aca2-3Galb1-3GlcNAcb1-3GalNAca-Sp14	15	1	8
393	Fuca1-2Galb1-4GlcNAcb1-3GalNAca-Sp14	21	6	26
394	Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	21	2	8
395	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	14	3	25
396	Gala1-4Galb1-3GlcNAcb1-2Mana1-6(Gala1-4Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	29	3	10
397	Gala1-4Galb1-4GlcNAcb1-2Mana1-6(Gala1-4Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	65	4	6
398	Gala1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	13	1	8
399	Galb1-3GlcNAcb1-6Galb1-4GlcNAcb-Sp0	23	2	8
400	Galb1-3GlcNAca1-6Galb1-4GlcNAcb-Sp0	14	5	34
401	GalNAcb1-3Gala1-6Galb1-4Glcb-Sp8	24	3	13
402	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)Glcb-Sp21	19	1	4
403	Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	12	5	45
404	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	34	7	22
405	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	21	5	24
406	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	12	3	27
407	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	7	2	27
408	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0	22	2	8
409	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	31	2	8
410	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	18	3	18
411	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	26	3	12
412	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	51	3	5
413	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	24	4	15
414	GlcNAcb1-2(GlcNAcb1-6)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	45	1	3
415	Fuca1-2Galb1-3GlcNAcb1-3GalNAc-Sp14	17	6	33
416	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	13	4	28
417	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	19	5	26
418	Gala1-3Galb1-3GlcNAcb1-3GalNAc-Sp14	20	3	16

419	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	30	5	16
420	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	29	3	10
421	Galb1-3GlcNAcb1-6(Galb1-3GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	32	5	15
422	Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	15	1	8
423	Fuca1-3GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	16	2	11
424	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	11	4	39
425	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	5	7	147
426	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	16	2	11
427	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	13	4	31
428	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	23	17	74
429	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	11	3	27
430	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	12	2	17
431	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	14	1	7
432	Galb1-4Galb-Sp10	18	5	25
433	Galb1-6Galb-Sp10	19	10	52
434	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb-Sp8	22	3	13
435	GalNAcb1-6GalNAcb-Sp8	17	5	29
436	(6S)Galb1-3GlcNAcb-Sp0	32	10	32
437	(6S)Galb1-3(6S)GlcNAc-Sp0	30	3	10
438	Fuca1-2Galb1-4 GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2(Fuca1-2Galb1-4GlcNAcb1-4)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	31	4	11
439	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-4(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	48	6	13
440	Galb1-4(Fuca1-3)GlcNAcb1-6GalNAc-Sp14	35	3	7
441	Galb1-4GlcNAcb1-2Mana-Sp0	28	2	6
442	Fuca1-2Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)GalNAc-Sp14	17	1	5
443	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	16	4	26
444	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	13	3	26
445	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp0	67	7	10
446	GalNAcb1-4Galb1-4Glcb-Sp0	30	5	15
447	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	34	4	10
448	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	22	2	11
449	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	18	1	5
450	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	25	1	2
451	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	23	2	8
452	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	15	2	12
453	Neu5Aca2-3Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	12	1	12
454	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	14	2	12
455	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	12	6	47

456	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	13	2	17
457	Neu5Aca2-6Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	14	4	26
458	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	17	1	6
459	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	14	2	14
460	Gala1-3(Fuca1-2)Galb1-3GalNAca-Sp8	26	3	10
461	Gala1-3(Fuca1-2)Galb1-3GalNAcb-Sp8	49	5	10
462	Glca1-6Glca1-6Glca1-6Glcb-Sp10	20	6	30
463	Glca1-4Glca1-4Glca1-4Glcb-Sp10	30	2	7
464	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	15	2	13
465	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	64	8	12
466	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	53	4	7
467	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	61	7	12
468	Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	38	2	6
469	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glcb-Sp21	15	1	9
470	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana-Sp0	40	7	16
471	Neu5Aca2-3Galb1-4GlcNAcb1-6GalNAca-Sp14	11	4	42
472	Neu5Aca2-6Galb1-4GlcNAcb1-6GalNAca-Sp14	25	6	24
473	Neu5Aca2-6Galb1-4 GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3)GalNAca-Sp14	17	1	8
474	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	45	1	2
475	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	50	1	3
476	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	40	3	8
477	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	59	6	10
478	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	12	4	37
479	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	15	6	41
480	Galb1-3GlcNAcb1-6GalNAca-Sp14	12	6	48
481	Gala1-3Galb1-3GlcNAcb1-6GalNAca-Sp14	11	3	30
482	Galb1-3(Fuca1-4)GlcNAcb1-6GalNAca-Sp14	32	9	27
483	Neu5Aca2-3Galb1-3GlcNAcb1-6GalNAca-Sp14	21	1	6
484	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp0	29	7	23
485	Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-6(Neu5Aca2-3Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	27	2	7
486	Fuca1-2Galb1-4GlcNAcb1-6GalNAca-Sp14	26	2	9
487	Gala1-3Galb1-4GlcNAcb1-6GalNAca-Sp14	11	3	29
488	Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	38	4	10
489	Fuca1-2(6S)Galb1-3GlcNAcb-Sp0	13	3	27
490	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAca-Sp14	20	5	27
491	Fuca1-2Galb1-4GlcNAcb1-2Mana-Sp0	14	8	54
492	Fuca1-2Galb1-3(6S)GlcNAcb-Sp0	31	4	12
493	Fuca1-2(6S)Galb1-3(6S)GlcNAcb-Sp0	37	11	31
494	Neu5Aca2-6GalNAcb1-4(6S)GlcNAcb-Sp8	21	6	29
495	GalNAcb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	23	4	19
496	(3S)GalNAcb1-4(Fuca1-3)GlcNAcb-Sp8	21	6	29
497	Fuca1-2Galb1-3GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)GalNAca-Sp14	25	2	7
498	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAca-Sp14	14	2	11
499	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	20	2	11
500	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	13	2	19
501	Galb1-3GlcNAca1-3Galb1-4GlcNAcb-Sp8	19	4	19

502	Galb1-3(6S)GlcNAcb-Sp8	23	8	34
503	(6S)(4S)GalNAcb1-4GlcNAc-Sp8	20	5	25
504	(6S)GalNAcb1-4GlcNAc-Sp8	11	5	49
505	(3S)GalNAcb1-4(3S)GlcNAc-Sp8	29	7	25
506	GalNAcb1-4(6S)GlcNAc-Sp8	34	2	4
507	(3S)GalNAcb1-4GlcNAc-Sp8	39	2	5
508	(4S)GalNAcb-Sp10	24	2	6
509	Galb1-4(6P)GlcNAcb-Sp0	16	1	6
510	(6P)Galb1-4GlcNAcb-SP0	8	3	34
511	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAc-Sp14	12	2	16
512	Neu5Aca2-6Galb1-4GlcNAcb1-2Man-Sp0	15	2	14
513	Gala1-3Galb1-4GlcNAcb1-2Mana-Sp0	17	3	16
514	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana-Sp0	13	7	56
515	GalNAca1-3(Fuca1-2)Galb1-4 GlcNAcb1-2Mana-Sp0	11	3	23
516	Galb1-3GlcNAcb1-2Mana-Sp0	35	2	7
517	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAc-Sp14	12	4	37
518	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana-Sp0	12	2	12
519	Gala1-3Galb1-3GlcNAcb1-2Mana-Sp0	14	1	4
520	GalNAcb1-4GlcNAcb1-2Mana-Sp0	18	2	12
521	Neu5Aca2-3Galb1-3GalNAcb1-4Galb1-4Glcb-Sp0	10	3	33
522	GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	12	3	28
523	Galb1-4GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	14	3	20
524	Galb1-4GlcNAcb1-2 Mana1-6(Galb1-4GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	11	3	31
525	Fuca1-4(Galb1-3)GlcNAcb1-2 Mana-Sp0	44	9	19
526	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	12	2	17
527	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAc-Sp0	12	3	22
528	GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	17	1	6
529	Gala1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	17	3	17
530	Galb1-3GalNAcb1-3Gal-Sp21	50	2	3
531	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	51	10	19
532	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	12	7	63
533	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-6(Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Sp12	8	4	46
534	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	57	4	6
535	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	32	5	16
536	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	6	2	39
537	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	40	4	9
538	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp25	25	4	17
539	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAc-Sp0	23	2	8
540	Neu5Aca2-8Neu5Gca2-3Galb1-4GlcNAc-Sp0	7	3	46
541	Neu5Gca2-8Neu5Aca2-3Galb1-4GlcNAc-Sp0	18	3	16
542	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAc-Sp0	15	1	9
543	Neu5Gca2-8Neu5Gca2-6Galb1-4GlcNAc-Sp0	19	1	3
544	Neu5Aca2-8Neu5Aca2-3Galb1-4GlcNAc-Sp0	7	2	25
545	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Man a1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp24	66	12	18
546	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Mana1-4GlcNAcb1-4GlcNAc-Sp24	38	8	22
547	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp24	52	7	14
548	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-3)GalNAca-Sp14	17	2	12
549	GalNAcb1-3GlcNAcb-Sp0	14	4	32

550	GalNAcb1-4GlcNAcb1-3GalNAcb1-4GlcNAcb-Sp0	17	3	16
551	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	39	5	14
552	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	42	8	19
553	GlcNAb1-3Galb1-3GalNAc-Sp14	8	5	64
554	Galb1-3GlcNAcb1-6(Galb1-3)GalNAc-Sp14	17	5	29
555	(3S)GlcAb1-3Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	11	3	23
556	(3S)GlcAb1-3Galb1-4GlcNAcb1-2Mana-Sp0	17	4	25
557	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	36	11	29
558	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	39	9	23
559	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp21	21	1	7
560	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	35	5	15
561	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	56	9	16
562	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	57	6	11
563	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	14	4	29
564	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	19	4	19
565	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	15	2	12
566	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	25	3	11
567	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	15	3	20
568	GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	15	2	13
569	GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	17	1	8
570	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	21	1	3
571	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	21	2	7
572	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	10	5	53
573	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	12	4	32
574	Galb1-4GlcNAcb1-3Galb1-3GalNAca-Sp14	5	2	43
575	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	16	5	32
576	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	17	2	10
577	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	15	1	5
578	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	24	4	18
579	GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	11	2	18
580	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	18	5	25
581	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	196	19	9
582	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	58	2	4
583	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	23	2	7
584	GlcNAcb1-3Fuca-Sp21	23	1	4
585	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glcb-Sp21	18	1	8

8. 7. Appendix VII: PDBe Fold search for Awp3A

#	Q-score	P-score	Z-score	RMSD	Nalgn	Nsse	Ngaps	Seq-%	Nmd	Nres-Q	Nsse-Q	Nres-T	Nsse-T	Query Target		
1	0.1998	0	8.19	2.971	196	19	16	0.09184	0	328	31	296	31	Awp3A	PDB	4i84:A
2	0.1794	0	5.957	3.085	174	19	12	0.1092	0	328	31	250	27	Awp3A	PDB	4peu:A
3	0.1718	0	8.444	2.739	183	19	21	0.07104	0	328	31	324	30	Awp3A	PDB	3vmv:A
4	0.1706	0	7.793	2.826	185	16	22	0.06486	0	328	31	324	28	Awp3A	PDB	3vmw:A
5	0.1688	0	5.167	3.286	202	16	30	0.1188	0	328	31	335	31	Awp3A	PDB	1czf:B
6	0.1649	0	5.677	3.372	191	21	23	0.1257	0	328	31	298	35	Awp3A	PDB	6kfn:A
7	0.1585	0	7.038	3.321	186	17	18	0.09677	0	328	31	299	29	Awp3A	PDB	5c1c:A
8	0.156	0	7.761	2.766	182	19	19	0.1099	0	328	31	350	31	Awp3A	PDB	1jrg:B
9	0.1554	0	7.706	2.88	188	18	18	0.117	0	328	31	361	34	Awp3A	PDB	1ooc:B
10	0.1549	0	7.485	2.952	190	18	17	0.1158	0	328	31	361	36	Awp3A	PDB	1ooc:A
11	0.1539	0	7.734	2.908	188	18	19	0.1223	0	328	31	361	35	Awp3A	PDB	1jta:A
12	0.1535	0	5.565	3.043	185	16	22	0.1189	0	328	31	335	25	Awp3A	PDB	1czf:A
13	0.1531	0	7.346	3.327	183	18	19	0.0929	0	328	31	299	29	Awp3A	PDB	5c1e:A
14	0.153	0	7.954	2.828	185	18	19	0.0973	0	328	31	361	34	Awp3A	PDB	1pe9:A
15	0.1522	0	8.063	2.747	182	18	20	0.1099	0	328	31	361	35	Awp3A	PDB	1pe9:B
16	0.1512	0	6.189	2.946	187	17	18	0.08556	0	328	31	359	31	Awp3A	PDB	1qcx:A
17	0.1508	0	6.158	2.857	184	17	18	0.07609	0	328	31	359	32	Awp3A	PDB	1idk:A
18	0.1473	0	7.151	2.865	196	16	19	0.08163	0	328	31	416	39	Awp3A	PDB	1vbl:A
19	0.1426	0	6.604	2.641	182	17	20	0.08791	0	328	31	399	35	Awp3A	PDB	2o04:A
20	0.1384	0	5.966	3.321	204	19	24	0.1176	0	328	31	412	42	Awp3A	PDB	5zkw:C
21	0.1382	0	6.114	3.288	204	19	25	0.1127	0	328	31	417	42	Awp3A	PDB	5zkw:A
22	0.1366	0	6.055	3.273	203	19	25	0.1133	0	328	31	420	42	Awp3A	PDB	5zkw:F
23	0.1365	0	6.085	3.289	204	19	24	0.1127	0	328	31	422	42	Awp3A	PDB	5zkw:B
24	0.1365	0	6.245	2.805	183	16	20	0.08743	0	328	31	399	35	Awp3A	PDB	2nzm:A
25	0.136	0	4.574	3.683	185	16	22	0.1405	1	328	31	306	29	Awp3A	PDB	5nxk:A

#	Q-score	P-score	Z-score	RMSD	Nalgn	Nsse	Ngaps	Seq-%	Nmd	Nres-Q	Nsse-Q	Nres-T	Nsse-T	Query Target		
26	0.1348	0	3.971	3.541	162	19	13	0.1296	0	328	31	248	24	Awp3A	PDB	4phb:A
27	0.1347	0	4.484	3.723	185	16	22	0.1514	1	328	31	305	30	Awp3A	PDB	5nxk:B
28	0.1331	0	5.95	2.779	169	16	19	0.1243	0	328	31	352	29	Awp3A	PDB	1plu:A
29	0.1322	0	5.769	2.878	169	17	19	0.1124	0	328	31	343	33	Awp3A	PDB	5ny0:A
30	0.1316	0	6.055	3.273	203	19	25	0.1133	0	328	31	436	42	Awp3A	PDB	5zkw:E
31	0.1311	0	6.114	3.269	203	19	25	0.1133	0	328	31	438	42	Awp3A	PDB	5zkw:D
32	0.1309	0	5.996	3.26	203	19	23	0.1133	0	328	31	440	42	Awp3A	PDB	5zks:A
33	0.1284	0	6.468	3.245	201	19	28	0.0995	0	328	31	442	41	Awp3A	PDB	5zku:B
34	0.1281	0	6.055	3.247	201	19	25	0.1144	0	328	31	443	42	Awp3A	PDB	5zku:F
35	0.1275	0	6.144	3.207	199	19	25	0.1156	0	328	31	442	42	Awp3A	PDB	5zku:C
36	0.1266	0	6.328	3.251	184	17	21	0.07065	0	328	31	375	34	Awp3A	PDB	6fi2:A
37	0.1263	0	5.996	3.235	199	19	25	0.1156	0	328	31	442	41	Awp3A	PDB	5zku:A
38	0.1251	0	5.812	3.264	199	18	26	0.1156	0	328	31	442	41	Awp3A	PDB	5zku:D
39	0.1239	0	5.985	3.201	196	18	24	0.1173	0	328	31	442	40	Awp3A	PDB	5zku:E
40	0.121	0	6.898	3.246	179	19	19	0.1285	0	328	31	372	38	Awp3A	PDB	2odl:A
41	0.1204	0	6.503	3.519	176	16	25	0.1477	0	328	31	330	28	Awp3A	PDB	2qxz:A
42	0.1186	0	6.663	3.463	173	16	23	0.1387	0	328	31	330	28	Awp3A	PDB	2qxz:B
43	0.1183	0	6.827	2.782	179	16	22	0.09497	0	328	31	444	38	Awp3A	PDB	3jur:A
44	0.1178	0	5.333	3.403	188	16	30	0.09043	0	328	31	400	32	Awp3A	PDB	1ru4:A
45	0.1171	0	2.825	3.024	83	10	11	0.04819	0	328	31	89	13	Awp3A	PDB	5jmc:H
46	0.1147	0	3.145	4.242	218	19	29	0.09174	0	328	31	421	37	Awp3A	PDB	3zpp:A
47	0.1144	0	6.803	2.782	176	16	23	0.09659	0	328	31	444	38	Awp3A	PDB	3jur:C
48	0.114	0	6.315	2.87	149	17	16	0.1074	0	328	31	310	33	Awp3A	PDB	5nxk:C
49	0.1107	0	5.205	3.08	162	16	16	0.1235	0	328	31	352	30	Awp3A	PDB	1air:A

#	Q-score	P-score	Z-score	RMSD	Nalgn	Nsse	Ngaps	Seq-%	Nmd	Nres-Q	Nsse-Q	Nres-T	Nsse-T	Query Target		
50	0.108	0	5.686	3.04	159	16	19	0.1195	0	328	31	352	29	Awp3A	PDB	2pec:A
51	0.1079	0	2.27	3.203	83	10	12	0.04819	0	328	31	91	13	Awp3A	PDB	5jmc:B
52	0.1036	0	4.446	4.192	182	16	19	0.1319	0	328	31	330	27	Awp3A	PDB	2qx3:A
53	0.1026	0	4.281	3.847	177	16	25	0.1412	0	328	31	352	30	Awp3A	PDB	2ewe:A
54	0.1018	0	1.965	4.292	213	20	30	0.1268	0	328	31	446	39	Awp3A	PDB	4mr0:B
55	0.1008	0	5.205	3.525	165	18	18	0.07273	0	328	31	346	36	Awp3A	PDB	1pxz:B
56	0.09956	0	5.232	3.525	164	18	18	0.07317	0	328	31	346	35	Awp3A	PDB	1pxz:A
57	0.09752	0	3.94	3.539	204	14	22	0.05882	0	328	31	544	42	Awp3A	PDB	6e1r:D
58	0.0958	0	4.482	3.776	210	13	23	0.09048	0	328	31	543	43	Awp3A	PDB	6e1r:B
59	0.09315	0	5.489	3.55	131	16	14	0.1145	0	328	31	234	27	Awp3A	PDB	4w8q:A
60	0.09297	0	4.882	3.727	137	16	15	0.1095	0	328	31	242	27	Awp3A	PDB	5keh:A
61	0.0919	0	5.028	3.414	194	15	26	0.0567	0	328	31	544	43	Awp3A	PDB	6e1r:A
62	0.09167	0	4.86	3.515	197	15	25	0.07107	0	328	31	544	43	Awp3A	PDB	6e1r:C
63	0.09088	0	5.222	3.479	195	14	27	0.1282	0	328	31	544	41	Awp3A	PDB	6e1r:F
64	0.0876	0	4.252	3.443	197	22	21	0.1117	0	328	31	583	44	Awp3A	PDB	5gqc:H
65	0.08326	0	2.909	3.698	203	17	26	0.07882	0	328	31	599	42	Awp3A	PDB	6g0x:A
66	0.08325	0	2.914	4.088	206	13	22	0.06311	0	328	31	544	40	Awp3A	PDB	6e1r:E
67	0.0814	0	2.552	3.67	199	17	28	0.0804	0	328	31	594	42	Awp3A	PDB	6gvp:A
68	0.06979	0	3.556	4.117	195	10	29	0.1231	0	328	31	576	33	Awp3A	PDB	5zru:A
69	0.05537	0	2.833	4.048	154	11	23	0.1039	0	328	31	463	32	Awp3A	PDB	6ixx:A
70	0.04982	0	4.49	2.785	102	11	12	0.05882	0	328	31	342	29	Awp3A	PDB	6oq6:A
71	0.04859	0	2.521	4.039	145	11	23	0.07586	0	328	31	469	34	Awp3A	PDB	5d7w:A
72	0.04606	0	2.379	4.97	183	11	27	0.04372	0	328	31	592	36	Awp3A	PDB	5ijc:A
73	0.04368	0	7.019	2.719	114	15	13	0.09649	0	328	31	498	34	Awp3A	PDB	6qvi:B
74	0.02459	-0	1834	6.146	139	10	29	0.04317	0	328	31	461	24	Awp3A	PDB	4kng:A

8. 8. Appendix VIII: Predicted glycosylation sites in Awp1

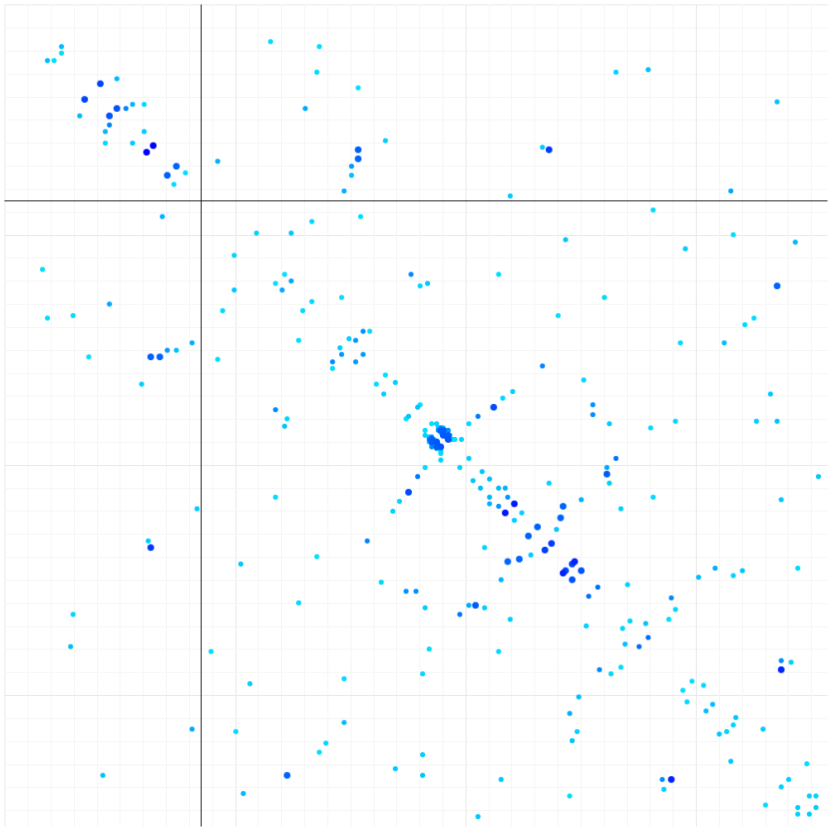
```
##gff-version 2
##source-version NetOGlyc 4.0.0.13
##date 20-2-14
##Type Protein
#seqname  source    feature   start     end       score     strand    frame  comment
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  2         2         0.0474092 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  5         5         0.0206563 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  14        14        0.0349508 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  18        18        0.0947298 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  23        23        0.0324545 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  25        25        0.334385  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  26        26        0.10867   .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  28        28        0.0960115 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  32        32        0.0572995 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  38        38        0.0899525 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  41        41        0.0507403 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  42        42        0.0752175 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  44        44        0.0555112 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  49        49        0.0264627 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  62        62        0.270531  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  63        63        0.215746  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  65        65        0.111394  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  67        67        0.20345   .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  75        75        0.182778  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  78        78        0.112795  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  80        80        0.231035  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  84        84        0.167706  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  85        85        0.16876   .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  86        86        0.198553  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  88        88        0.306797  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  90        90        0.162227  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  92        92        0.135444  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  96        96        0.370059  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  98        98        0.225932  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  100       100       0.233292  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  103       103       0.190258  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  104       104       0.105821  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  111       111       0.130017  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  118       118       0.0658696 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  120       120       0.0429707 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  122       122       0.025517  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  124       124       0.0253782 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  128       128       0.158607  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  130       130       0.101523  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  132       132       0.0958882 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  133       133       0.0555037 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  134       134       0.0655286 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  135       135       0.0834856 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  142       142       0.0902631 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  146       146       0.0938189 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  148       148       0.0897877 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  153       153       0.12575   .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  156       156       0.108738  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  158       158       0.176642  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  162       162       0.397331  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  164       164       0.147972  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  169       169       0.244889  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  172       172       0.413194  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  173       173       0.331418  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  176       176       0.101528  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  182       182       0.147715  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  190       190       0.0815793 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  196       196       0.0308309 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  198       198       0.0326156 .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  201       201       0.103725  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  203       203       0.196986  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  222       222       0.165715  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  226       226       0.146127  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  227       227       0.352403  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  228       228       0.166187  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  229       229       0.117982  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  235       235       0.512886  .      .  #POSITIVE
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  238       238       0.291915  .      .
SEQUENCE  netOGlyc-4.0.0.13  CARBOHYD  254       254       0.600399  .      .  #POSITIVE
```

SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	257	257	0.350422	.	.	
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	258	258	0.680968	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	260	260	0.462802	.	.	
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	262	262	0.703792	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	265	265	0.632744	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	267	267	0.792522	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	271	271	0.537528	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	273	273	0.916835	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	274	274	0.8811	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	279	279	0.332395	.	.	
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	292	292	0.5	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	297	297	0.539974	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	299	299	0.690481	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	307	307	0.468967	.	.	
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	318	318	0.806975	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	321	321	0.682005	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	326	326	0.966148	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	327	327	0.918819	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	328	328	0.95372	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	330	330	0.931663	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	332	332	0.979644	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	333	333	0.948497	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	334	334	0.975229	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	338	338	0.986949	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	339	339	0.917258	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	340	340	0.965553	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	344	344	0.982516	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	345	345	0.888577	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	346	346	0.937962	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	350	350	0.982434	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	351	351	0.893938	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	352	352	0.940061	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	356	356	0.983089	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	357	357	0.891346	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	358	358	0.935663	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	362	362	0.980239	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	363	363	0.890119	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	364	364	0.93364	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	368	368	0.981892	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	369	369	0.886518	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	370	370	0.929268	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	374	374	0.981733	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	375	375	0.871627	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	376	376	0.928963	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	380	380	0.980455	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	381	381	0.878075	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	382	382	0.929186	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	386	386	0.981266	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	387	387	0.881276	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	388	388	0.929819	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	392	392	0.979766	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	393	393	0.878199	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	394	394	0.928535	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	398	398	0.978503	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	399	399	0.869696	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	400	400	0.925009	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	404	404	0.980148	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	405	405	0.873972	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	406	406	0.923783	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	410	410	0.980493	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	411	411	0.880639	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	412	412	0.929349	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	416	416	0.979767	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	417	417	0.880406	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	418	418	0.924733	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	422	422	0.979832	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	423	423	0.884471	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	424	424	0.923612	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	428	428	0.980608	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	429	429	0.882493	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	430	430	0.923458	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	434	434	0.978974	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	435	435	0.868393	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	436	436	0.923986	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	440	440	0.980005	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	441	441	0.87255	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	442	442	0.922715	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	446	446	0.969655	.	.	#POSITIVE

SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	447	447	0.868473	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	448	448	0.947537	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	452	452	0.976897	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	453	453	0.954395	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	458	458	0.960711	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	459	459	0.878447	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	460	460	0.937865	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	464	464	0.973509	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	465	465	0.93421	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	466	466	0.97519	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	472	472	0.903062	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	476	476	0.964996	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	477	477	0.869721	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	478	478	0.95073	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	481	481	0.956826	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	483	483	0.939565	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	487	487	0.97397	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	488	488	0.928628	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	489	489	0.973262	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	490	490	0.950894	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	493	493	0.990194	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	494	494	0.976391	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	496	496	0.975136	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	499	499	0.987951	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	500	500	0.988891	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	501	501	0.992027	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	503	503	0.991491	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	505	505	0.994624	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	506	506	0.984542	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	508	508	0.966405	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	511	511	0.98689	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	512	512	0.981603	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	513	513	0.99213	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	515	515	0.991315	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	517	517	0.994951	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	518	518	0.984159	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	520	520	0.971848	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	523	523	0.972378	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	524	524	0.969729	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	525	525	0.990149	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	527	527	0.97241	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	529	529	0.987762	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	530	530	0.968146	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	532	532	0.923461	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	533	533	0.975793	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	535	535	0.963124	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	536	536	0.972926	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	537	537	0.976463	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	541	541	0.879796	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	542	542	0.853951	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	546	546	0.82925	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	549	549	0.94059	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	550	550	0.933681	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	552	552	0.883382	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	557	557	0.990777	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	558	558	0.946093	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	559	559	0.953479	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	564	564	0.959828	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	565	565	0.9663	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	566	566	0.968567	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	648	648	0.956568	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	652	652	0.92966	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	656	656	0.980694	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	663	663	0.950719	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	665	665	0.901281	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	671	671	0.9297	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	673	673	0.9215	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	674	674	0.968207	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	675	675	0.939822	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	677	677	0.939957	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	684	684	0.922094	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	689	689	0.935541	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	694	694	0.921681	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	704	704	0.953968	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	714	714	0.93509	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	727	727	0.904215	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	729	729	0.93806	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	732	732	0.894278	.	.	#POSITIVE

SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	738	738	0.874486	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	745	745	0.88941	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	751	751	0.982921	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	753	753	0.851239	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	755	755	0.932913	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	765	765	0.952878	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	767	767	0.898677	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	769	769	0.950579	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	771	771	0.95853	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	773	773	0.878874	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	779	779	0.899936	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	782	782	0.96499	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	793	793	0.852128	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	795	795	0.893325	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	798	798	0.973699	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	799	799	0.938209	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	803	803	0.985727	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	810	810	0.953561	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	811	811	0.95624	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	813	813	0.973935	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	815	815	0.989972	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	822	822	0.979175	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	824	824	0.97369	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	825	825	0.988112	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	826	826	0.986838	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	829	829	0.991146	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	830	830	0.99374	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	836	836	0.970364	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	840	840	0.93823	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	844	844	0.821688	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	845	845	0.636551	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	851	851	0.0973235	.	.	

8. 9. Appendix IX: *GREMLIN* contact map and residue-residue interactions



i	j	gene	i_id	j_id	r_sco	s_sco	prob	I_prob
61	64	A	61_T	64_E	0.132	3.022	0.999	N/A
217	221	B	132_F	136_F	0.118	2.696	0.996	N/A
242	247	B	157_G	162_L	0.114	2.597	0.995	N/A
289	337	B	204_V	252_L	0.113	2.589	0.995	N/A
63	236	AB	63_K	151_H	0.103	2.339	0.986	0.954
234	237	B	149_Q	152_D	0.101	2.31	0.985	N/A
34	41	A	34_C	41_C	0.097	2.209	0.978	N/A
243	246	B	158_N	161_I	0.097	2.205	0.978	N/A
175	212	B	90_F	127_N	0.094	2.154	0.973	N/A
45	48	A	45_S	48_L	0.092	2.108	0.969	N/A
189	192	B	104_F	107_K	0.091	2.084	0.966	N/A
246	250	B	161_I	165_V	0.091	2.066	0.963	N/A
187	190	B	102_R	105_P	0.089	2.029	0.958	N/A
204	261	B	119_T	176_W	0.088	2.007	0.955	N/A
63	153	AB	63_K	68_F	0.087	1.994	0.953	0.873
185	190	B	100_Y	105_P	0.087	1.987	0.952	N/A
187	192	B	102_R	107_K	0.087	1.976	0.950	N/A
185	189	B	100_Y	104_F	0.086	1.963	0.947	N/A
70	74	A	70_N	74_Y	0.085	1.937	0.942	N/A
227	231	B	142_G	146_T	0.085	1.936	0.942	N/A
122	335	B	37_N	250_P	0.085	1.931	0.941	N/A
67	153	AB	67_T	68_F	0.085	1.93	0.941	0.847
223	241	B	138_C	156_C	0.084	1.916	0.938	N/A
218	242	B	133_A	157_G	0.084	1.911	0.937	N/A
275	279	B	190_L	194_K	0.081	1.852	0.923	N/A
253	257	B	168_A	172_A	0.08	1.833	0.918	N/A
197	265	B	112_L	180_L	0.078	1.782	0.902	N/A
184	189	B	99_F	104_F	0.077	1.752	0.892	N/A
179	205	B	94_G	120_L	0.076	1.735	0.886	N/A
45	52	A	45_S	52_A	0.073	1.658	0.854	N/A
157	233	B	72_T	148_W	0.072	1.65	0.850	N/A
258	289	B	173_Y	204_V	0.072	1.646	0.848	N/A
117	176	B	32_W	91_F	0.072	1.633	0.842	N/A
187	191	B	102_R	106_A	0.071	1.617	0.834	N/A
185	192	B	100_Y	107_K	0.07	1.594	0.821	N/A

142	155	B	57_L	70_D	0.07	1.593	0.821	N/A
178	255	B	93_R	170_G	0.07	1.588	0.818	N/A
285	337	B	200_M	252_L	0.069	1.585	0.816	N/A
174	255	B	89_R	170_G	0.069	1.565	0.805	N/A
185	188	B	100_Y	103_V	0.068	1.548	0.795	N/A
214	218	B	129_S	133_A	0.068	1.545	0.793	N/A
146	152	B	61_K	67_P	0.066	1.514	0.774	N/A
70	150	AB	70_N	65_F	0.066	1.498	0.763	0.536
152	155	B	67_P	70_D	0.065	1.494	0.761	N/A
188	192	B	103_V	107_K	0.064	1.46	0.737	N/A
120	124	B	35_L	39_F	0.062	1.423	0.709	N/A
45	130	AB	45_S	45_A	0.062	1.42	0.707	0.46
81	315	AB	81_K	230_K	0.062	1.413	0.702	0.453
201	261	B	116_I	176_W	0.062	1.407	0.697	N/A
245	308	B	160_N	223_T	0.061	1.386	0.680	N/A
81	147	AB	81_K	62_D	0.06	1.371	0.668	0.412
68	92	AB	68_T	7_T	0.06	1.362	0.661	0.403
215	250	B	130_F	165_V	0.059	1.351	0.652	N/A
210	217	B	125_V	132_F	0.059	1.35	0.651	N/A
103	343	B	18_L	258_L	0.058	1.332	0.635	N/A
210	214	B	125_V	129_S	0.058	1.323	0.628	N/A
43	55	A	43_C	55_C	0.057	1.31	0.616	N/A
32	48	A	32_S	48_L	0.057	1.308	0.615	N/A
28	279	AB	28_A	194_K	0.057	1.289	0.598	0.333
147	312	B	62_D	227_N	0.056	1.288	0.597	N/A
249	301	B	164_W	216_L	0.056	1.287	0.596	N/A
206	210	B	121_V	125_V	0.056	1.284	0.593	N/A
304	307	B	219_I	222_F	0.056	1.278	0.588	N/A
18	24	A	18_P	24_C	0.056	1.273	0.584	N/A
74	150	AB	74_Y	65_F	0.055	1.265	0.576	0.311
269	278	B	184_A	193_K	0.054	1.242	0.555	N/A
187	193	B	102_R	108_S	0.054	1.242	0.555	N/A
42	335	AB	42_I	250_P	0.054	1.24	0.554	0.289
175	179	B	90_F	94_G	0.054	1.237	0.551	N/A
246	320	B	161_I	235_L	0.054	1.232	0.546	N/A
181	335	B	96_I	250_P	0.054	1.225	0.540	N/A
203	207	B	118_W	122_F	0.054	1.221	0.536	N/A
121	183	B	36_S	98_L	0.053	1.218	0.533	N/A
102	243	B	17_I	158_N	0.053	1.214	0.530	N/A
99	124	B	14_I	39_F	0.053	1.213	0.529	N/A
310	317	B	225_T	232_I	0.053	1.211	0.527	N/A
184	188	B	99_F	103_V	0.053	1.211	0.527	N/A
188	191	B	103_V	106_A	0.053	1.204	0.520	N/A
215	337	B	130_F	252_L	0.053	1.201	0.518	N/A
228	239	B	143_L	154_G	0.052	1.197	0.514	N/A
205	353	B	120_L	268_S	0.052	1.197	0.514	N/A
169	332	B	84_L	247_V	0.052	1.19	0.508	N/A
315	329	B	230_K	244_D	0.052	1.189	0.507	N/A
83	219	AB	83_D	134_V	0.052	1.176	0.495	0.236
349	352	B	264_P	267_L	0.051	1.174	0.493	N/A
55	60	A	55_C	60_C	0.051	1.174	0.493	N/A
62	233	AB	62_I	148_W	0.051	1.173	0.492	0.233
184	190	B	99_F	105_P	0.051	1.171	0.490	N/A
182	262	B	97_I	177_L	0.051	1.167	0.486	N/A
106	295	B	21_F	210_I	0.051	1.165	0.484	N/A
189	193	B	104_F	108_S	0.051	1.163	0.483	N/A
208	262	B	123_N	177_L	0.051	1.155	0.475	N/A
344	349	B	259_L	264_P	0.051	1.152	0.472	N/A
219	267	B	134_V	182_F	0.05	1.152	0.472	N/A
164	169	B	79_F	84_L	0.05	1.151	0.471	N/A
145	149	B	60_S	64_W	0.05	1.151	0.471	N/A
181	326	B	96_I	241_I	0.05	1.147	0.468	N/A
59	165	AB	59_E	80_A	0.05	1.146	0.467	0.213
252	270	B	167_A	185_L	0.05	1.145	0.466	N/A
286	341	B	201_M	256_R	0.05	1.141	0.462	N/A
337	340	B	252_L	255_F	0.05	1.138	0.459	N/A
221	224	B	136_F	139_Q	0.05	1.137	0.459	N/A
99	109	B	14_I	24_V	0.049	1.124	0.447	N/A
248	316	B	163_A	231_D	0.049	1.124	0.447	N/A
29	265	AB	29_I	180_L	0.049	1.124	0.447	0.197
344	352	B	259_L	267_L	0.049	1.122	0.445	N/A
182	187	B	97_I	102_R	0.049	1.119	0.442	N/A
197	201	B	112_L	116_I	0.049	1.119	0.442	N/A
189	198	B	104_F	113_G	0.049	1.115	0.438	N/A
313	316	B	228_P	231_D	0.049	1.113	0.437	N/A
168	220	B	83_L	135_L	0.049	1.111	0.435	N/A
208	236	B	123_N	151_H	0.048	1.103	0.428	N/A
268	271	B	183_P	186_M	0.048	1.096	0.421	N/A

127	260	B	42_F	175_L	0.048	1.093	0.418	N/A
122	180	B	37_N	95_S	0.048	1.091	0.417	N/A
163	251	B	78_F	166_A	0.048	1.09	0.416	N/A
263	291	B	178_L	206_V	0.048	1.09	0.416	N/A
186	189	B	101_L	104_F	0.047	1.083	0.410	N/A
171	216	B	86_T	131_F	0.047	1.082	0.409	N/A
214	281	B	129_S	196_I	0.047	1.074	0.401	N/A
181	291	B	96_I	206_V	0.047	1.072	0.400	N/A
94	133	B	9_I	48_T	0.047	1.07	0.398	N/A
139	321	B	54_A	236_C	0.047	1.066	0.394	N/A
296	303	B	211_I	218_T	0.047	1.066	0.394	N/A
147	293	B	62_D	208_V	0.047	1.065	0.393	N/A
189	195	B	104_F	110_N	0.047	1.063	0.392	N/A
184	280	B	99_F	195_K	0.046	1.057	0.386	N/A
245	344	B	160_N	259_L	0.046	1.056	0.386	N/A
100	316	B	15_A	231_D	0.046	1.056	0.386	N/A
189	194	B	104_F	109_G	0.046	1.054	0.384	N/A
129	133	B	44_A	48_T	0.046	1.052	0.382	N/A
21	24	A	21_A	24_C	0.046	1.05	0.380	N/A
182	201	B	97_I	116_I	0.046	1.05	0.380	N/A
185	191	B	100_Y	106_A	0.046	1.05	0.380	N/A
29	135	AB	29_I	50_V	0.046	1.049	0.379	0.148
117	214	B	32_W	129_S	0.046	1.048	0.379	N/A
43	60	A	43_C	60_C	0.046	1.046	0.377	N/A
142	158	B	57_L	73_M	0.046	1.042	0.373	N/A
267	288	B	182_F	203_F	0.046	1.042	0.373	N/A
136	325	B	51_N	240_A	0.046	1.038	0.370	N/A
18	136	AB	18_P	51_N	0.045	1.037	0.369	0.142
73	78	A	73_S	78_V	0.045	1.036	0.368	N/A
16	115	AB	16_V	30_F	0.045	1.026	0.360	0.135
92	154	B	7_T	69_D	0.045	1.026	0.360	N/A
161	165	B	76_R	80_A	0.045	1.025	0.359	N/A
182	185	B	97_I	100_Y	0.045	1.024	0.358	N/A
89	281	B	4_I	196_I	0.045	1.023	0.357	N/A
330	348	B	245_V	263_L	0.045	1.023	0.357	N/A
135	240	B	50_V	155_H	0.045	1.02	0.355	N/A
127	146	B	42_F	61_K	0.044	1.013	0.349	N/A
36	153	AB	36_P	68_F	0.044	1.01	0.346	0.127
174	180	B	89_R	95_S	0.044	1.008	0.344	N/A
294	298	B	209_M	213_L	0.044	1.008	0.344	N/A
117	121	B	32_W	36_S	0.044	1.004	0.341	N/A
218	247	B	133_A	162_L	0.044	0.997	0.335	N/A
274	284	B	189_N	199_G	0.044	0.996	0.334	N/A
134	138	B	49_G	53_W	0.044	0.995	0.334	N/A
330	336	B	245_V	251_C	0.044	0.994	0.333	N/A
294	297	B	209_M	212_S	0.043	0.99	0.330	N/A
107	339	B	22_A	254_S	0.043	0.989	0.329	N/A
297	300	B	212_S	215_R	0.043	0.989	0.329	N/A
249	253	B	164_W	168_A	0.043	0.985	0.325	N/A
64	126	AB	64_E	41_F	0.043	0.984	0.325	0.114
20	24	A	20_C	24_C	0.043	0.984	0.325	N/A
248	309	B	163_A	224_R	0.043	0.981	0.322	N/A
105	339	B	20_L	254_S	0.043	0.98	0.321	N/A
195	314	B	110_N	229_T	0.043	0.979	0.321	N/A
303	308	B	218_T	223_T	0.043	0.978	0.320	N/A
111	256	B	26_T	171_I	0.043	0.977	0.319	N/A
335	342	B	250_P	257_L	0.043	0.972	0.315	N/A
281	310	B	196_I	225_T	0.043	0.971	0.314	N/A
98	188	B	13_V	103_V	0.043	0.97	0.313	N/A
105	109	B	20_L	24_V	0.042	0.969	0.313	N/A
293	301	B	208_V	216_L	0.042	0.969	0.313	N/A
221	226	B	136_F	141_I	0.042	0.967	0.311	N/A
147	233	B	62_D	148_W	0.042	0.966	0.310	N/A
220	226	B	135_L	141_I	0.042	0.965	0.309	N/A
197	255	B	112_L	170_G	0.042	0.965	0.309	N/A
203	351	B	118_W	266_V	0.042	0.965	0.309	N/A
262	283	B	177_L	198_G	0.042	0.961	0.306	N/A
67	177	AB	67_T	92_V	0.042	0.961	0.306	0.103
305	347	B	220_N	262_M	0.042	0.961	0.306	N/A
255	291	B	170_G	206_V	0.042	0.957	0.303	N/A
94	98	B	9_I	13_V	0.042	0.956	0.302	N/A
116	330	B	31_W	245_V	0.042	0.952	0.299	N/A
61	248	AB	61_T	163_A	0.042	0.951	0.298	0.099
63	233	AB	63_K	148_W	0.042	0.951	0.298	0.099
143	334	B	58_G	249_C	0.042	0.948	0.296	N/A
248	304	B	163_A	219_I	0.041	0.947	0.295	N/A
93	101	B	8_F	16_V	0.041	0.947	0.295	N/A
53	57	A	53_T	57_K	0.041	0.946	0.295	N/A

62	235	AB	62_I	150_G	0.041	0.945	0.294	0.096
17	68	A	17_L	68_T	0.041	0.942	0.292	N/A
275	333	B	190_L	248_I	0.041	0.939	0.289	N/A
179	209	B	94_G	124_V	0.041	0.939	0.289	N/A
106	127	B	21_F	42_F	0.041	0.938	0.288	N/A
33	115	AB	33_K	30_F	0.041	0.936	0.287	0.092
288	291	B	203_F	206_V	0.041	0.934	0.285	N/A
31	42	A	31_Q	42_I	0.041	0.932	0.284	N/A
209	213	B	124_V	128_L	0.041	0.932	0.284	N/A
102	289	B	17_I	204_V	0.041	0.931	0.283	N/A
275	326	B	190_L	241_I	0.041	0.931	0.283	N/A
199	228	B	114_R	143_L	0.041	0.931	0.283	N/A
64	146	AB	64_E	61_K	0.041	0.931	0.283	0.09
37	307	AB	37_T	222_F	0.041	0.927	0.280	0.089
41	241	AB	41_C	156_C	0.041	0.927	0.280	0.089
42	48	A	42_I	48_L	0.041	0.926	0.279	N/A
35	77	A	35_D	77_G	0.041	0.926	0.279	N/A
114	279	B	29_Y	194_K	0.041	0.925	0.279	N/A
120	176	B	35_L	91_F	0.041	0.925	0.279	N/A
72	249	AB	72_T	164_W	0.041	0.924	0.278	0.088
252	301	B	167_A	216_L	0.04	0.923	0.277	N/A
93	321	B	8_F	236_C	0.04	0.923	0.277	N/A
273	278	B	188_L	193_K	0.04	0.922	0.276	N/A
273	307	B	188_L	222_F	0.04	0.918	0.274	N/A
201	265	B	116_I	180_L	0.04	0.915	0.271	N/A
215	248	B	130_F	163_A	0.04	0.915	0.271	N/A
219	304	B	134_V	219_I	0.04	0.915	0.271	N/A
102	206	B	17_I	121_V	0.04	0.915	0.271	N/A
169	320	B	84_L	235_L	0.04	0.914	0.271	N/A
113	256	B	28_A	171_I	0.04	0.912	0.269	N/A
54	169	AB	54_A	84_L	0.04	0.912	0.269	0.083
117	124	B	32_W	39_F	0.04	0.912	0.269	N/A
76	104	AB	76_C	19_R	0.04	0.912	0.269	0.083
293	312	B	208_V	227_N	0.04	0.909	0.267	N/A
170	266	B	85_Y	181_P	0.04	0.909	0.267	N/A
296	320	B	211_I	235_L	0.04	0.909	0.267	N/A
291	295	B	206_V	210_I	0.04	0.908	0.266	N/A
165	169	B	80_A	84_L	0.04	0.907	0.265	N/A
278	282	B	193_K	197_M	0.04	0.902	0.262	N/A
115	211	B	30_F	126_Y	0.04	0.902	0.262	N/A
297	308	B	212_S	223_T	0.039	0.901	0.261	N/A
262	282	B	177_L	197_M	0.039	0.901	0.261	N/A
45	59	A	45_S	59_E	0.039	0.901	0.261	N/A
292	296	B	207_A	211_I	0.039	0.899	0.260	N/A
285	289	B	200_M	204_V	0.039	0.897	0.258	N/A
41	223	AB	41_C	138_C	0.039	0.897	0.258	0.077
188	193	B	103_V	108_S	0.039	0.896	0.258	N/A
246	281	B	161_I	196_I	0.039	0.896	0.258	N/A
125	134	B	40_G	49_G	0.039	0.895	0.257	N/A
182	190	B	97_I	105_P	0.039	0.895	0.257	N/A
45	56	A	45_S	56_V	0.039	0.895	0.257	N/A
151	155	B	66_V	70_D	0.039	0.894	0.256	N/A
101	115	B	16_V	30_F	0.039	0.893	0.256	N/A
84	110	AB	84_S	25_L	0.039	0.892	0.255	0.076
34	241	AB	34_C	156_C	0.039	0.892	0.255	0.076
147	156	B	62_D	71_V	0.039	0.891	0.254	N/A
21	338	AB	21_A	253_P	0.039	0.891	0.254	0.075
41	349	AB	41_C	264_P	0.039	0.89	0.253	0.075
206	281	B	121_V	196_I	0.039	0.888	0.252	N/A
106	279	B	21_F	194_K	0.039	0.888	0.252	N/A
162	328	B	77_L	243_L	0.039	0.888	0.252	N/A
131	345	B	46_V	260_R	0.039	0.887	0.251	N/A
8	89	AB	8_L	4_I	0.039	0.887	0.251	0.074
203	206	B	118_W	121_V	0.039	0.887	0.251	N/A
28	52	A	28_A	52_A	0.039	0.886	0.251	N/A
181	310	B	96_I	225_T	0.039	0.886	0.251	N/A
116	325	B	31_W	240_A	0.039	0.886	0.251	N/A
146	308	B	61_K	223_T	0.039	0.886	0.251	N/A
177	352	B	92_V	267_L	0.039	0.885	0.250	N/A
330	333	B	245_V	248_I	0.039	0.885	0.250	N/A
19	131	AB	19_A	46_V	0.039	0.884	0.249	0.073
170	178	B	85_Y	93_R	0.039	0.881	0.247	N/A
250	254	B	165_V	169_T	0.039	0.88	0.246	N/A
180	253	B	95_S	168_A	0.039	0.879	0.246	N/A
174	204	B	89_R	119_T	0.038	0.878	0.245	N/A
17	95	AB	17_L	10_G	0.038	0.877	0.244	0.071
167	252	B	82_M	167_A	0.038	0.877	0.244	N/A
175	335	B	90_F	250_P	0.038	0.875	0.243	N/A

14	229	AB	14_Q	144_F	0.038	0.875	0.243	0.07
130	134	B	45_A	49_G	0.038	0.875	0.243	N/A
203	263	B	118_W	178_L	0.038	0.875	0.243	N/A
147	294	B	62_D	209_M	0.038	0.874	0.242	N/A
263	288	B	178_L	203_F	0.038	0.873	0.242	N/A
82	337	AB	82_S	252_L	0.038	0.872	0.241	0.069
26	354	AB	26_E	269_T	0.038	0.872	0.241	0.069
179	208	B	94_G	123_N	0.038	0.872	0.241	N/A
152	281	B	67_P	196_I	0.038	0.87	0.240	N/A
227	337	B	142_G	252_L	0.038	0.87	0.240	N/A
28	298	AB	28_A	213_L	0.038	0.869	0.239	0.068
45	49	A	45_S	49_N	0.038	0.868	0.238	N/A
271	284	B	186_M	199_G	0.038	0.867	0.238	N/A
182	192	B	97_I	107_K	0.038	0.866	0.237	N/A
298	349	B	213_L	264_P	0.038	0.866	0.237	N/A
33	108	AB	33_K	23_R	0.038	0.866	0.237	0.067
252	278	B	167_A	193_K	0.038	0.864	0.236	N/A
243	279	B	158_N	194_K	0.038	0.864	0.236	N/A
259	280	B	174_D	195_K	0.038	0.863	0.235	N/A
208	258	B	123_N	173_Y	0.038	0.862	0.234	N/A
20	60	A	20_C	60_C	0.038	0.86	0.233	N/A
83	315	AB	83_D	230_K	0.038	0.86	0.233	0.065
214	311	B	129_S	226_V	0.038	0.86	0.233	N/A
164	220	B	79_F	135_L	0.038	0.86	0.233	N/A
91	95	B	6_A	10_G	0.038	0.858	0.232	N/A
278	328	B	193_K	243_L	0.038	0.858	0.232	N/A
174	211	B	89_R	126_Y	0.038	0.857	0.231	N/A
174	178	B	89_R	93_R	0.038	0.856	0.230	N/A
243	314	B	158_N	229_T	0.037	0.855	0.230	N/A
319	326	B	234_Q	241_I	0.037	0.854	0.229	N/A
207	261	B	122_F	176_W	0.037	0.853	0.228	N/A
34	223	AB	34_C	138_C	0.037	0.852	0.228	0.063
273	299	B	188_L	214_V	0.037	0.852	0.228	N/A
7	10	A	7_L	10_L	0.037	0.85	0.226	N/A
190	193	B	105_P	108_S	0.037	0.85	0.226	N/A
179	283	B	94_G	198_G	0.037	0.849	0.226	N/A
125	310	B	40_G	225_T	0.037	0.848	0.225	N/A
202	205	B	117_Q	120_L	0.037	0.848	0.225	N/A
70	239	AB	70_N	154_G	0.037	0.848	0.225	0.061
179	183	B	94_G	98_L	0.037	0.847	0.224	N/A
237	340	B	152_D	255_F	0.037	0.846	0.224	N/A
50	254	AB	50_S	169_T	0.037	0.846	0.224	0.061
92	117	B	7_T	32_W	0.037	0.845	0.223	N/A
229	321	B	144_F	236_C	0.037	0.845	0.223	N/A
11	89	AB	11_D	4_I	0.037	0.845	0.223	0.061
174	262	B	89_R	177_L	0.037	0.845	0.223	N/A
243	308	B	158_N	223_T	0.037	0.845	0.223	N/A
176	302	B	91_F	217_K	0.037	0.844	0.223	N/A
341	349	B	256_R	264_P	0.037	0.844	0.223	N/A
79	324	AB	79_T	239_S	0.037	0.842	0.221	0.06
108	270	B	23_R	185_L	0.037	0.842	0.221	N/A
214	329	B	129_S	244_D	0.037	0.841	0.221	N/A
101	123	B	16_V	38_L	0.037	0.839	0.219	N/A
183	187	B	98_L	102_R	0.037	0.839	0.219	N/A
179	201	B	94_G	116_I	0.037	0.838	0.219	N/A
199	203	B	114_R	118_W	0.037	0.838	0.219	N/A
47	226	AB	47_A	141_I	0.037	0.836	0.217	0.058
99	231	B	14_I	146_T	0.037	0.836	0.217	N/A
62	148	AB	62_I	63_V	0.037	0.836	0.217	0.058
226	325	B	141_I	240_A	0.037	0.835	0.217	N/A
99	130	B	14_I	45_A	0.037	0.834	0.216	N/A
134	221	B	49_G	136_F	0.037	0.833	0.216	N/A
92	348	B	7_T	263_L	0.037	0.833	0.216	N/A
175	183	B	90_F	98_L	0.037	0.833	0.216	N/A
235	270	B	150_G	185_L	0.036	0.832	0.215	N/A
56	345	AB	56_V	260_R	0.036	0.83	0.214	0.056
92	96	B	7_T	11_L	0.036	0.83	0.214	N/A
110	184	B	25_L	99_F	0.036	0.829	0.213	N/A
259	266	B	174_D	181_P	0.036	0.828	0.212	N/A
276	331	B	191_H	246_G	0.036	0.828	0.212	N/A
106	161	B	21_F	76_R	0.036	0.827	0.212	N/A
130	227	B	45_A	142_G	0.036	0.827	0.212	N/A
159	201	B	74_V	116_I	0.036	0.827	0.212	N/A
256	301	B	171_I	216_L	0.036	0.827	0.212	N/A
301	316	B	216_L	231_D	0.036	0.826	0.211	N/A
131	134	B	46_V	49_G	0.036	0.826	0.211	N/A
319	322	B	234_Q	237_L	0.036	0.825	0.211	N/A
115	216	B	30_F	131_F	0.036	0.825	0.211	N/A

133	227	B	48_T	142_G	0.036	0.825	0.211	N/A
28	165	AB	28_A	80_A	0.036	0.825	0.211	0.055
10	89	AB	10_L	4_I	0.036	0.824	0.210	0.055
294	312	B	209_M	227_N	0.036	0.822	0.209	N/A
56	82	A	56_V	82_S	0.036	0.822	0.209	N/A
119	338	B	34_D	253_P	0.036	0.821	0.208	N/A
209	238	B	124_V	153_H	0.036	0.821	0.208	N/A
245	297	B	160_N	212_S	0.036	0.82	0.208	N/A
257	316	B	172_A	231_D	0.036	0.819	0.207	N/A
92	213	B	7_T	128_L	0.036	0.819	0.207	N/A
49	251	AB	49_N	166_A	0.036	0.819	0.207	0.053
158	213	B	73_M	128_L	0.036	0.818	0.206	N/A
63	234	AB	63_K	149_Q	0.036	0.818	0.206	0.053
201	205	B	116_I	120_L	0.036	0.818	0.206	N/A
301	307	B	216_L	222_F	0.036	0.818	0.206	N/A
107	123	B	22_A	38_L	0.036	0.817	0.206	N/A
202	261	B	117_Q	176_W	0.036	0.817	0.206	N/A
112	337	B	27_K	252_L	0.036	0.816	0.205	N/A
310	318	B	225_T	233_V	0.036	0.815	0.205	N/A
40	221	AB	40_S	136_F	0.036	0.815	0.205	0.052
117	295	B	32_W	210_I	0.036	0.814	0.204	N/A
313	350	B	228_P	265_H	0.036	0.814	0.204	N/A
200	203	B	115_V	118_W	0.036	0.813	0.203	N/A
325	331	B	240_A	246_G	0.036	0.812	0.203	N/A
53	127	AB	53_T	42_F	0.036	0.812	0.203	0.052
155	350	B	70_D	265_H	0.036	0.812	0.203	N/A
228	231	B	143_L	146_T	0.036	0.81	0.202	N/A
172	306	B	87_A	221_Q	0.035	0.81	0.202	N/A
156	333	B	71_V	248_I	0.035	0.81	0.202	N/A
116	342	B	31_W	257_L	0.035	0.809	0.201	N/A
215	222	B	130_F	137_Q	0.035	0.809	0.201	N/A
246	260	B	161_I	175_L	0.035	0.809	0.201	N/A
60	66	A	60_C	66_L	0.035	0.807	0.200	N/A
90	141	B	5_Y	56_I	0.035	0.807	0.200	N/A
45	273	AB	45_S	188_L	0.035	0.806	0.199	0.05
222	275	B	137_Q	190_L	0.035	0.805	0.199	N/A
165	228	B	80_A	143_L	0.035	0.805	0.199	N/A
100	120	B	15_A	35_L	0.035	0.804	0.198	N/A
98	126	B	13_V	41_F	0.035	0.803	0.197	N/A
69	73	A	69_K	73_S	0.035	0.802	0.197	N/A
127	151	B	42_F	66_V	0.035	0.802	0.197	N/A
82	342	AB	82_S	257_L	0.035	0.801	0.196	0.049
78	169	AB	78_V	84_L	0.035	0.8	0.196	0.048
233	312	B	148_W	227_N	0.035	0.8	0.196	N/A
91	346	B	6_A	261_R	0.035	0.8	0.196	N/A
84	184	AB	84_S	99_F	0.035	0.8	0.196	0.048
168	310	B	83_L	225_T	0.035	0.8	0.196	N/A
262	288	B	177_L	203_F	0.035	0.799	0.195	N/A
59	280	AB	59_E	195_K	0.035	0.799	0.195	0.048
156	321	B	71_V	236_C	0.035	0.799	0.195	N/A
204	207	B	119_T	122_F	0.035	0.799	0.195	N/A
42	181	AB	42_I	96_I	0.035	0.798	0.194	0.048
211	329	B	126_Y	244_D	0.035	0.797	0.194	N/A
155	243	B	70_D	158_N	0.035	0.796	0.193	N/A
81	150	AB	81_K	65_F	0.035	0.793	0.192	0.047
97	146	B	12_A	61_K	0.035	0.793	0.192	N/A
171	220	B	86_T	135_L	0.035	0.793	0.192	N/A
73	172	AB	73_S	87_A	0.035	0.792	0.191	0.047
7	186	AB	7_L	101_L	0.035	0.792	0.191	0.047
40	128	AB	40_S	43_G	0.035	0.791	0.190	0.046
318	321	B	233_V	236_C	0.035	0.791	0.190	N/A
52	165	AB	52_A	80_A	0.035	0.791	0.190	0.046
22	132	AB	22_L	47_F	0.035	0.791	0.190	0.046
208	211	B	123_N	126_Y	0.035	0.79	0.190	N/A
170	259	B	85_Y	174_D	0.035	0.79	0.190	N/A
263	302	B	178_L	217_K	0.035	0.789	0.189	N/A
344	355	B	259_L	270_S	0.035	0.789	0.189	N/A
64	74	A	64_E	74_Y	0.035	0.789	0.189	N/A
63	151	AB	63_K	66_V	0.035	0.788	0.189	0.046
144	318	B	59_Q	233_V	0.035	0.788	0.189	N/A
147	150	B	62_D	65_F	0.035	0.788	0.189	N/A
29	137	AB	29_I	52_I	0.035	0.787	0.188	0.045
58	215	AB	58_A	130_F	0.035	0.787	0.188	0.045
27	218	AB	27_T	133_A	0.034	0.787	0.188	0.045
275	281	B	190_L	196_I	0.034	0.786	0.188	N/A
77	144	AB	77_G	59_Q	0.034	0.786	0.188	0.045
143	159	B	58_G	74_V	0.034	0.786	0.188	N/A
175	219	B	90_F	134_V	0.034	0.784	0.186	N/A

214	310	B	129_S	225_T	0.034	0.784	0.186	N/A
7	194	AB	7_L	109_G	0.034	0.784	0.186	0.045
24	43	A	24_C	43_C	0.034	0.784	0.186	N/A
68	250	AB	68_T	165_V	0.034	0.783	0.186	0.044
20	293	AB	20_C	208_V	0.034	0.783	0.186	0.044
58	254	AB	58_A	169_T	0.034	0.783	0.186	0.044
304	350	B	219_I	265_H	0.034	0.782	0.185	N/A
281	306	B	196_I	221_Q	0.034	0.782	0.185	N/A
171	332	B	86_T	247_V	0.034	0.781	0.185	N/A
14	337	AB	14_Q	252_L	0.034	0.781	0.185	0.044
62	310	AB	62_I	225_T	0.034	0.781	0.185	0.044
38	326	AB	38_D	241_I	0.034	0.781	0.185	0.044
24	41	A	24_C	41_C	0.034	0.78	0.184	N/A
115	305	B	30_F	220_N	0.034	0.78	0.184	N/A
177	331	B	92_V	246_G	0.034	0.78	0.184	N/A
22	267	AB	22_L	182_F	0.034	0.778	0.183	0.043
18	286	AB	18_P	201_M	0.034	0.778	0.183	0.043
289	292	B	204_V	207_A	0.034	0.777	0.183	N/A
58	347	AB	58_A	262_M	0.034	0.776	0.182	0.043
273	330	B	188_L	245_V	0.034	0.774	0.181	N/A
33	303	AB	33_K	218_T	0.034	0.774	0.181	0.042
53	162	AB	53_T	77_L	0.034	0.771	0.179	0.042
127	172	B	42_F	87_A	0.034	0.771	0.179	N/A
157	289	B	72_T	204_V	0.034	0.77	0.179	N/A
178	215	B	93_R	130_F	0.034	0.77	0.179	N/A
186	192	B	101_L	107_K	0.034	0.77	0.179	N/A
261	345	B	176_W	260_R	0.034	0.77	0.179	N/A
97	329	B	12_A	244_D	0.034	0.77	0.179	N/A
184	187	B	99_F	102_R	0.034	0.768	0.178	N/A
196	260	B	111_K	175_L	0.034	0.768	0.178	N/A
71	75	A	71_I	75_M	0.034	0.767	0.177	N/A
201	204	B	116_I	119_T	0.034	0.765	0.176	N/A
50	295	AB	50_S	210_I	0.034	0.765	0.176	0.041
318	322	B	233_V	237_L	0.034	0.765	0.176	N/A
120	163	B	35_L	78_F	0.034	0.765	0.176	N/A
247	270	B	162_L	185_L	0.034	0.765	0.176	N/A
227	237	B	142_G	152_D	0.034	0.764	0.176	N/A
82	265	AB	82_S	180_L	0.033	0.764	0.176	0.04
170	223	B	85_Y	138_C	0.033	0.763	0.175	N/A
58	350	AB	58_A	265_H	0.033	0.763	0.175	0.04
54	282	AB	54_A	197_M	0.033	0.762	0.174	0.04
210	242	B	125_V	157_G	0.033	0.762	0.174	N/A
129	310	B	44_A	225_T	0.033	0.762	0.174	N/A
266	287	B	181_P	202_F	0.033	0.761	0.174	N/A
107	343	B	22_A	258_L	0.033	0.761	0.174	N/A
63	185	AB	63_K	100_Y	0.033	0.761	0.174	0.04
324	334	B	239_S	249_C	0.033	0.76	0.173	N/A
146	150	B	61_K	65_F	0.033	0.759	0.173	N/A
17	202	AB	17_L	117_Q	0.033	0.759	0.173	0.039
65	175	AB	65_S	90_F	0.033	0.759	0.173	0.039
189	227	B	104_F	142_G	0.033	0.759	0.173	N/A
332	351	B	247_V	266_V	0.033	0.758	0.172	N/A
33	96	AB	33_K	11_L	0.033	0.758	0.172	0.039
326	341	B	241_I	256_R	0.033	0.758	0.172	N/A
130	248	B	45_A	163_A	0.033	0.758	0.172	N/A
259	294	B	174_D	209_M	0.033	0.758	0.172	N/A
59	116	AB	59_E	31_W	0.033	0.757	0.172	0.039
81	312	AB	81_K	227_N	0.033	0.757	0.172	0.039
217	238	B	132_F	153_H	0.033	0.756	0.171	N/A
29	301	AB	29_I	216_L	0.033	0.755	0.171	0.038
147	151	B	62_D	66_V	0.033	0.755	0.171	N/A
59	206	AB	59_E	121_V	0.033	0.755	0.171	0.038
78	135	AB	78_V	50_V	0.033	0.754	0.170	0.038
27	92	AB	27_T	7_T	0.033	0.754	0.170	0.038
178	283	B	93_R	198_G	0.033	0.754	0.170	N/A
97	235	B	12_A	150_G	0.033	0.754	0.170	N/A
8	189	AB	8_L	104_F	0.033	0.753	0.170	0.038
236	270	B	151_H	185_L	0.033	0.752	0.169	N/A
171	325	B	86_T	240_A	0.033	0.752	0.169	N/A
64	125	AB	64_E	40_G	0.033	0.752	0.169	0.038
198	274	B	113_G	189_N	0.033	0.752	0.169	N/A
222	286	B	137_Q	201_M	0.033	0.752	0.169	N/A
93	96	B	8_F	11_L	0.033	0.752	0.169	N/A
74	77	A	74_Y	77_G	0.033	0.751	0.169	N/A
39	190	AB	39_L	105_P	0.033	0.751	0.169	0.038
312	338	B	227_N	253_P	0.033	0.751	0.169	N/A
283	288	B	198_G	203_F	0.033	0.751	0.169	N/A
28	54	A	28_A	54_A	0.033	0.75	0.168	N/A

185	193	B	100_Y	108_S	0.033	0.75	0.168	N/A
323	327	B	238_W	242_E	0.033	0.75	0.168	N/A
349	355	B	264_P	270_S	0.033	0.75	0.168	N/A
136	163	B	51_N	78_F	0.033	0.748	0.167	N/A
53	69	A	53_T	69_K	0.033	0.748	0.167	N/A
72	260	AB	72_T	175_L	0.033	0.747	0.167	0.037
115	220	B	30_F	135_L	0.033	0.747	0.167	N/A
134	142	B	49_G	57_L	0.033	0.747	0.167	N/A
153	275	B	68_F	190_L	0.033	0.746	0.166	N/A
243	330	B	158_N	245_V	0.033	0.746	0.166	N/A
124	307	B	39_F	222_F	0.033	0.746	0.166	N/A
112	255	B	27_K	170_G	0.033	0.746	0.166	N/A
53	339	AB	53_T	254_S	0.033	0.745	0.166	0.037
129	173	B	44_A	88_T	0.033	0.745	0.166	N/A
80	288	AB	80_P	203_F	0.033	0.744	0.165	0.036
48	52	A	48_L	52_A	0.033	0.744	0.165	N/A