

**Structural and functional studies of  
mucin-interacting  
adhesion domains  
from *Candida glabrata*  
& *Helicobacter pylori***

A dissertation submitted  
in fulfillment of the requirements for the degree of  
DOCTOR IN NATURAL SCIENCES  
(Dr. Rer. Nat.)

Presented at  
the Department of Chemistry  
Philipps University Marburg  
by

Manuel Maestre Reyna  
from  
Valencia, Spain

Marburg/Lahn 2011

Received on the \_\_\_\_\_ at the Chemistry Department, Philipps  
University Marburg.

First advisor: Prof. Dr. Lars-Oliver Essen (Department of Chemistry, Marburg).

Second advisor: Prof. Dr. Hans-Ulrich Mösch (Department of Biology, Marburg).

Date of Examination: \_\_\_\_\_



I declare that to the best of my knowledge and belief, this thesis with the title:

**“Structural and functional studies of mucin-interacting adhesion domains from *Candida glabrata* & *Helicobacter pylori*”**

contains no material previously published or written by another person, except where due reference has been made. It also does not contain any material which has been accepted for the award of any other degree or diploma in any University .

Marburg, March 29<sup>th</sup> 2011,

---

Manuel Maestre Reyna



Further publications:

Psakis, G., Nitschkowski, S., Holz, C., Kress, D., **Maestre-Reyna, M.**, Polaczek, J., Illing, G., Essen, L.-O. (2007). *Expression screening of integral membrane proteins from Helicobacter pylori* 26695. *Protein Sci.* **16**,12,2667-76

Posters:

Psakis, G., Nitschkowski, S., Hölscher, S., Neuhaus, C., Kress, D., **Maestre-Reyna, M.**, Essen, L.-O. (2007). *High-throughput expression and crystallisation of membrane proteins from Helicobacter pylori and homologues*. *Structural Proteomics of Membrane Proteins*, Rauschholzhäuser.

*To my family and  
other animals\**

---

\*As Gerald Durrell would have put it.



# Table of Contents

1	Summary/Zusammenfassung.....	1
1.1	Epithelial adhesins from <i>Candida glabrata</i> .....	1
1.2	Adhesins of <i>Helicobacter pylori</i> .....	2
1.3	Epitheladhäsine von <i>C. glabrata</i> .....	2
1.4	Adhäsine von <i>H. pylori</i> .....	3
2	Introduction.....	5
2.1	Mucins are primary targets for pathogenic adhesins.....	5
2.1.1	Mucin structure and function.....	5
2.1.2	Mucin tissue localization.....	6
2.1.3	Mucin glycosylation.....	9
2.2	<i>Candida glabrata</i> .....	12
2.2.1	The genetic background of <i>C. glabrata</i> .....	12
2.2.2	The cell wall of <i>C. glabrata</i> .....	14
2.2.3	The adhesins of <i>C. glabrata</i> .....	15
2.2.4	Pathogenicity of <i>C. glabrata</i> .....	17
2.3	<i>Helicobacter pylori</i> .....	19
2.3.1	Colonization and pathogenesis of <i>H. pylori</i> .....	19
2.3.2	Adhesion in <i>H. pylori</i> .....	23
2.4	The Structural features of glycan binding proteins from pathogenic origin.....	25
2.4.1	The PA14 domain.....	26
2.4.2	Type V protein secretion: autotransport.....	30
3	Aim.....	33
4	Materials.....	34
4.1	Chemicals, consumables and equipment.....	34
4.2	Vectors and Microorganisms.....	38
4.2.1	Vectors.....	38
4.2.2	<i>E. coli</i> strains.....	39
4.3	Primers.....	41
4.3.1	Primers for <i>de novo</i> cloning:.....	41
4.3.2	Mutagenesis primers:.....	41
4.4	Media, buffers and Stock solutions.....	42
4.4.1	Media:.....	42
4.4.2	Buffers:.....	43
4.4.3	Stock solutions:.....	45
5	Methods.....	48
5.1	Molecular biology.....	48
5.1.1	DNA synthesis methods.....	48
5.1.2	Cloning.....	50
5.1.3	DNA Isolation methods.....	51
5.1.4	Agarose gel electrophoresis.....	52
5.1.5	Cell preparation methods.....	52
5.2	Biochemical methods.....	54
5.2.1	Recombinant gene expression.....	54
5.2.2	Cell lysis.....	55
5.2.3	Protein purification.....	57
5.2.4	Protein concentration.....	59
5.2.5	Protein refolding.....	59
5.2.6	Estimation of protein concentration.....	62

5.2.7 Protein precipitation.....	63
5.3 Analytical Methods.....	64
5.3.1 Western blotting.....	64
5.3.2 CD spectroscopy.....	65
5.3.3 Analytical gel filtration.....	66
5.3.4 High throughput glycan binding studies at the Consortium for Functional Glycomics..	67
5.3.5 Fluorescence spectroscopy .....	69
5.4 Crystallographic Methods.....	73
5.4.1 Crystallization screening.....	73
5.4.2 Optimizing and reproducing crystallization.....	74
5.4.3 Diffractometric measurements and sample preparation.....	75
5.5 Computer assisted methods.....	77
5.5.1 Homology model building methods.....	77
5.5.2 Structure elucidation methods.....	88
5.5.3 Protein properties calculation methods.....	92
6 Results.....	96
6.1 Epithelial Adhesins from <i>Candida glabrata</i> (Epa).....	96
6.1.1 Purification of the Epa1A domain.....	97
6.1.2 Structure determination of the Epa1A domain.....	98
6.1.3 Generation of Epa1A variants.....	102
6.1.4 Expression of Epa variants.....	104
6.1.5 Purification of Epa variants.....	104
6.1.6 Structural analyses of subclass-converted Epa1A variants.....	105
6.1.7 Carbohydrate binding of Epa A domains.....	110
6.1.8 Quantitative carbohydrate binding assays via fluorescence titration.....	119
6.2 Lewis antigen binding adhesins from <i>H. pylori</i> .....	122
6.2.1 Bioinformatic studies.....	123
6.2.2 Refolding of BabA fragments.....	124
6.2.3 CD spectroscopy of BabA 235.....	125
6.2.4 pH dependent analytical gel filtration of BabA 235 fragment.....	130
6.2.5 High-throughput semi-quantitative binding assays of BabA 235.....	131
6.2.6 Other Strategies for BabA235.....	135
6.2.7 Recombinant expression and purification of SabA extracellular fragments.....	137
7 Discussion.....	139
7.1 <i>H. pylori</i> adhesin BabA.....	140
7.1.1 The BabA extracellular domain can suffer profound pH-dependent conformational changes.....	140
7.1.2 <i>H. pylori</i> interacts with the host in a pH dependent manner.....	144
7.1.3 BabA specificity shift towards LewisX and LewisY blood group antigens.....	145
7.1.4 Outlook.....	147
7.2 Epithelial adhesins from <i>C. glabrata</i> .....	148
7.2.1 Comparison of the A domain from Epa adhesins with <i>S. cerevisiae</i> Flo5 flocculin.....	149
7.2.2 The binding pocket of Epa1A.....	151
7.2.3 A structural model for carbohydrate binding in Epa1A.....	153
7.2.4 The specificity of Epa1A and its variants.....	154
7.2.5 Epa A domains do not bind lactose.....	159
7.2.6 Outlook.....	160
8 Acknowledgements.....	162
Bibliography.....	I
9 Appendices.....	1

9.1 Appendix I: Glycan Arrays.....	1
9.1.1 Array V3.1.....	1
9.1.2 Glycan array V4.1.....	13
9.2 Appendix II: Protein sequences.....	29
9.2.1 Sequences of Epa1A and variants.....	29
9.2.2 Sequences of BabA extracellular fragments from <i>H. pylori</i> 26695.....	31
9.2.3 Sequences of BabA extracellular fragments from <i>H. pylori</i> J99.....	32
9.2.4 Sequences of BabA extracellular fragments from <i>H. pylori</i> P12.....	33
9.2.5 Sequences of BabA extracellular fragments from <i>H. pylori</i> G27.....	35
9.3 Appendix III: fluorescence spectroscopy data.....	38



# 1 Summary/Zusammenfassung

## 1.1 Epithelial adhesins from *Candida glabrata*

Epithelial adhesins (Epa) are crucial proteins in the colonization, pathogenesis and virulence of *Candida glabrata*. These adhesins have a similar modular structure to *Saccharomyces cerevisiae* flocculins, with an N-terminal adhesive A domain, a central neck-like B domain, and an anchorage C-terminal C domain. A hallmark of many fungal adhesins is the presence of a calcium binding PA14 domain within the A domain. The PA14 domain is responsible for carbohydrate binding in a calcium dependent manner, which allows for the classification of these proteins as C-type lectins[1].

In this study, it was possible to elucidate the crystal structures of Epa1A and three variants at resolutions from 1.4 to 2.0 Å. The latter were meant to emulate the specificities of Epa2, Epa3 and Epa6 adhesive domains. The results yielded a profound knowledge of the binding pocket of Epa1A and the mechanisms through which specificity is controlled in the Epa A domain. Especially surprising was the fact that, even though the proteins were crystallized in the presence of lactose, the protein co-crystals never showed the aforementioned sugar. Instead, a galactose $\beta$ 1-3glucose disaccharide unit could be modeled into the electron density. The disaccharide is commonly found on cell surfaces and milk derivatives, from which the employed lactose was obtained[2].

Epa1A, Epa1 $\rightarrow$ 2A, Epa1 $\rightarrow$ 3A and Epa1 $\rightarrow$ 6A were also functionally characterized by semi-quantitative, high-throughput methodologies. In collaboration with the consortium for functional glycomics, fluorescently labeled proteins were set in contact with large-scale glycan arrays. The results showed a marked preference for galactose $\beta$ 1-3 terminal oligosaccharides in the case of Epa1A. For the other proteins, varying degrees of promiscuity were noted. Epa1 $\rightarrow$ 6A presented a very similar binding profile to the one presented by Zupancic *et. al.* in 2008 for Epa6A, demonstrating the validity of the method. Epa1 $\rightarrow$ 2A and Epa1 $\rightarrow$ 3A were much less active, and presented a preference for sulfated glycans, along with terminal galactose. Fluorescence titrations showed for Epa1A a  $\sim$ 20 time stronger affinity for the T antigen (galactose $\beta$ 1-3N-acetyl-galactosamine) than for the milk-derived lactose, showing how marked the adhesin preference for  $\beta$ 1-3 linkages is, as compared to  $\beta$ 1-4 glycosidic bonds.

## **1.2 Adhesins of *Helicobacter pylori***

The adhesins of *H. pylori* have been shown to be critical for the colonization and immune recognition of the bacterium during gastric invasion and disease development[3]. BabA and SabA figure prominently, as the former is the primary adhesin during early stage colonization, while the latter binds strongly to inflamed tissue[4]. Both of them are autotransporters, with a C-terminal, membrane bound translocation unit and an N-terminal passenger domain which contains the adhesive portion of the protein[5].

Pure, soluble passenger domains of BabA and SabA were successfully overproduced by recombinant expression in *Escherichia coli*. BabA could be functionally characterized by the same method as the Epa proteins. The results showed that BabA activity was strongly pH dependent, with a ~100 times stronger activity at pH 5.8 than at pH 2.5. This behavior could be further characterized through circular dichroism spectroscopy and size exclusion chromatography, which showed that BabA is in a reversible molten globule-like, aggregation-prone and relaxed conformation at pH 2.5. At pH 5.8, on the other hand, the protein is in a much more compact, defined conformation with a strong tendency to precipitate. *H. pylori* has been shown to present many pH dependent virulence factors, like the urea transporter, but up to now no direct biochemical data had been presented supporting pH dependent conformational changes in its adhesins.

## **1.3 Epitheladhäsine von *C. glabrata***

Die Epitheladhäsine (Epa) sind zentrale Proteine in der Kolonisierung, Pathogenese und Virulenz von *C. glabrata*. Diese Adhäsine haben einen ähnlichen modularen Aufbau wie *S. cerevisiae* Flokuline, mit einer N-terminalen adhäsiven A-Domäne, einer halsartigen B-Domäne und einer C-terminalen C-Domäne, die als Verankerung an der Zellwand dient. Flokulin-artige Adhäsine können in vielen Fällen eine Calcium-bindende PA14 Domäne innerhalb der A-Domäne beinhalten. Diese Proteinen wurden als C-Typ Lektine eingestuft, da die PA14 Domäne für eine Calcium abhängige Glycan-Bindung verantwortlich ist[1].

In der folgenden Arbeit war es möglich, die Kristallstrukturen von Epa1A und drei Varianten mit Auflösungen von 1,4 bis 2,0 Å zu lösen. Die Varianten sollten die Spezifitäten der Epa2-, Epa3- und Epa6- adhäsiven A-Domänen emulieren. Diese Erkenntnisse erlaubten eine detaillierte Beschreibung der Bindungstasche von Epa1A und die Mechanismen, durch die die Spezifität in der Epa A- Domäne bestimmt werden. Besonders überraschend war die Tatsache, dass, obwohl die

Proteine in Gegenwart von Laktose kristallisiert wurden, die Co-Kristalle keine Laktose zeigten. Stattdessen konnte nur ein Galaktose $\beta$ 1-3glukose Disaccharid in die Elektronendichte modelliert werden. Dieses Disaccharid findet sich für gewöhnlich auf Zelloberflächen und in Milchderivaten[2], aus denen die verwendete Laktose stammte.

Die funktionelle Charakterisierung von Epa1A, Epa1 $\rightarrow$ 2A, Epa1 $\rightarrow$ 3A und Epa1 $\rightarrow$ 6A verlief über semiquantitative, Hochdurchsatz Analysen in Zusammenarbeit mit dem *Consortium for Functional Glycomics* (CFG). Mit Fluorophoren markierte Proteine wurden mit Glykan Arrays in Kontakt gesetzt. Die Ergebnisse zeigten eine deutliche Präferenz im Falle von Epa1A für Oligosaccharide mit einer terminalen Galaktose $\beta$ 1-3 Verknüpfung. Für die anderen Proteinen konnte gezeigt werden, dass die Spezifität in unterschiedlichen Massen verringert war. Das Epa1 $\rightarrow$ 6A Bindungsprofil konnte mit der von Zupancic *et. al.* 2008 publizierten Daten für Epa6A verglichen werden und zeigte große Ähnlichkeit. Epa1 $\rightarrow$ 2A und 3A waren deutlich weniger aktiv, zeigten aber eine Präferenz für sulfatierte Glykane sowie für terminale Galaktosen. Quantitative Fluoreszenztitrationen ergaben ungefähr zwanzigfach höhere Affinitäten für das T-Antigen als für die aus Milch gewonnene Laktose, wodurch die Präferenz für  $\beta$ 1-3 über  $\beta$ 1-4 Verbindungen weiter bestätigt werden konnte.

#### ***1.4 Adhäsine von Helicobacter pylori***

Die Adhäsine von *H. pylori* zeigten sich als kritisch bei der Kolonisierung und Immunerkennung des Bakteriums während des Eindringens in den Magen und der Entstehung und Entwicklung der Krankheit[3]. BabA und SabA sind zwei besonders wichtige Adhäsine, da ersteres bei einer frühen Kolonisierung als Primäradhäsin wirkt, während SabA stark an entzündetes Gewebe bindet[4]. Es handelt sich bei beiden um Autotransporter, die mit einer C-terminalen, membrangebundenen Translokationseinheit und einer N-terminalen, adhäsiven transportierte Domäne versetzt sind.

Reine, lösliche transportierte Domänen von BabA und SabA wurden in großen Mengen in *Escherichia coli* rekombinant überproduziert. BabA konnte, wie die Epa Proteine, mit Hilfe von Glykan Arrays funktionell charakterisiert werden. BabA-Aktivität erwies sich als stark pH-abhängig, da sich bei pH 5.8 eine etwa hundertfach stärkere Bindung als bei pH 2.5 zeigte. Dieses Verhalten konnte durch Circular-Dichroismus Spektroskopie und chromatographische Analyse weiter untersucht werden. Es zeigte sich, dass BabA bei pH 2.5 sich in eine relaxierten, molten-globule ähnlichen Konformation befand, die leicht aggregierte. Bei pH 5.8 andererseits

befand sich das Protein in einer viel kompakteren, klarer definierten Konformation mit einer starken Tendenz zur Präzipitat-Entstehung. Es konnten schon mehrere *H. pylori* Virulenzfaktoren gefunden werden, deren Aktivität pH abhängig ist, aber bis jetzt gab es noch keine biochemischen Daten, die pH abhängige Konformationsänderungen bei ihren Adhäsinen belegten.

## 2 Introduction

### 2.1 *Mucins are primary targets for pathogenic adhesins*

Mucins are the main non-cellular component of the mammalian mucosa[6]. Mucins are highly glycosylated glycoproteins, which are generally classified as tethered mucins or secreted mucins, depending on whether they are bound on the cell surface or secreted into the intercellular space[7]. The glycans decorating mucins are the target for the adhesins of several microbial adhesins, including those of *Helicobacter pylori* and *Candida glabrata*[8][9].

#### 2.1.1 *Mucin structure and function*

Around 20 different mucin genes have been characterized through data mining, RT PCR and other methods[6] (Fig. 2.1). Mucins show an extremely diverse and tissue specific splicing and glycosylation behavior, which gives rise to a wide variety of gene products, that are unevenly spread through the different tissues of the mammalian organism[10]. All mucins have a similar “bottle brush” structure with a long, extended protein core unto which mucin-type O-glycans are bound (Fig. 2.1), which constitute at least 50% of the molecular mass of the glycoprotein. The hallmark of all mucins is the presence of “variable number tandem repeats” (VNTR), serine threonine rich regions that act as O-glycan binding sites. Cysteine rich regions and Von Willenbrand factor like D-domains may act in the polymerization of mucins[11]. Most tethered mucins have also been shown to contain N-glycosylation sites near their C-terminal regions (Fig. 2.1).

Secreted mucins act as a buffering system, protect the epithelium from invasion of pathogens, keep it lubricated and protect it from the action of proteases[12][13]. Tethered mucins, on the other hand, have been shown to have many different functions beyond the aforementioned; they are involved in cellular signaling [14], adhesion and anti-adhesion between mammalian cells[15], and also in disease, both neoplastic[16], genetic [17][18] and infectious[19]. Although mucins act to prevent infection, it has been shown that at least in the case of *Pseudomonas aeruginosa*[19], *Helicobacter pylori*[20][7], and *Escherichia coli*[21], mucins can act as adhesion targets during tissue invasion.

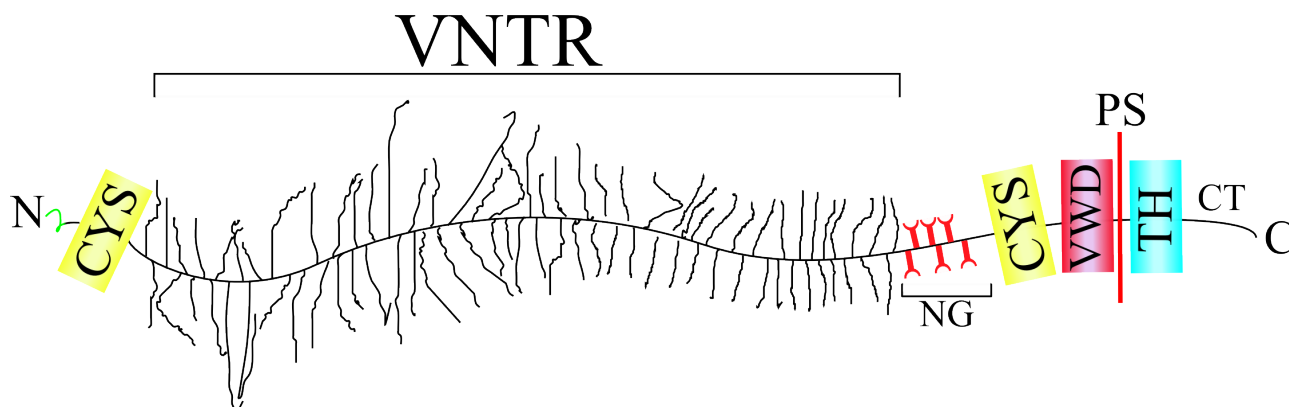


Figure 2.1: Simplified structure of an archetypal mucin. Mucins show extended peptidic cores which are covered with long mucin-type O-glycans. O-glycosylated regions characteristically have high amounts of serin, threonin and prolin and are called Variable Number Tandem Repeats (VNTR). Most tethered mucins also contain an N-glycosylated (NG) region in the extracellular C-terminus. Glycosylated regions may be flanked by either cysteine rich regions (CYS), von Willenbrand factor like D domains (VWD) or both. Both features help in the polymerization of mucins in the extracellular space, either through covalent bonds (CYS) or non-covalent interactions (VWD). All mucins present an N-terminal signal peptide, shown in green, which targets the proteins for export through the plasma membrane. Tethered mucins are proteolyzed upon export in a specific proteolysis site (PS), which gives rise to either the corresponding secreted mucin or the mature tethered mucin. Mature tethered mucins are formed as heterodimers from the large, glycosylated subunit and the small, membrane bound subunit resulting from proteolysis of the precursor. Tethered mucins contain a single transmembrane helix (TH) and a C-terminal cytoplasmic tail (CT) with numerous phosphorylation sites. Drawing adapted from [6] and [11].

### 2.1.2 Mucin tissue localization

Out of the ~20 identified mucins, only a small subset is to be found in any given tissue[6]. There has been extensive research to determine the localization of the different mucin gene products[22][23][24] and a general tissue specificity has been determined (Table 2.1). Mucins are extremely variable in their splicing and glycosylation pattern, and these two variables seem to be also tissue dependent and of an even greater importance than the mere coding gene for function and tissue recognition[16]. Unfortunately, our knowledge of the mucin glycosylation and splicing patterns remains sketchy at best[6]. It has been shown that healthy mucins are mostly glycosylated with core 2 mucin-type O-glycans, while neoplastic mucins tend to present core 1 mucin-type O-glycans (Figure 2.2)[25], a general distribution for sialylated and sulfated mucins has been established and several carbohydrates have been characterized as being presented by different mucins (Table 2.2)[7], but until now, the tissue specificity of carbohydrate distribution has remained elusive.

Several pathogens have been shown to interact specifically with the human host tissue, including *H. pylori*[4] and *C. glabrata*[8]. Even though *C. glabrata* is relatively unspecific in the type of mucosa it invades, including oral, vaginal and urethral (Table 2.1), the surface-displayed adhesin array changes strongly between the different niches[26][27]. Such polymorphic behavior hints at the

possibility of very specific pathogen-mucin interactions, which need to be varied depending on the mucin catalog presented by each tissue. On the other hand, *H. pylori* is very specific in just colonizing the antral mucosa and, very rarely, other regions of the stomach (Table 2.1)[4]. Closely related species, like *Campylobacter jejunii*, do not share adhesin orthologes with *H. pylori*, and correspondingly do colonize different tissues in the human gastrointestinal tract[28]. Tissue localization and glycosylation patterns of mucins seem therefore crucial in the elucidation of the adhesion specificity of several pathogens. Conversely, the information obtained from adhesins specifically targeting certain mucins in certain tissues might provide information on the glycosylation patterns of the human mucosal proteins.

Human tissues in which sulfo-mucins have been detected
--

Submandibular glands

Sublingual glands

Lung tissue

Esophagus

Stomach

Small intestine

Colon

Uterus

Prostate

Table 2.1: Sulfo-mucin positive tissues in *Homo sapiens* [7]. Sulfo-mucin mediated adhesion has been proven to be of great importance for *H. pylori*[29]

Mucin	Family	Tissue specificity	Possible target of
MUC1	both	ubiquitous	<i>C. glabrata</i> <i>H. pylori</i>
MUC2	secreted	Colon, middle ear epithelium	
MUC3	tethered	Gastrointestinal and middle ear epithelia, male urogenital epithelial	<i>C. glabrata</i>
MUC4	secreted	Male urogenital and middle ear epithelia, cervix, vagina	<i>C. glabrata</i>
MUC5	secreted	Male urogenital and middle ear epithelia, cervix, saliva	<i>C. glabrata</i>
MUC6	secreted	Cervix, endometrium, uterus, male urogenital and middle ear epithelia, parotid glands, heart, fundic neck cells, pyloric glands, duodenum, jejunum, ileum, gall bladder glands, common bile duct, pancreas.	<i>C. glabrata</i> <i>H. pylori</i>
MUC7	secreted	middle ear epithelia, saliva	<i>C. glabrata</i> <i>H. pylori</i>
MUC8	secreted	Sperm, respiratory tract, endometrium, cervix, middle ear epithelia.	<i>C. glabrata</i>
MUC9	secreted	Oviduct middle ear epithelium	<i>C. glabrata</i>
MUC10	secreted	Saliva	<i>C. glabrata</i> <i>H. pylori</i>
MUC11	secreted	Middle ear epithelium, colon and stomach	<i>H. pylori</i>
MUC12	tethered	Middle ear epithelium, colon and stomach	<i>H. pylori</i>
MUC13	tethered	Middle ear epithelium, gastrointestinal and respiratory tracts, male urogenital epithelia	<i>C. glabrata</i> <i>H. pylori</i>
MUC14	secreted	Heart, kidney and lung	
MUC15	tethered	Milk fat globules, lungs, gastrointestinal tract, mammary gland, hematopoietic tissues, gonads, middle ear epithelium	
MUC16	secreted	Middle ear epithelium	
MUC17	tethered	Gastrointestinal tract, male urogenital epithelia	<i>C. glabrata</i> <i>H. pylori</i>
MUC18	secreted	Middle ear epithelium	
MUC19	secreted	Middle ear epithelium, saliva.	<i>C. glabrata</i> <i>H. pylori</i>
MUC20	tethered	Proximal renal tubuli, male urogenital and middle ear epithelia.	<i>C. glabrata</i>

*Table 2.2: Tissue specific distribution of mucins. Historically, MUC has been used to denominate human mucins, while muc is the denomination of non-human, mammal mucins. All specificities have been compiled for Homo sapiens. According to the tissue specificity of the pathogen, the mucins have also been assigned as possible targets for C. glabrata (yellow), H. pylori (blue) or both (green). Data compiled from: [6][15][22][30][31][32][33][34].*



### 2.1.3 Mucin glycosylation

Mucins show very specific glycosylation patterns in the VNTR region, the so called mucin type O-glycans[36]. Mucin-type O-glycans are always bound to either a serine or a threonine residue through an N-acetyl-galactosamine reductive end[11]. Mucin type O-glycans can be classified via their core, which commonly corresponds to a di- or trisaccharide that is bound by its first N-acetyl-galactosamine (Fig. 2.2)[25] [11]. Beyond the cores, several relevant antigens, relating to blood groups, amongst others, have been identified as forming part of the branched O-glycans (Figure 2.7)[37][38][39][35]. Finally, some tethered mucins contain a few sites for N-glycosylation along their C-terminal, extracellular domains (Figure 2.2 f))[6][40].

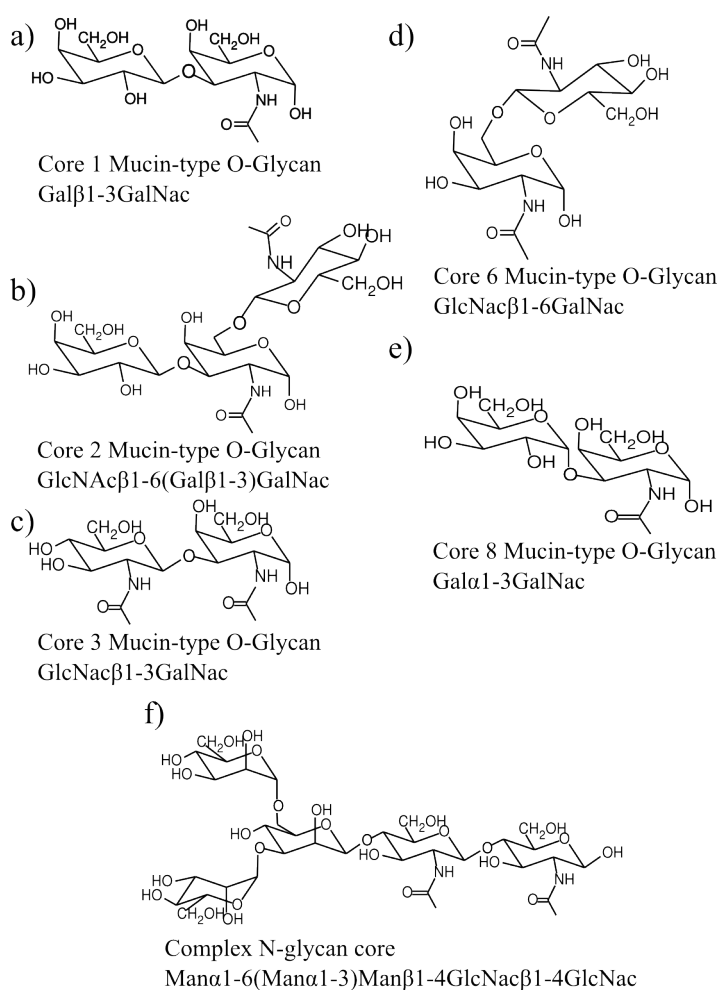


Figure 2.2: Core oligosaccharide structures present in Mucins. a) – e) Examples of mucin-type O-glycan cores. f) Complex N-glycan core common to all N-glycans.  $\beta$ 1-3 glycosidic bonds such as the ones presented in a) and c) are preferred by the *C. glabrata* adhesin Epa1. [11][35].

Even though the glycosylation patterns of mucins remain largely unknown, knowledge derived from the binding specificities of adhesins presented by pathogens might help in understanding what the crucial determinants are for tissue specific adhesion. For example, it has been shown that the Epa1 adhesin from *C. glabrata* has a marked preference for  $\beta$ 1-3 galactosides[41], pointing at a preference for core 1 and core 3 O-glycans (Figure 2.2 a and c). On the other hand, *H. pylori* pathogenicity and the ultimate disease produced by the organism has been linked to the presence of several blood group antigens, especially the Lewis antigens (Figure 2.3). Polymorphisms in the blood group antigen binding adhesin (BabA) are determinant in the successful colonization of *H. pylori* in human populations where the O-blood type is more strongly presented[42]. The intimate co-evolution that is inferred from these observations indicates that the study of pathogenic adhesion molecules needs to be linked to the study of the glycosylation of mucins, but that the opposite is

also true, the study of pathogenic adhesion molecules may help in the mapping of the glycosylation patterns within the human body.

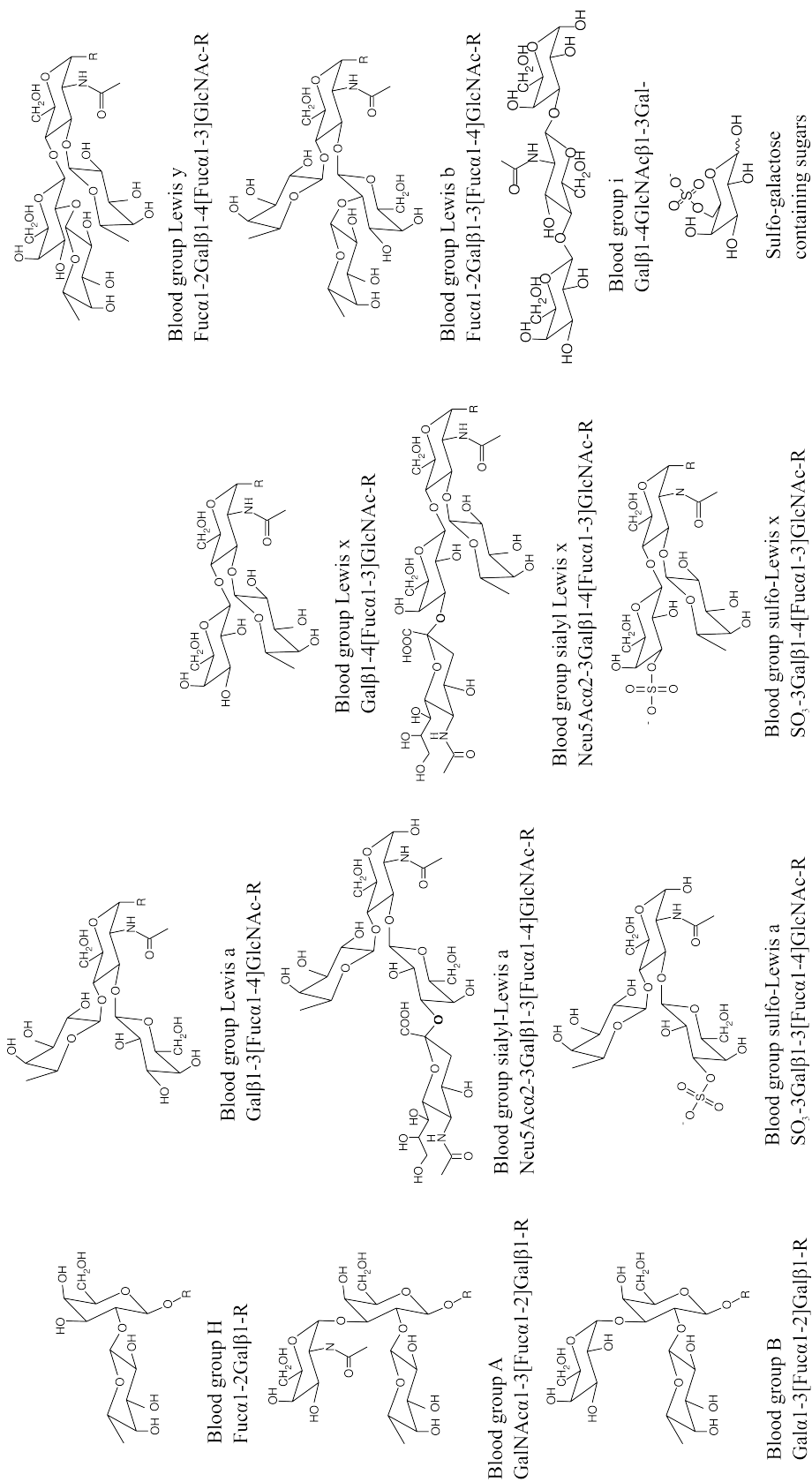


Figure 2.3: Blood group related oligosaccharides found on mucins as branches in O-glycans. Lewis antigens have been linked to *H. pylori* colonization, although the B and A antigens also seem to be important for it [11][7][40].

## 2.2 *Candida glabrata*

*C. glabrata* is a strict mammalian commensal fungus[43], which can act as a health care associated opportunistic pathogen, normally on immunocompromised patients, with severe results, related to systemic infections and sepsis[44]. *C. glabrata* can also be the causative agent of several less severe, mucosa related, oral or urogenital diseases[45][46]. Due to its relative resistance against traditional antimycotics[27], the pathogen is considered to be highly emergent[45], and new therapeutic tools are being actively sought. One of these therapeutic targets is the epithelial adhesin (Epa) machinery, which is represented in *C. glabrata* by up to 23 different genes, of which the Epa1 gene is considered to be responsible for adhesion of fungal cells to mammalian surfaces *in vitro*[47].

### 2.2.1 *The genetic background of C. glabrata*

The *C. glabrata* genome is 12.3 Mbp long, and is divided into 13 chromosomes[48]. The genome shows a relatively high similarity to the baker's yeast *S. cerevisiae*, with paralogous genes sharing around 65% average sequence identity[48]. In reflection of its commensal life-style, *C. glabrata* has lost many metabolic genes. Therefore it is auxotrophic for at least nicotinic acid, thiamine and pyridoxine[43]. Despite *C. glabrata* being mainly haploid, mating loci have been found in its genome[49], which together with non-clonal recombination, point to the possibility of seldom occurring sexual reproduction[50].

An interesting characteristic of the *C. glabrata* genome, which is shared with *S. cerevisiae* but not with *C. albicans*, is the fact that a large part of the adhesion apparatus coding genes are located in sub-telomeric regions, where they are subject to regulation of expression by gene silencing[43][51]. It has been proposed that the proteins of the Sir silencing machinery, Rap1 and Rif1 are involved in the silencing of epithelial adhesin coding genes (Epa genes). In an analogous model to *S. cerevisiae* Rap1 and Hdf1 are telomere associated and recruit the proteins encoded by the genes *SIR2*, *SIR3* and *SIR4*[52]. The Sir2 protein is an NAD<sup>+</sup>-dependent histone deacetylase that increases the positive charge of histones H3 and H4 through deacetylation of amine groups, thus increasing attraction between the nucleosomes and effectively compacting the chromatin[53][54](Fig. 2.4). An NAD<sup>+</sup>-dependent Sir silencing model for the regulation of expression of Epa genes is further strengthened by the observed behavior of *C. glabrata*, which, under limiting concentrations of nicotinic acid, immediately increases the expression of Epa6[26]. As previously mentioned, *C.*

*glabrata* is an auxotroph for nicotinic acid, which is a direct precursor of  $\text{NAD}^+$ . At limiting nicotinic acid concentrations, insufficient  $\text{NAD}^+$  would be directed to the Sir machinery, which would not be able to counteract background acetylation by acetylases like Sas2[55]. Predominant acetylation would then initiate a de-compaction of the subtelomeric chromatin, and an increased transcription rate of the *Epa* genes (Fig. 2.4).

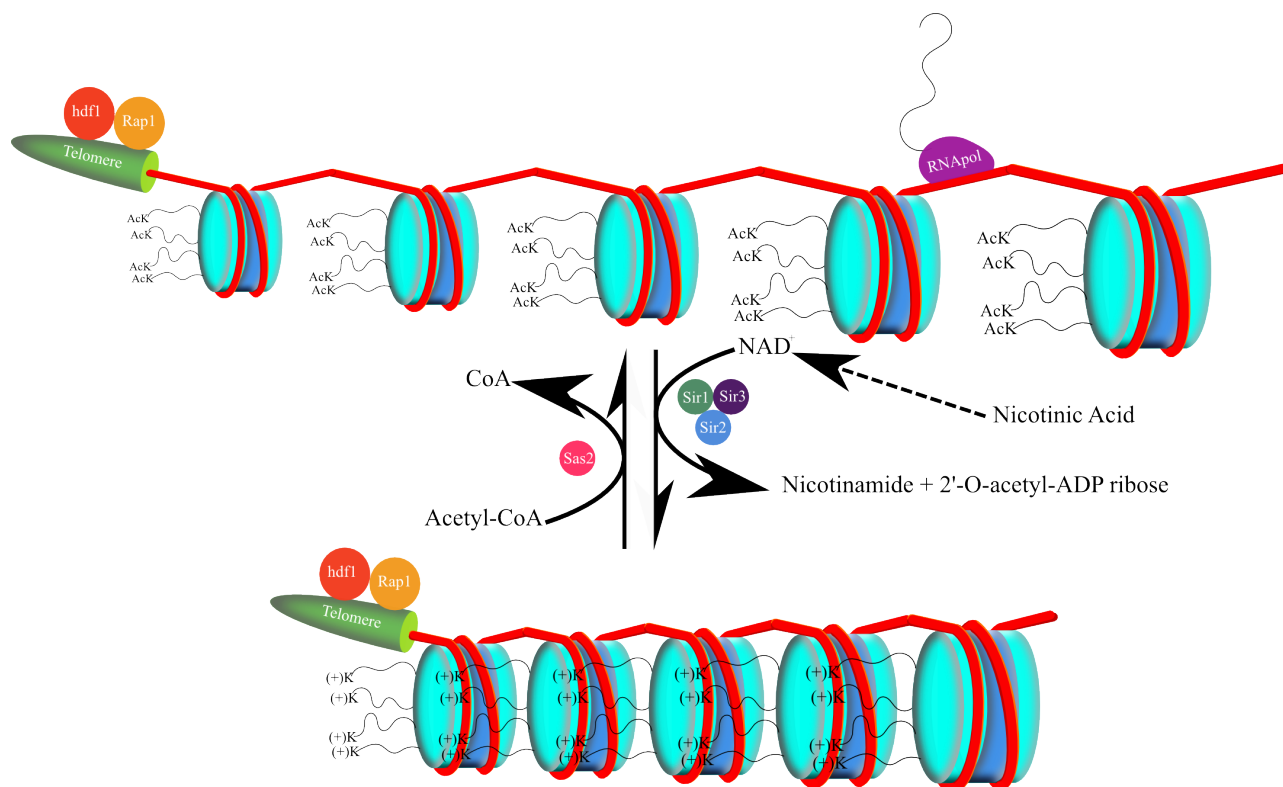


Figure 2.4: Mechanism for the regulation of gene expression through silencing. Histones (light and dark blue drums) form octameres around which DNA coils, forming nucleosomes. Histones contain N-terminal tails which are lysine rich, and which interact either with DNA or with other histones, yielding very compact heterochromatin[54]. Sas2 is a histone acetylase, which consumes Acetyl-CoA to acetylate N-terminal lysines in histones. The result is euchromatin, which can be transcribed, as shown in the picture by the RNAPol. On the other hand, the Sir trimer acts as a deacetylase, and is recruited to subtelomeric regions by *hdf1* and *Rap1*[52]. While the activity of Sas2 is apparently constitutive, the Sir complex is highly dependent on the presence of  $\text{NAD}^+$ , which it consumes during de-acetylation[55]. Since *C. glabrata* is a nicotinic acid (NA) auxotroph, it depends on external sources of it to maintain  $\text{NAD}^+$  levels. If insufficient quantities of NA are supplied,  $\text{NAD}^+$  will become limiting in deacetylation, and therefore Sas2 mediated acetylation will yield more euchromatin in subtelomeric regions. The result in *C. glabrata* would be an increased expression of *Epa* genes, as observed [26].

### 2.2.2 The cell wall of *C. glabrata*

All yeasts have a cell wall, an outer shell composed of different polysaccharides, glycolipids and glycoproteins (Fig. 2.5), which protects and shapes the cell, and interacts with the environment[11]. In the case of *C. glabrata*, the cell wall is around 100 – 200 nm in thickness, shows two electron dark regions sandwiching an electron translucent mid-section in electron micrographs, and corresponds to around 19% of the dry weight of the cell[56]. This layered arrangement is typical of ascomycetous yeasts[57]. In *S. cerevisiae*, the outer, dark layers are composed mainly of mannoproteins and other glycoproteins, while the inner, translucent layer corresponds to a polysaccharide rich region[58][59].

The cell wall of *C. glabrata* contains mainly  $\beta$ 1-3- and  $\beta$ 1-6-glucans with traces of chitin (Figure 2.5 b, c and a, respectively) and an unusually high amount of mannoproteins ( Figure 2.5 d and e) [56]. There appears to be less cross-linking between the glucans and chitin, with a denser mannoprotein outer layer, than in *C. albicans*[56]. It has been suggested that such an arrangement might facilitate host-evasion by reducing recognizance via the dectin-1 host receptor, as is the case with *C. albicans*[60].

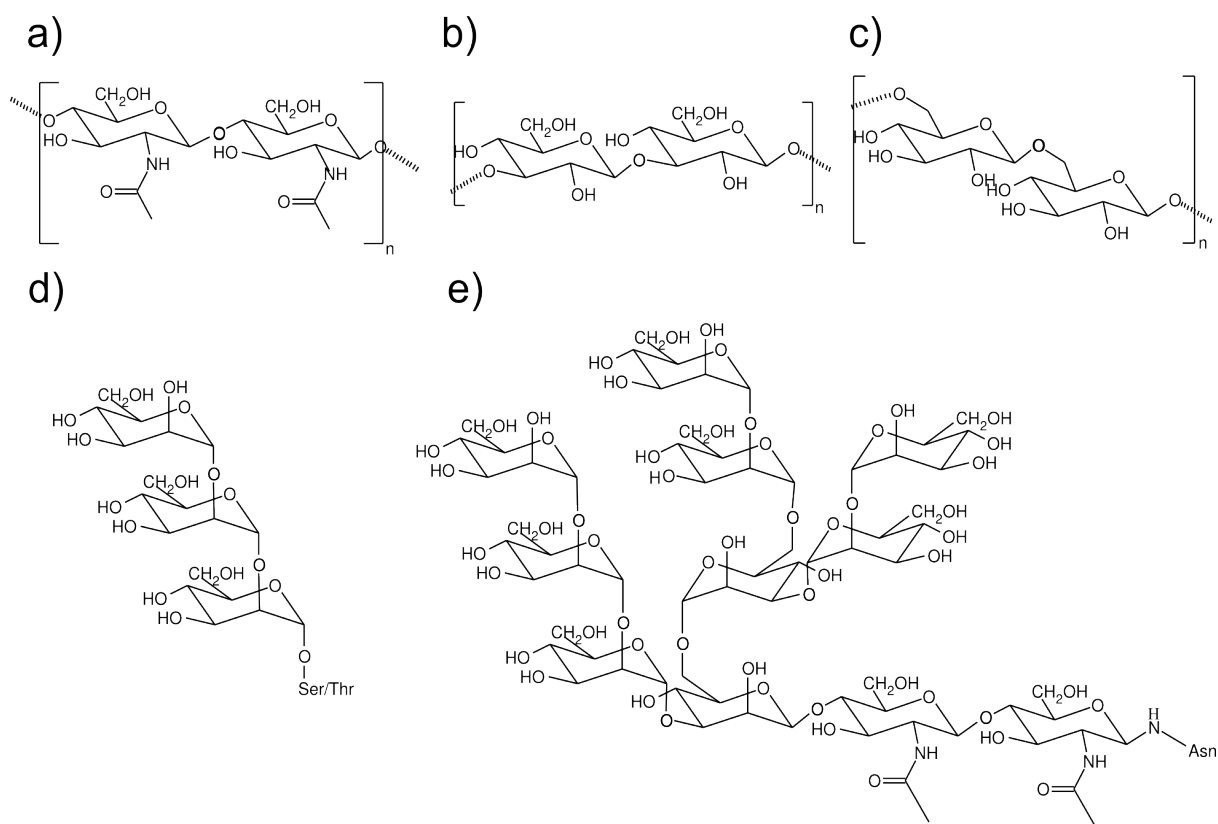


Figure 2.5: Glycans found within the cell wall of *C. glabrata*. a) Chitin b)  $\beta$ 1-3-glucan c)  $\beta$ 1-6-glucan d) O-mannan and e) N-mannan commonly found in fungal glycoproteins.

The mannoprotein layer of the *C. glabrata* cell wall has proven to be highly dynamic, with proteins both GPI-anchored and cross-linked to the  $\beta$ 1-3-glucan[56]. Among these, many are structural proteins that cross-link the cell wall, as demonstrated for an orthologue of the Pir protein family found in *S. cerevisiae*[57], but there are also many proteins that are crucial in pathogenicity and virulence, which are expressed on the cell surface, especially adhesins. The adhesin Epa1 is presented on the surface of the cell wall at very early stages of cell growth, and is then cleaved by aspartic proteases[61], while *EPA6* is expressed during the stationary phase *in vitro*[56]. Under limiting nicotinic acid conditions, *EPA6* along with *EPA7* and *EPA1* is transcriptionally active[26]. In general, it has been shown that *EPA* genes remain relatively silent under non-limiting growth conditions[51], but that any kind of danger or limiting conditions, including the presence of antimycotics[27], can lead to their activation and presentation on the cell wall.

The cell wall of *C. glabrata* is therefore a highly dynamic structure, with the capacity to both adapt to a commensal life-style or a pathogenic one, and even to regulate the probable binding specificity and hydrophobicity of the cell surface according to the ecological niche occupied at any given moment.

### 2.2.3 The adhesins of *C. glabrata*

Adhesion plays a fundamental role in the pathogenesis of both *C. glabrata* and *C. albicans*, as it is the first step in the colonization of the mucosal surface[8]. In both cases, adhesion is mediated by GPI-anchored mannoproteins called adhesins, of which the Als (agglutinin-like sequence) family in *C. albicans* and the Epa (epithelial adhesin) family in *C. glabrata* are some of the best characterized examples. Both protein families share a very similar modular structure -with N-terminal domains responsible for the adhesive capabilities (A domain), central serine and threonine rich regions which are highly glycosylated *in vivo* (B domain)- and maintain an extended structure and a C-terminal GPI anchor which cross-links the protein to the cell wall (C domain)(Fig. 2.6 a))[47][62].

The Ser/Thr rich region (B domain) is not directly involved in adhesion, but is highly glycosylated and maintains the N-terminal adhesive domain well outside the cell wall, allowing it to interact with its ligands[63]. The N-terminal adhesive domain (A domain) of all members of the Epa family has been characterized as being a PA14 domain[64][65] involved in sugar binding in several adhesins and sharing a very similar domain architecture in *S. cerevisiae*, as typified by Flo1p[1][62].





Recent, semiquantitative functional characterization of Epa1, Epa6 and Epa7 shows that the specificity for these three proteins is similar, but not identical. While all three proteins require a terminal galactose, Epa6 can bind to carbohydrates where the galactose is bonded either  $\alpha$  or  $\beta$  over a 1-3 or 1-4 glycosidic bond to glucose or galactose or their N-acetylated derivatives. In contrast, Epa1 shows a narrower specificity, mainly excluding the  $\alpha$ -bonded carbohydrates. Finally, Epa7 is the most specific, almost exclusively binding galactose $\beta$ 1-3galactose or lactose (galactose $\beta$ 1-4glucose)[41].

The large number of epithelial adhesins present in the *Candida* genome, their differential expression and specificity point towards a specialization of Epa function. This specialization could be tissue or situation specific, or both.

#### 2.2.4 Pathogenicity of *C. glabrata*

*C. glabrata* is responsible for around 15% of all infections caused by the *Candida* genus[43]. There have not been many comprehensive attempts to set all virulence factors of *C. glabrata* in a whole-organism context. Instead, many isolated pieces of evidence point towards a generally similar pathogenicity mechanism as *C. albicans*[67]. Major similarities are the large adhesin array, composed of the Epa family and others in *C. glabrata*[43], as well as the presence of secreted aspartic proteases[61], which have been described as very important for *C. albicans* pathogenicity[68]. The capacity for phenotype switching, which allows the *Candida* population to change its outward appearance without any major genotypic events, and is therefore crucial in host evasion, is also a feature present in both organisms[69][70].

Major differences, on the other hand, include the fact that pseudohyphal forms, although possible[71] in *C. glabrata*, could not be isolated from clinical samples[43]. In *C. albicans*, the filamentous form dramatically increases invasiveness, and is considered to be essential during pathogenic colonization[72]. *C. glabrata's* tightly regulated adhesion machinery[51] is also an important difference, as *C. albicans* has basically a much more adhesive cell wall[73]. Finally, it seems that secreted proteases and phospholipases, although genotypically present[61][74], are not as important in *C. glabrata* as in *C. albicans*, as no significant activities could be found extracellularly[74].

The mannose rich cell wall of *C. glabrata* could also be an important difference, as it has been proposed to be more difficult to identify by the host immune system[60]. Finally, *C. glabrata* has

been characterized as being naturally resistant to azole based antimycotics, due to an ABC efflux transporter[75].

A basally less adhesive cell wall might be responsible for the lower infection rate of *C. glabrata* compared to *C. albicans*. On the other hand, the mannoprotein-rich cell wall and the natural antimycotic resistance might account for the more severe nature of *C. glabrata* infections, and the increased difficulty in treating them[67]. The importance of the different regulation of adhesin expression and the significance of proteases and phospholipases have not yet been assessed.

## 2.3 *Helicobacter pylori*

*H. pylori* is a flagellated Gram-negative eubacterium with helical morphology, which was discovered in 1982 by Drs Marshall and Warren in Perth, Australia[76]. Their results challenged the then-established paradigm for the etiology of peptic ulcer, and were rewarded with the Nobel prize for medicine in 2005. It is the only bacterium to persistently colonize the human gastric mucus[77], and infects around 50% of the human population[78][79]. Even with such a large part of humanity infected by *H. pylori*, only ~15% of carriers will develop peptic ulceration (~80% of all peptic ulcers), and ~0.2 – 0.3% gastric adenocarcinoma[4]. Mucosa associated lymphoid tissue (MALT) can also develop from any ulcerous process, but if no malignancy is developed, it normally recedes with successful antibacterial treatment[4]. The reasons why some carriers develop a disease, while others do not, and whether the pathological process develops into either peptic ulcer or cancer is yet unclear, although it appears that one type of process protects from the other (i.e. *H. pylori* induced peptic ulcer patients do not suffer gastric adenocarcinoma)[80].

### 2.3.1 Colonization and pathogenesis of *H. pylori*

*H. pylori* probably colonizes the human stomach during the host's early childhood[4]. In the rare cases of colonization during adulthood, gastritis accompanied with hypochlorhydria ensues[81][82]. It is not known whether childhood infection carries any changes in gastric homeostasis. No colonization of the esophagus or the duodenum is possible unless they have undergone gastric metaplasia[4]. Accordingly, there only appears to be an active niche within the stomach and an oral reservoir.

In a first step of colonization, *H. pylori* must chemotactically (Fig. 2.7) locate its target within the stomach. One CheA histidine kinase and two CheY regulator proteins have been isolated in *H. pylori*, and they are essential for chemotaxis and colonization of the gastric mucosa[83].

While swimming in the gastric juices, *H. pylori* needs to be able to survive the extremely harsh environment, specifically, it must neutralize the low pHs of its surrounding. As proof of the importance of acid-dependent regulation of all *H. pylori* functions, it could be shown that more than 300 genes are acid-regulated or acid-affected[84]. Among these, the gene coding for the most abundant protein within *H. pylori* can be found, namely, urease. Urease is an intracellular enzyme which hydrolyzes urea to ammonia and carbon dioxide. The ammonia raises the pH within the cytoplasm, which buffers the periplasm. The result is a strong buffering system surrounding

*H. pylori*, which encourages survival at extremely acidic pHs. Urea import is controlled by a pH-gated urea channel (UreI), which allows non-genetic, direct coupling of pH and urease activity[85][86].

Once the mucus producing cells of the antral region are located, *H. pylori* will bind to the mucus directly above them, mainly to mucins MUC1 and MUC5A[87]. The largest part of the *H. pylori* population will grow within the mucus layer, and only a small part, if at all, will actually attach itself to the epithelial cells forming the antral mucosa[11][88][89]. Attachment and adhesion during the process of colonization and invasion of the mucous and epithelial layers of the mucosa is very complex, and requires very distinct modes of adhesion[90][91].

If the bacterium manages to colonize the epithelial surface, it is possible for a pathogenic process to be initiated. The factors involved in the development of illness and the corresponding symptomatology are very diverse, depend both on the bacterium and on the host, and are not yet fully understood. Amongst the better understood host factors are the *mucl* genotype, where short *mucl* alleles could be linked to *Helicobacter*-linked gastritis and gastric adenocarcinoma[92]. Murine *mucl* knockouts showed a much higher propensity towards severe *H. pylori* infections and were easier to colonize[93]. Another important host factor is the inflammation-dependent differential glycosylation of the gastric mucins. Upon contact with *H. pylori*, the epithelium releases an innate immune response. As part of the response, the glycosylation patterns of the gastric mucins change, with an increase in sialic acid-containing glycoconjugates. *H. pylori* is capable of adapting to such changes by differentially expressing different sets of adhesins[91][94].

The bacterial factors leading to virulence and severe disease are much better understood. The cytotoxin associated gene pathogenicity island (*cag* PAI) is a large ensemble of approximately 30 genes obtained by *H. pylori* through an unknown source and inserted within its glutamate racemase gene[95]. Different *H. pylori* strains may contain the entire PAI (*cag*+), none (*cag*-) or only part of it, which normally leads to a *cag*- phenotype[96].

The *cag* PAI is composed of genes that code for a secretory apparatus[95] that allows *H. pylori* to inject the *cagA* gene product (also part of the *cag* PAI) into epithelial cells[97]. Once within epithelial cells, CagA acts as a target for host cell Src family tyrosine kinases[98]. Phosphorylated CagA then activates several other regulatory proteins within the MAP-kinase pathway leading to cell proliferation[99], motility and cytoskeletal changes[100]. The most potent effect of CagA injection is the elongation of the epithelial cells in the so called hummingbird phenotype[100].

Interestingly, MAP-kinase activation has been linked to c-Fos, c-Jun[99] and the c-Met receptor[101], which are all pro-oncogenes. CagA can also disrupt gap-junctions between epithelial cells, decreasing tissue cohesiveness and probably facilitating nutrient uptake[102]. The secretory apparatus itself also causes an increased production of interleukin IL8, recruiting a higher number of neutrophils and leading to an enhanced inflammatory response to any cag<sup>+</sup> strain[103]. The *cag* PAI is necessary for disease, but not sufficient, as some cag<sup>+</sup> strains are minimally pathogenic, as CagA is polymorphous and may contain no phosphorylation sites[103].

The other very important pathogenicity factor of *H. pylori* is the vacuolating cytotoxin A (*vacA*). VacA is produced in *H. pylori* as a pre-toxin, which is secreted by an autotransporter mechanism[104][105][106]. The toxin is then subjected to N- and C-terminal cleavage, generating a functional heterodimer[107]. VacA's toxicity relies on it binding several epithelial factors, the receptor protein tyrosine phosphatase (RPTP)  $\alpha$  and  $\beta$ [108], the epidermal growth factor receptor (EGFR)[109] and a (GPI)-anchored protein associated with lipid rafts[110]. Interaction of VacA with a receptor is followed by acid-dependent internalization through endocytosis[111]. In an acidic pH dependent manner[112], VacA is activated after internalization, and oligomerizes into hexameric pores[113]. The pore is selective to anions and small neutral molecules[114], and causes osmotic swelling of late endosomal compartments, which become large acidic vacuoles[115]. Vacuolation is only observed *in vitro*, however, and it is not clear what the mechanism for toxicity is within the gastric mucosa, although it appears to affect both the epithelium and immune cells[92]. VacA is also highly polymorphic, with so-called S2 variants yielding minimally pathogenic strains, while S1 variants vary in pathogenicity, but are always pathogenic to some extent[116].

Pathogenicity in *H. pylori* is very complex, but also much better studied than in *C. glabrata*. *H. pylori* is extremely well adapted to an environment where pH ranges are extreme, going within a few microns from extremely acidic to neutral. Its virulence factors reflect such an adaptation, while they also show a tight co-evolution with its human host.

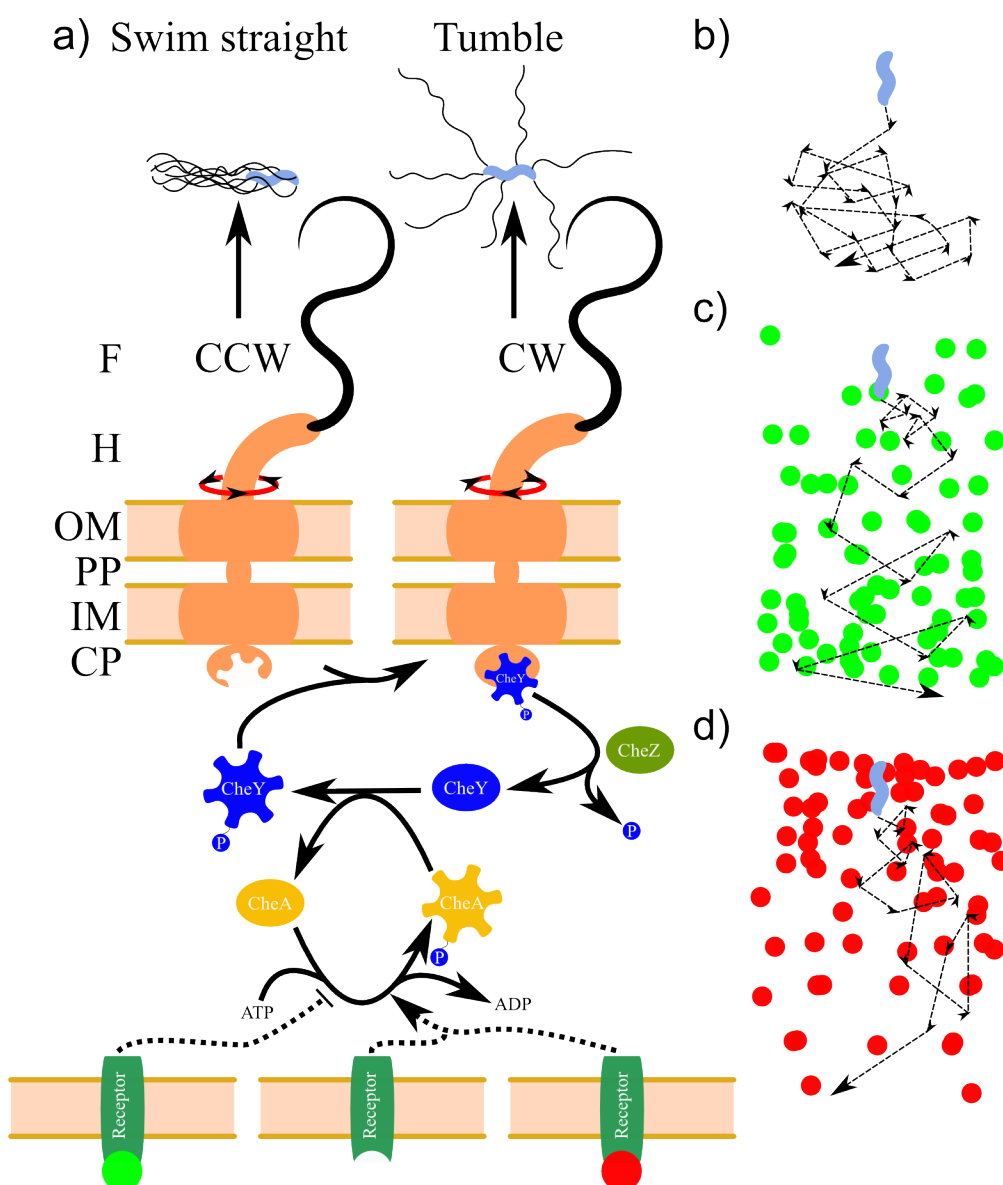


Figure 2.7: *CheA/CheY dependent Chemotaxis.* a) The flagellar apparatus (orange), intracellular regulatory cascade and corresponding receptors. The flagellar apparatus consists of a cytoplasmic (CP) regulatory region, an inner (IM) and an outer (OM) membrane regions, a periplasmic (PP) region, the Hook (H) and the flagellum (F). Due to the orientation of the Hook, whenever the flagellum rotates counterclockwise (CCW), the bacterium swims straight. If the flagellum rotates clockwise (CW), then the flagella separate, and the bacterium initiates a tumble. Whenever CheY phosphate is bound to the regulatory region, the flagellum turns clockwise, causing a tumble. CheY phosphate is produced by transfer of a phosphate group from CheA phosphate, which is itself an autophosphorylating kinase, and depleted through CheZ. Chemotactic receptors act on the transduction cascade depending on their state. When free, they activate CheA activity, yielding basal tumbling (b). When bound to an attractant (green), they inhibit CheA activity, increasing straight swimming paths with increasing attractant concentrations (c). If they bind a repellent (red), they up-regulate CheA activity beyond basal levels, causing more tumbling at higher concentration levels, and longer paths at lower concentrations (d). Upon constant activation, chemotactic receptors become desensitized, and tumble frequency returns to basal levels.

### 2.3.2 Adhesion in *H. pylori*

Adhesion in *Helicobacter pylori* has been proven to be one of the key virulence factors, but it appears that the adhesion machinery of *H. pylori* is very complex and polymorphic, with a high recombination and modification rate[117]. It seems that the binding to the Lewis b (Le<sup>b</sup>) antigen present on gastric mucins (section 2.1) is of special importance during colonization, as *H. pylori* strains lacking Le<sup>b</sup> binding capacities are less prone to produce any of the described diseases[118][119].

The blood group antigen-binding adhesin (BabA) is mainly responsible for the binding of the Le<sup>b</sup> antigen[120]. The BabA protein is highly polymorphic, and has been shown to lose and regain its Le<sup>b</sup> binding activity repeatedly during infection[120][121][122]. Even in certain human populations where Le<sup>b</sup> secretors are very rare, it has been shown that BabA has adapted to bind other, more common ligands[42]. The rapid and functional modification of BabA has been traced back, at least in part, to three gene loci *babA1*, *babA2* and *babB*, of which only *babA2* is expressed. *babA2* can rapidly recombine with *babB* and perhaps *babA1* to lose or gain function[121].

BabA is ~80 kDa and ~750 amino-acid long outer membrane protein (OMP) of the HP\_OMP family[123], which is composed of a C-terminal  $\beta$ -sheet rich region and an N-terminal globular domain, that is putatively considered to be responsible for adhesion. The C-terminal domain fulfills all prerequisites of a typical transmembrane  $\beta$ -barrel like structure typical of OMPs[89][124]. The N-terminal domain is composed of a very conserved N-terminal sequence and a C-terminal region of high variability[42]. Insertion into the outer membrane of the C-terminal domain would leave the N-terminal domain within the periplasm, which further supports the claim that BabA works as an autotransporter[105][106][125][120].

Secondary to BabA-Le<sup>b</sup> mediated adhesion, the Lewis X (Le<sup>x</sup>) antigen has also been observed to be an important ligand to *H. pylori*[126]. Especially in its sialylated form, Le<sup>x</sup> is bound by the sialic acid binding adhesin (SabA), which is very similar in biophysical characteristics to BabA. SabA is apparently most important after the unspecific immune response has started, as Le<sup>x</sup> expression is up-regulated in inflamed tissue[91]. Other putative adhesins, like AlpA and AlpB, might also be involved, but it is not quite clear what their cognate host ligands are[91].

Other modes of adhesion have been identified for *H. pylori*, including one that is mediated by electrostatic interactions and which only happens at low pH[90] for which no single protein has been identified, and one in which the LeX and LeY antigens are involved. These glycans must be

present both at the bacterial lipopolysaccharide (LPS) and at the mucosal surface of the host[127]. Anti-LeX auto-antigens produced in response of *H. pylori* colonization have been implicated in this adhesin-independent adhesion, which is strong enough to reestablish wild-type infectivity in bab<sup>-</sup> mutants[127].

Adhesion of *H. pylori* has been shown to be highly pH dependent[87]. *H. pylori* almost exclusively colonizes slime producing cells of the gastric antrum[128], which is the least acidic region of the stomach. Also, *ex vivo* binding to pig gastric mucins was pH dependent[9]. Specifically for BabA, it could be demonstrated that BabA-dependent activity was only to be found at mildly acidic to neutral pHs, and not at extremely acidic pHs[90][87]. Therefore, BabA is important during epithelial adhesion in gastroduodenal disease[129] and for the maintenance of an oral reservoir[130], but not during travel of the bacterium within the gastric fluids[90].

The adhesion mechanisms of *H. pylori* are extremely complex, polymorphic and subject to strong variations, with apparently many rescue pathways to maintain adhesion to the host. Such a complex and tightly secured system indicates the extreme importance of adhesion in *H. pylori* infection.



## ***2.4 The Structural features of glycan binding proteins from pathogenic origin***

The surface of epithelial cells is covered by the glycocalyx, which is composed of different glycoproteins and oligosaccharides, including tethered mucins. The glycocalyx has long been proposed to be responsible for tissue specificity during pathogenic colonization, due to very specific recognition of glycan determinants by the invasive organisms. Genetic and functional studies demonstrated that microbial pathogens present several surface lectins, which were deemed responsible for glycan binding[131][132]. The structural characterization of the lectin-glycan interactions soon followed in some cases, but current information is still limited.

Even though the folds are very diverse, all known structures are  $\beta$ -strand rich, with  $\beta$ -sandwiches, either straight or in jellyroll fold[133], lectin-like domains[134],  $\beta$ -propellers [135][136]and ricin-B trefoil domains being the recurrent motifs[137](Fig. 2.8).  $\beta$ -propeller type lectins are multivalent, with many binding sites to procure a high avidity, and follow a similar mode of action as their animal and plant lectins counterparts. Most of the other microbial lectins are monovalent.[136] Accordingly, lectins from pathogenic organisms have, as a rule, a much higher affinity for their ligands than their non-pathogenic counterparts[138].

Within each fold type, structural conservation is quite high, but sequence identity can be as low as ~20%. Perhaps as a result, different lectins presenting the same fold do not necessarily bind their ligands in the same way. Even the position of the binding pocket within a single fold might not be well conserved, and can vary strongly from one protein to another[139].

Fungal flocculins, like Flo5, have a modular structure with an adhesive A domain at their apex (Fig. 2.6). It has been determined that the Flo5A domain is folded into a PA14 domain[1], which contains a jellyroll fold. It has been proposed that the Epa1 A domain will share that architecture and a similar binding mode with Flo5A. On the other hand, not much is known on the structural features of the BabA adhesin, besides it having an autotransporter modular architecture[106]. Secondary structure predictions indicate that the extracellular fragment is  $\beta$ -strand rich, in accordance with other bacterial adhesins, which bind blood group antigens[140].

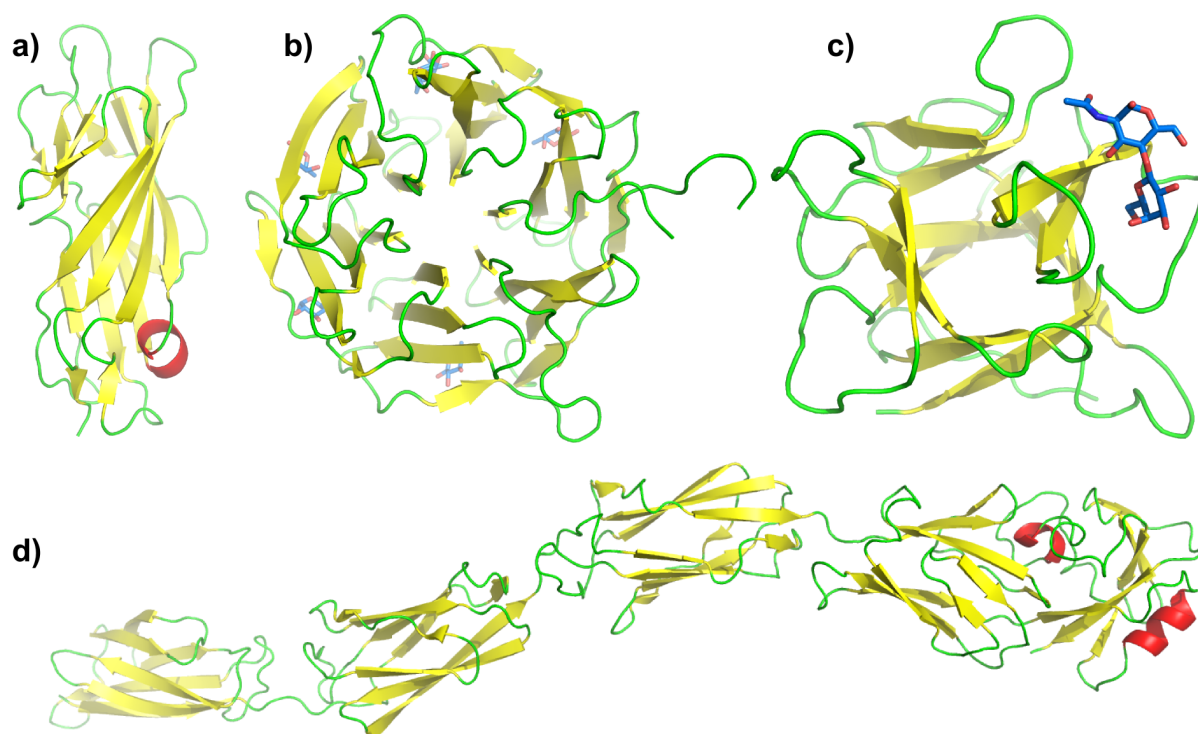


Figure 2.8 a) Jellyroll fold from *E. coli* FimH chaperone-usher adhesin[133]. (PDBID: 1QUN) b) Multivalent  $\beta$ -propeller lectin from the fungus *Aleuria aurantia* (AAL) [135](PDBID: 1OFZ). c) Trefoil fold of the LSL adhesin from *Laetiporus sulphureus* [137] (PDBID: 1W3F). d) the Invasin structure of *Yersinia pseudotuberculosis*, composed of 3 N-terminal Ig-like domains (left) and one C-terminal Lectin-like domain (right)[134] (PDBID: 1CWV)

#### 2.4.1 The PA14 domain

The PA14 domain was first discovered in the crystal structure of the anthrax protective antigen toxin (PA) (PDB accession code 1ACC)[141], but was defined much later[64]. The PA14 domain is contained within the PA20 pro-peptide of immature PA (Fig. 2.9 a), and is called PA14 due to the fact that it weighs 14 kDa. In *Bacillus anthracis*, PA is part of the trimeric anthrax toxin, which is composed by the protective antigen, the edema factor and the lethal factor[142]. *In vivo*, PA associates with the eukaryotic membrane as a monomeric immature protein, from which the PA20 pro-peptide is then cleaved. The resulting protein, PA63, oligomerizes into heptamers, which insert themselves into the eukaryotic membrane to produce a large pore, through which the edema and the lethal factors are translocated[141].

With the use of multiple sequence alignments and other bioinformatic methods[143][144], a

minimum theoretical domain structure within PA20 was defined[64], composed of amino-acids 43 to 179 from the PA (SWISS-PROT: P13423). This sequence corresponds to amino-acids 14-150 of the 1ACC PDB entry (shown in red in Fig. 2.9 a and c), and define the PA14 domain. The general fold described for the PA14 domain is a  $\beta$ -sandwich like structure with jelly-roll topology, composed of a total of 10 anti-parallel  $\beta$ -strands.

Although no carbohydrate binding capacity has been characterized for the original PA14 domain, there are many examples of glycan binding proteins which have been shown or are predicted to employ a PA14 domain for ligand binding[64]. All Epa proteins have been predicted to present PA14 domains as A domains[64]. There are several glycan binding proteins for which crystal structures are available showing the PA14 domain, like the  $3\beta$ -glucosidase from *Kluyveromyces marxianus*[145] or, more importantly, the A domain of the *S. cerevisiae flocculin* (Flo5A) (Fig. 2.9 b)[1].

The Flo5A structure is extremely interesting for the characterization of the Epa A domains, as the PA14 domain has been shown to be structurally quite well-conserved between them despite their very low sequence identity[1]. The Flo5A core structure is quite similar to the PA14 structure from 1ACC, as it is composed of 11 anti-parallel  $\beta$ -strands forming a  $\beta$ -sandwich. The main differences stem from the loops, the so-called subdomain and the calcium binding loops 1 and 2 (CBL1 and CBL2). The Flo5 sub-domain is composed of the loop spanning between  $\beta$ -strand 4 and  $\beta$ -strand 10 of Flo5A. It contains 5 short  $\beta$ -strands that interact with each other and 4 cysteines, which generate two disulfide bridges within the sub-domain structure (Fig. 2.8). The CBL1 and CBL2 loops are located between  $\beta$ -sheets 11 and 12 and  $\beta$ -sheets 15 and 16 respectively. CBL1 contains the unique *DcisD* motif (Fig. 2.9), which presents an unusual cis peptide bond between two aspartates, which enables them to perfectly coordinate a calcium cation required for carbohydrate binding[1]. CBL2 is positioned below the same calcium cation and interacts with it through the loop's peptide backbone[1]. It is within these four features (the subdomain, CBL1, CBL2 and the calcium cation), that the carbohydrate ligand binds. Flo5A has been shown to present affinity for  $\alpha$ 1-2mannobiose containing oligosaccharides in the mM range[1] (Fig. 2.9).

Interestingly, no subdomain is predicted to be present in Epa1A (Fig. 2.9), but the specificity defining penta-peptide present within Epa1A[41] aligns well with the Flo5A CBL2[1]. Additionally, although no *DcisD* motif can be found in the PA14 domain, and indeed the DN peptide substituting it is in a trans conformation in PA14 (Fig. 2.9), the motif can be found and

aligned within Epa1. Both pieces of evidence point to a similar protein-carbohydrate interaction for Epa1 and Flo5A.

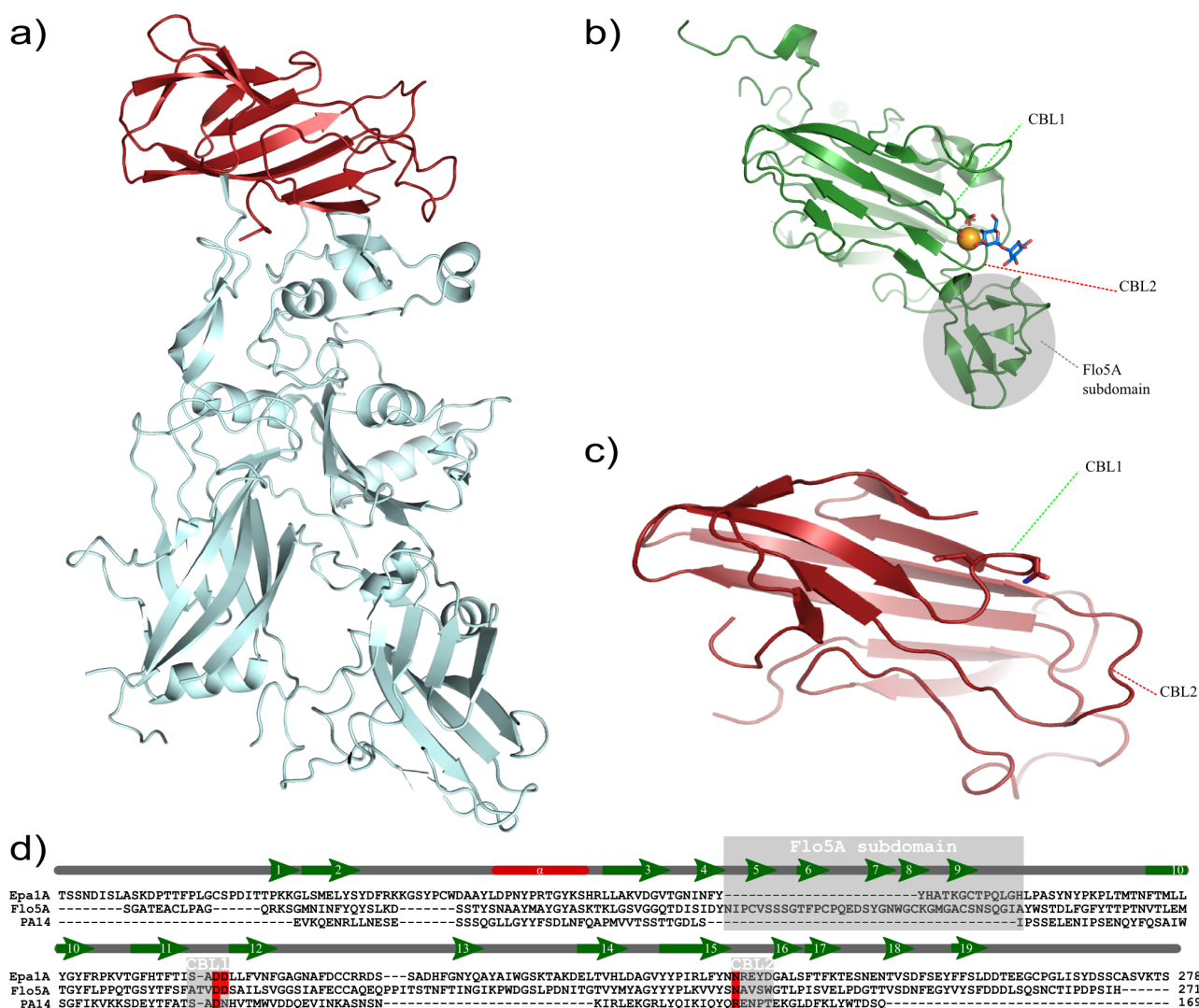


Figure 2.9: PA14 domains in different proteins. a) The *B. anthracis* protective antigen. The PA is an 83 KDa protein divided into 4 different domains. Marked in red is the PA14 domain, which is part of the PA20 propeptide, which is in turn part of domain 1 of the immature PA. The PA14 domain shows a  $\beta$ -sandwich like structure with jelly-roll topology, composed of a total of 10 anti-parallel  $\beta$ -sheets [64]. b) The crystal structure of the *S. cerevisiae* flocculin Flo5A domain (Flo5A). The structure of Flo5A is strikingly similar to that of the PA14 domain, despite their poor sequence similarity (Fig. 2.8 c) and d). It presents a Flo5A subdomain, which is marked in grey and the calcium binding loops 1 and 2 (CBL1 and CBL2). CBL1 and CBL2 bind calcium (shown in orange), CBL1 through two aspartates joined by a cis peptide bond (the DcisD motif) and CBL2 through its peptide backbone. The combination of the subdomain, CBL1, CBL2 and the calcium cation form the binding pocket. In the present structure, Flo5A is bound to  $\alpha$ -1-2mannobiose, shown in blue. c) The PA14 domain from PA. The CBL2 loop follows a different path than in Flo5A, while in CBL1, one of the aspartates has been substituted by an asparagine, resulting in no interaction with calcium. This is probably also the reason why the peptide bond between the aspartate and the asparagine is in trans, and not cis. [1] d) Multiple alignment between Epa1A, Flo5A and PA14. No Flo5A subdomain is expected within Epa1A, but the DcisD motif is. Part of the Epa1A specificity defining penta-peptide [41] is aligned with CBL2. The alignment hints therefore at the possibility of a similar mechanism in ligand binding between Epa1A and Flo5A. The alignment is based on the one presented by Veelders et. al. [1], and the secondary structure elements are derived from the Flo5A structure.

### 2.4.2 Type V protein secretion: autotransport.

In order to secrete proteins into the extracellular space, Gram-negative bacteria have various systems that allow for the secreted molecules to pass both the inner and the outer membranes (IM and OM, respectively). Secretion types I to IV need accessory proteins to traverse both the IM, the OM and the periplasm (PP). These systems are normally involved in export of soluble proteins, like toxins or constituents of pili or fimbria, which have no more contact with the OM after export. On the other hand, during type V secretion, accessory proteins are only needed to traverse the inner membrane[146]. Transport through the outer membrane is accomplished by a C-terminal autotransporter domain, which is composed of a porin-like  $\beta$ -barrel. After insertion of the  $\beta$ -barrel into the OM, an N-terminal passenger domain is exported to the extracellular space. The passenger domain may then either stay attached to the  $\beta$ -barrel (type V<sub>a</sub> secretion), or separate from it for full secretion (type V<sub>b</sub> secretion)[5]. A third option is that the  $\beta$ -barrel is formed through the  $\beta$ -sheets of several molecules, generating a membrane bound oligomere (type V<sub>c</sub> secretion, Fig. 2.10 a)[5]. Chaperones might be needed for passage through the periplasm, but are not absolutely necessary in most cases.

Autotransporters are consistently composed of three regions; a signal peptide which is compatible either with the Sec or SRP translocation systems, an N-terminal passenger domain, and a C-terminal translocation unit domain (Fig. 2.10 b)[147].

During inner membrane transport, SecB has been postulated to associate with the autotransporter, and to bind to SecA, which translocates the autotransporter to the periplasm[148]. On the other hand, SRP dependent association with the translocon allows for co-translational translocation of the autotransporter[5] (Fig. 2.10 c). Once in the periplasm, it has been hypothesized that the  $\beta$ -barrel of the translocation unit immediately starts insertion into the OM, which leaves the passenger domain in an extended conformation within the periplasm[5]. Evidence has been found for both, a pre-folding before translocation[149] and for some intramolecular chaperones which maintain the passenger domain in an extended, yet stable conformation[150]. It is therefore unclear whether there is a universal mechanism for periplasmic transport, or whether some autotransporters behave in one fashion, and others in a different one.

Protein passage through the outer membrane is also a matter of discussion[125]. Although it was initially proposed that the  $\beta$ -barrel barrel of the translocation unit would autonomously insert itself[146], it could be shown that several factors at least facilitate such insertion. The bacterial

lipopolysaccharide (LPS)[151] and the periplasmic chaperone Skp[152] seem to have some function in the facilitation of membrane insertion of the  $\beta$ -barrel. Also, the outer membrane  $\beta$ -barrel assembly machinery (Bam complex or Omp85) seems to be associated with the translocation unit during insertion[153]. The Bam complex is essential for the export of OMPs, and it has been shown to interact directly with them during membrane insertion[154].

Finally, the passenger domain needs to cross the outer membrane, to be exposed to the medium. A hairpin model seems to be the most prevalent in describing how the passenger domain is translocated through the translocation unit domain[125]. In it, the C-terminal region of the passenger domain is first translocated through the  $\beta$ -barrel pore, yielding an extracellular hairpin. The rest of the passenger domain is 'pulled out' by this hairpin. How this happens without the coupling of an energy carrier such as nucleotide triphosphates is still unclear(Fig 2.10 d)[105]. A weakness of this model is the fact that fully folded, disulfide bridge comprising passenger domains should not be exported, although such transport has been observed *in vivo*[149]. Other models have been proposed (Fig 2.10 d), but none of them can fully explain the translocation behavior.

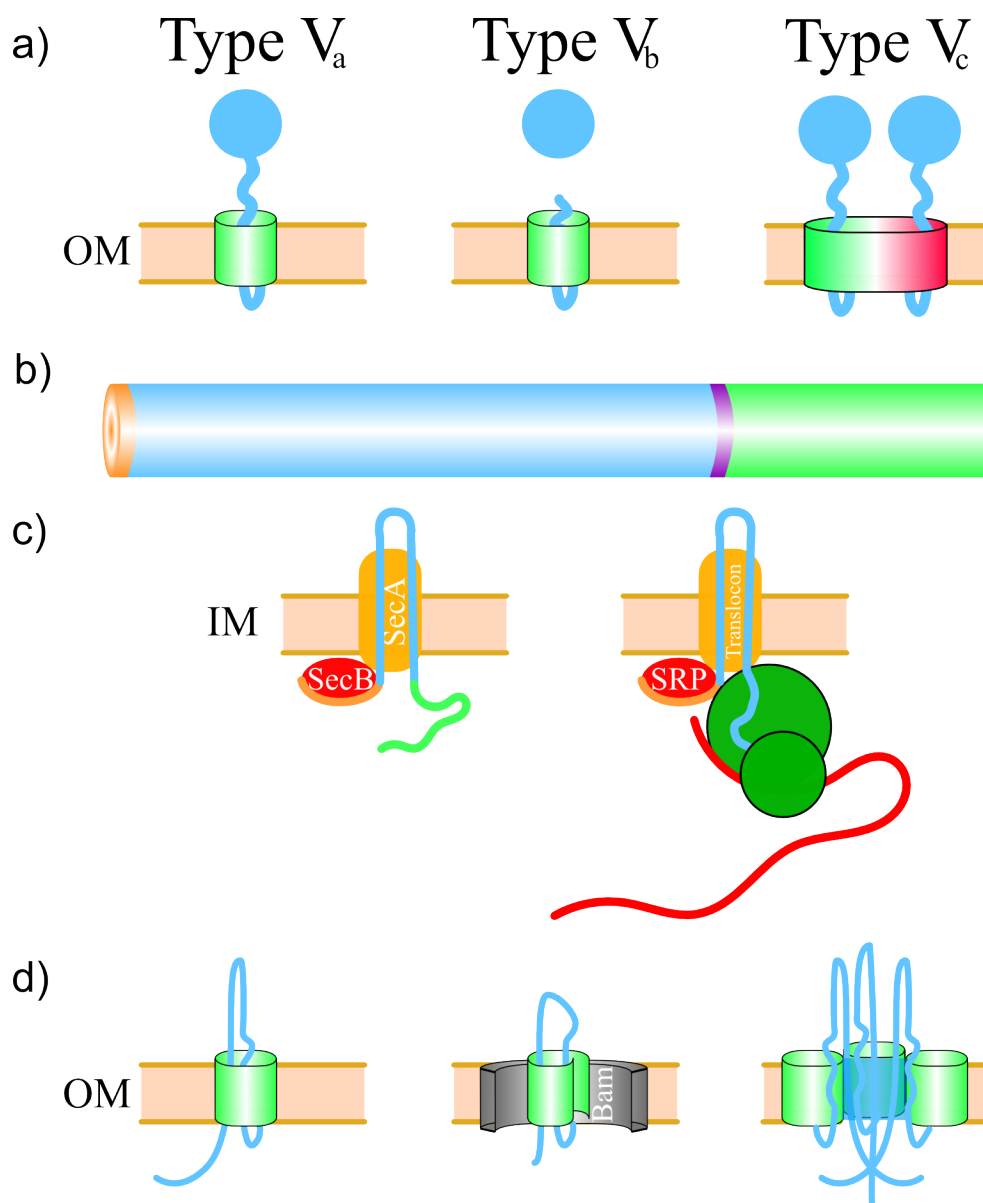


Figure 2.10: Autotransporters in Gram-negative procaryotes. a) Autotransport is also known as type  $V$  secretion. There are three subclasses of type  $V$  secretion. Type  $V_a$ , in which the passenger domain (blue) remains attached to the translocation unit (green). Type  $V_b$ , where the passenger domain is cleaved and released to the exterior. In type  $V_c$  two or more domains build the functional pore (in green and red gradient), yielding a multimeric autotransporter[5]. b) The domain architecture of an autotransporter is composed of a minimum of three regions, an N-terminal signal peptide (orange), the passenger domain (blue) and the C-terminal translocation unit (green). An intramolecular chaperone domain (purple), might be present, but is not necessary[125]. c) Translocation through the inner membrane. Two pathways are available for autotransporter translocation through the inner membrane, either post-translationally through the Sec system[148], or co-translational through the SRP-translocon system[5]. (Ribosome in green, mRNA in red). d) Possible translocation mechanisms of the passenger domain through the outer membrane. Left: classical hairpin mode[105]. Center: the  $\beta$ -barrel pre-forms in the periplasm and is inserted in a semi-folded state into the OM by the Bam system. This mode would allow for fully folded passenger domains to be passed through the membrane. Right: Several translocation units associate in the outer membrane, producing a large central hydrophilic pore. The pore would then allow for the semi-folded passenger domains to pass through[105].



### 3 Aim

The aim of this project was to characterize pathogenic adhesins, of microbial origin which interact with human mucins. *C. glabrata* adhesin Epa1 and *H. pylori*'s BabA and SabA were chosen due to their medical relevance. The A domain of Epa1A was to be functionally characterized and its structure elucidated. Furthermore, variants of Epa1A were to be generated to shift its specificity along Epa1's nearest phylogenetic relatives within the Epa family. These variants were to be also functionally and structurally characterized as well.

The recombinant production of the passenger domain of the *H. pylori* autotransporter adhesins BabA and SabA was to be established. To obtain soluble protein, refolding methods and extensive expression screenings were to be employed. The resulting fragments were to be biophysically and functionally characterized.

## 4 Materials

### 4.1 Chemicals, consumables and equipment

Product	Manufacturer
Ampicillin	<i>Applichem</i>
Bacto <sup>®</sup> yeast extract	<i>Difco</i>
Bacto <sup>®</sup> trypton	<i>Difco</i>
Coomasie-Brilliant blue	<i>Serva</i>
DNA ladder	<i>Fermentas</i>
DTT	<i>Biomol</i>
IPTG	<i>Duchefa</i>
peqGOLD Protein-Marker I	<i>PEQLAB</i>
Restriction enzymes	<i>New England Biolabs</i>
carbohydrates	<i>Dextra Labs</i>
Lacto-N-biose	
Lewis <sup>x</sup> trisaccharide	
Lewis <sup>b</sup> tetrasaccharide	
Lewis <sup>a</sup> tetrasaccharide	
D-Galactose-6-O-sulfate	
N-Acetyl-D-galactosamine-6-O-sulfate	
N-Acetyl-D-galactosamine	
Unstained protein molecular weight marker	<i>Fermentas</i>
Virkon <sup>®</sup>	<i>Antec International</i>

Table 4.1: chemicals.

Product	Model (manufacturer)
Crystallization consumables	
Plates	Cover slides $\varnothing = 22$ mm ( <i>Hampton Research</i> ), 24 well crystallization tool ( <i>Nextal/Qiagen</i> ), 24 well VDX ( <i>Hampton Research</i> ), 96 well Innovaplate™ SD-2 ( <i>Hampton Research</i> ), SilverSeal ( <i>Greiner Bio-one</i> ), VIEWseal™ adhesive sheets ( <i>Greiner Bio-one</i> )
Screens	NEXTAL crystallization kits ( <i>Qiagen</i> )
Tools	Cryotong, Crystal Wand™, Micro-Tools ( <i>Hampton Research</i> )
Cryoloops	Crystal Cap™ Colored ( <i>Hampton Research</i> ), CrystalCap HT™ ( <i>Hampton Research</i> ), Micromounts™ ( <i>MiTeGen</i> )
Dialysis cassettes	Slide-A-Lyzer® Mini Dialysis Units 10000 MWCO ( <i>Pierce</i> )
Dialysis membrane	Pore size 0.025 $\mu$ m, $\varnothing = 25$ mm ( <i>Millipore</i> ), Snakeskin, MWCO 3.5 Kda ( <i>Thermo Scientific</i> )
Disposable cuvettes	Cuvettes REF 67.742 ( <i>Sarstedt</i> )
Desalting columns	PD-10 Desalting Columns ( <i>GE Healthcare</i> ) Microspin G50 Desalting columns ( <i>GE Healthcare</i> )
Protein concentrators	Amicon-Ultra4 ( <i>Millipore</i> )
Sterile filters	Ultrafree® -MC ( <i>Millipore</i> ), Filtropur S 0.2 ( <i>Sarstedt</i> ), Steritop bottle top filter, 22 $\mu$ m express plus ( <i>Millipore</i> )
LC consumables	Columns and materials ( <i>GE Healthcare</i> )

Table 4.2: consumables.

Device	Model (manufacturer)
Autoclave	T-line™ ( <i>Fedegari</i> )
CD spectrometer	
Digital Camera	Powershot A70 ( <i>Canon</i> ), D60 ( <i>Nikon</i> )
Documentation systems	
Agarose gels	Computer E.A.S.Y with digital camera ( <i>UVP</i> ), UV-transilluminator ( <i>Herolab</i> ), thermal printer UP-D 895 ( <i>Sony</i> )
Crystallization	Crystal Pro HT 110 ( <i>TriTek Corporation</i> )
Fluidizer	EmulsiFlex-C5 ( <i>Avestin</i> )
Fluorescence spectrophotometer	
French Press	French-pressure cell-Version 5.1 ( <i>Sim Aminco</i> ) French-pressure cell ( <i>Philipps Universität Marburg workshop for precision engineering</i> )
Gel electrophoresis	
SDS-PAGE	Mighty Small II SE250 ( <i>HoeferScientific Instruments</i> ), <i>PerfectBlue two-gel system Twin M (PEQLAB)</i>
Power boxes	Model 453 ( <i>ISCO</i> ), EPS 300 ( <i>GE Health care</i> )
Gel shaker	Model 3013 ( <i>GFL</i> ), Köttermann 4010 ( <i>Köttermann</i> )
Incubator	Incubator ( <i>Heraeus</i> ), BFED-53 ( <i>WTB Binder</i> )
Crystallization robot	Cartesian Microsys™ SQ4000 ( <i>Genomic Solutions</i> )
LC-Systems	ÄKTA prime™ P100 ( <i>GE Health care</i> ) ÄKTA purifier™ 100 ( <i>GE Health care</i> ) FPLC System 250 ( <i>GE Health care</i> )
Liquid handling system	Lissy ( <i>Zinsser Analytic</i> )
pH-Meter	Basic Meter PB-11 ( <i>Sartorius</i> ), Seven Easy ( <i>Mettler Toledo</i> )
Photometer	U-2000 ( <i>Hitachi</i> ), Ultrospec™ 3100pro ( <i>Biochrom</i> ), Uvikon spectrophotometer 930 ( <i>Kontron</i> ), Nicolet Evolution e100 ( <i>Thermo Electron Corporation</i> )
Ultrapure water dispenser	Seralpur Pro90CN ( <i>Seral</i> ), Milli-Q ( <i>Millipore</i> )
X-ray diffractometer	Rotating anode AXS FR591 ( <i>Bruker</i> ), MAR345dtb detector ( <i>MAR Research</i> ). Oxford cryosystems ( <i>Oxford</i> )
Peristaltic pump	P1 ( <i>GE Health care</i> )
Shakers	Certomat® IS ( <i>Sartorius</i> ), G25 ( <i>New Brunswick Scientific</i> ), Multitron 2 ( <i>Infors</i> ), TH 25 ( <i>Edmund Bühler</i> )
Dish washer	Professional G7883 ( <i>Miele</i> )
Sonicator	Sonoplus HD2200 with UW200 type needle ( <i>Bandelin</i> )

Device	Model (manufacturer)
	<i>electronic)</i>
Scales	DeltaRange® PC2200 ( <i>Mettler</i> ), Toledo Labstyle 54 ( <i>Mettler</i> )
Centrifuges	2K15 ( <i>Sigma</i> ), Biofuge 13R ( <i>Heraeus</i> ), Biofuge fresco/pico ( <i>Heraeus</i> ), Centrifuge 5417C/5910R ( <i>Eppendorf</i> ), Discovery™ 90SE ( <i>Sorvall</i> ), EvolutionRC ( <i>Sorvall</i> ), J2-HS / J2-21 M/E ( <i>Beckmann</i> ), Minifuge GL ( <i>Heraeus Christ</i> )
X-Ray diffractometer	Rotating anode AXS FR591 ( <i>Bruker</i> ) MAR345dtb detector ( <i>Mar Research</i> ) Oxford Cryosystems ( <i>Oxford</i> )

*Table 4.3: Technical equipment.*

Products, chemicals and consumables mentioned in this work which are not listed in these tables were ordered from either *Merck*, *Sigma*, *Roth* or *Fluka* with the highest possible purity.

## 4.2 Vectors and Microorganisms

### 4.2.1 Vectors

#### 4.2.1.1 The pET28a Vector

The pET28a vector system was employed in combination with different *E. coli* strains for the cloning, mutagenesis and heterologous expression of all genes presented in this work. pET28a is a low copy plasmid equipped with a kanamycin resistance and an expression cassette under the control of the T7 promoter and the lac operator. These features ensure selection and tight control of the expression.

The pET28a Vector formed the basis for the two constructs upon which this work is based (Figure 4.1).

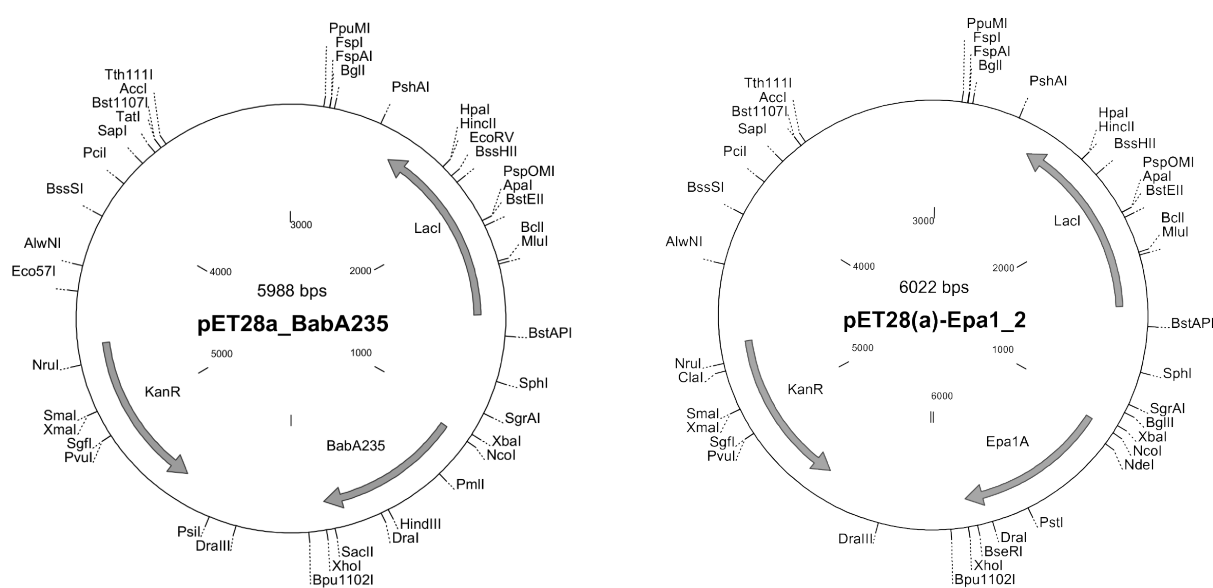


Figure 4.1: Left: pET28a\_BabA235 vector, which contained the putative passenger domain of BabA. It contains both a kanamycin resistance gene, the lactose repressor gene LacI and the BabA235 gene. Right : pET28a-Epa1\_2. This vector contained the Epa1A gene plus the same features as pET28a\_BabA235. The pET28a-Epa1\_2 vector was adopted from [155], where it got its original denomination.

#### 4.2.1.2 The pCR 2.1-Topo plasmid

The pCR 2.1-Topo plasmid is a commercially available plasmid which encodes for both a kanamycin and an ampicillin resistance. The cloning site for the insert is found within the *LacZa* gene. If an insert is successfully cloned within the vector, the  $\beta$ -galactosidase activity of the gene product will be disrupted. A blue-white screening for fragment insertion can then be undertaken by

administering X-gal to the LB-agar plates where transformants are to be grown.

This vector was used primarily in combination with the *Invitrogen* TOPO-TA cloning kit for purposes of subcloning.

#### 4.2.2 *E. coli* strains

##### ***E. coli* DH5 $\alpha$**

Genotype: F<sup>-</sup>  $\Phi$ 80*lacZ* $\Delta$ M15  $\Delta$ (*lacZYA-argF*) U169 *recA1 endA1 hsdR17* (r<sup>k</sup>, m<sup>k+</sup>) *phoA supE44*  $\lambda$ -*thi-1 gyrA96 relA*

The *E. coli* DH5 $\alpha$  (*Invitrogen*) strain is especially well suited for the isolation and production of recombinant DNA, as it is not capable of recombinant expression of T7 promoter-controlled genes and can achieve very high plasmid replication rates.

##### ***E. coli* BL21 (DE3) Gold:**

Genotype: B F<sup>-</sup> *ompT hsdS*(r<sup>B-</sup> m<sup>B-</sup>) *dcm*<sup>+</sup> *Tet*<sup>r</sup> *gal*  $\lambda$ (DE3) *endA Hte*

The BL21 (DE3) Gold strain is equipped with a  $\lambda$  pro-phage containing the T7 RNA polymerase gene, which works together with the *lac* operator, to allow for IPTG mediated gene expression. The system is under the control of the *lac* promoter[156].

##### ***E. coli* Origami2 (DE3):**

Genotype:  $\Delta$ (*ara-leu*)7697  $\Delta$ *lacX74*  $\Delta$ *phoA PvuII phoR araD139 ahpC galE galK rpsL*

F'<sup>+</sup>[*lac*<sup>+</sup> *lacI*<sup>q</sup> *pro*] (DE3) *gor522::Tn10 trxB* (Str<sup>R</sup>, Tet<sup>R</sup>)

*E. coli* Origami 2 (DE3) cells are a K12 derived strain containing the DE3 element and mutations in both the thioredoxin reductase (*TrxB*) and the glutathione oxidoreductase (*Gor*). The mutations within *trxB* and *gor* raise the oxidation potential of the cytoplasm enough to allow for the generation of disulfide bonds[157]. *E. coli* Origami 2 (DE3) cells are therefore well suited for the T7-regulated synthesis of proteins containing disulfide bonds.

##### ***E. coli* Shuffle T7:**

Genotype: F' *lac, pro, lacI*<sup>Q</sup> /  $\Delta$ (*ara-leu*)7697 *araD139 fhuA2 lacZ::T7 gene1*  $\Delta$ (*phoA*)*PvuII phoR ahpC\* galE (or U) galK* *latt::pNEB3-r1-cDsbC* (Spec<sup>R</sup>, *lacI*<sup>q</sup>)  $\Delta$ *trxB rpsL150*(Str<sup>R</sup>)  $\Delta$ *gor*  $\Delta$ (*malF*)

The *E. coli Shuffle T7* (*New England Biolabs*) strain contains deletions within the *gor* and *trxB* genes which, as in the case of the Origami 2 cells, grants it the capability to synthesize disulfide bonds intracellularly. Furthermore, it contains the *DsbC* gene, which codes for a disulfide bridge isomerase[158]. DsbC acts by rearranging energetically suboptimal disulfide bridges, thus increasing in most cases the soluble yield of the produced recombinant protein, as compared with the Origami 2 strain.



### 4.3 Primers

Primers were designed for both *de novo* cloning and mutagenesis of cloned genes.

#### 4.3.1 Primers for *de novo* cloning:

Gene/fragment	Primer ( <i>orientation/Restriction enzyme</i> )	T <sub>M</sub> (°C)
<i>SabA uni</i>	gtgtgCCATGGGGgtgagcgccggctatc ( <i>Fwd/NcoI</i> )	70.05
<i>SabA 180</i>	gttgtgCTCGAGgtttgatggtgagtctgtg ( <i>Rev/XhoI</i> )	62.3
<i>SabA 260</i>	gttgtgCTCGAGagatagggtgtggttgc ( <i>Rev/XhoI</i> )	63.8
<i>SabA 325</i>	gttgtcCTCGAGcgttgaaccgcattggg ( <i>Rev/XhoI</i> )	67.4

Table 4.4: *de novo* cloning primers. *Uni*: universal forward primer, to be used in all amplifications. *Fwd*: forward orientation, *Rev*: reverse orientation, T<sub>M</sub>: melting temperature. Fragment sizes are given in number of amino-acids. Restriction sites are highlighted in uppercase.

#### 4.3.2 Mutagenesis primers:

Construct/Mutation	Primer ( <i>orientation/diagnostic restriction</i> )	T <sub>M</sub> (°C)
BabA144/ Frame correction	cctccaggcttatactgTTCGAActgcaagtgatggaag ( <i>Fwd/BstBI</i> ) CttccatcacttgcaTTCGAAcagggtataaacctggagg ( <i>Rev/BstBI</i> )	80.1
BabA235/ Frame correction	GatgatggctgCCGCGGggcaccaggccactatc ( <i>Fwd/SacII</i> ) gatagtggcctggtgccCCGCGGCagccatcatc ( <i>Rev/SacII</i> )	82.3
Epa1→6A/ E220D Y221N	ccctattaggtattttataataaacagagataaatgatGCGCGCtcagttttac ( <i>Fwd/BssHII</i> ) gtaaaactgaGCGCGCcatcattatctctgttattataaaataacctaataaggg ( <i>Rev/BssHII</i> )	78.4
Epa1→2A/ E220D Y221N D222N	ccctattaggtattttataataaacagagataaataggtgctcagttttac ( <i>Fwd/deleted BssHII</i> ) gtaaaactgagcgcaccatttatctctgttattataaaataacctaataaggg	
Epa1→3A/ R219I E220G Y221K	ccctattaggtattttataataacataggatatgatggcgcactcagttttac ( <i>Fwd_1</i> ) gtaaaactgagtgcgccatcatatcctatgttattataaaataacctaataaggg ( <i>Rev_1</i> ) ccctattaggtattttataataacataggaaggatGCGCGCtcagttttac ( <i>Fwd_2/BssHII</i> ) gtaaaactgaGCGCGCcatccttctctatgttattataaaataacctaataaggg ( <i>Rev_2/BssHII</i> )	77.1 78.6

Table 4.5: Mutagenesis primers. Mutations are marked in red, mutations accumulated from previous rounds of mutagenesis are indicated in green, diagnostic restriction sites marked with capital letters.

#### 4.4 Media, buffers and Stock solutions

##### 4.4.1 Media:

Both Luria Bertani (LB) and Terrific Broth (TB) media were employed for heterologous expression; only LB medium was used during molecular biology experiments, both in liquid cultures and in agar plates. All media were autoclaved immediately after preparation (30 min, 121 °C, 1.5 bar) and were kept under sterile conditions until inoculation. Kanamycin (35 µg/ml en concentration) was added prior to inoculation in order to select for the pET28a vector. Any further selection requirements imposed by the corresponding strain were also fulfilled.

LB-Medium	
10 g/L	Bacto-Trypton
5 g/L	Bacto Yeast Extract
10 g/L	NaCl
0.4 mL/L	10 M NaOH Solution

Table 4.6: LB-Medium.

TB-Medium	
12 g/L	Bacto-Trypton
24 g/L	Bacto-Yeast Extract
4 mL	Glycerol (87%)
Ad 900 mL	Distilled H <sub>2</sub> O
Phosphate buffer	
0.17 M	KH <sub>2</sub> PO <sub>4</sub>
0.72 M	K <sub>2</sub> HPO <sub>4</sub>
Ad 100 mL	

Table 4.7: TB-Medium.

The TB-medium and the phosphate buffer were autoclaved separately and mixed later under sterile conditions.

#### 4.4.2 Buffers:

##### 4.4.2.1 Protein buffers:

Protein buffers were used to purify and store native or refolded proteins.

AM buffer (pH 8)	AMI buffer
20 mM Tris/HCl	AM buffer
200 mM NaCl	500 mM imidazole
AME buffer	AML buffer
AM buffer	AM buffer
10 mM EDTA	50 mM lactose
Phosphate buffer pH 5.8	Phosphate buffer pH 2.5
50 mM KH <sub>2</sub> PO <sub>4</sub> /K <sub>2</sub> HPO <sub>4</sub> pH 5.8	50 mM KH <sub>2</sub> PO <sub>4</sub> /HCl pH 2.5
200 mM NaCl	200 mM NaCl

Table 4.8: protein buffers.

##### 4.4.2.2 Washing buffers:

Washing buffers were used for the isolation and enrichment of inclusion bodies.

Washing buffer 1
20 mM Tris/HCl pH 8
200 mM NaCl
1 mM EDTA
0.3 mM PMSF
0.75 % (v/v) LDAO
Washing buffer 2
20 mM Tris/HCl pH 8
200 mM NaCl
0.1 mM EDTA
0.3 mM PMSF

Table 4.9: Washing buffers.

#### 4.4.2.3 Denaturing buffers:

Denaturing buffers were used to solubilize and purify protein from inclusion bodies.

Denaturing buffer 1
AM buffer
6 M guanidinium chloride
Denaturing buffer 2
AM buffer
6 M guanidinium chloride
500 mM imidazole

*Table 4.10: Denaturing buffers.*

#### 4.4.2.4 Analytical buffers:

Analytical buffers were used during experiments to increase stability under experimental conditions, or to provide the corresponding ligands or metal ions necessary for binding experiments.

Fluorescence buffer
0.1 M $K_2HPO_4/KH_2PO_4$ pH 5.2
5 mM $CaCl_2$
1 mM EDTA
Transition buffer (pH 8 – 2.5)
50 mM $K_2HPO_4/KH_2PO_4/HCl$

*Table 4.11: Analytical buffers.*

#### 4.4.2.5 Refolding buffer:

The refolding buffer provided adequate conditions for preparative folding of denatured protein by rapid dilution.

Refolding buffer
50 mM HEPES pH 7.5
0.5 mM Arginine
6 mM Glutathione (GSH)
4 mM Di-glutathione (GSSG)

Table 4.12: Refolding buffer.

#### 4.4.2.6 Buffers for the preparation of chemically competent cells

TFB I	TFB II
100 mM RbCl	10 mM MOPS
50 mM MnCl <sub>2</sub> ·2H <sub>2</sub> O	10 mM RbCl
30 mM KOAc	75 mM CaCl <sub>2</sub> ·2H <sub>2</sub> O
10 mM CaCl <sub>2</sub> ·2H <sub>2</sub> O	15 % (v/v) Glycerol
15 % (v/v) Glycerol	

Table 4.13: Buffers for preparation of chemically competent cells.

#### 4.4.3 Stock solutions:

##### 4.4.3.1 Agarose gel electrophoresis stock solutions:

Agarose gel	6x DNA Loading buffer
1 % (w/v) Agarose	0.25 % (w/v) Xylencyanol FF
TBE buffer	1.5 % Ficoll Type 400
10x TBE buffer	Ethidium bromide bath
1 M Tris	200 ml distilled H <sub>2</sub> O
1 M Boric acid	200 µg Ethidium bromide
20 mM EDTA	

Table 4.14: Solutions for agarose gels.

## 4.4.3.2 SDS PAGE stock solutions:

Acrylamide solution
30 % (w/v)
Stacking gel buffer
0.625 M Tris/HCl pH8
Resolving gel buffer
1.125 Tris/HCl pH 8.0
30 % (w/v) Sucrose
Running buffer
0.25 M Tris
2 M Glycine
1 % (w/v) SDS
Protein loading buffer
62.5 mM Tris/HCl pH6.8
15 % (w/v) Glycerol
4 % (w/v) SDS
0.001 % (w/v) Bromophenol blue
4 % (v/v) $\beta$ -Mercaptoethanol
Staining solution
0.25 % (w/v) Coomassie brilliant blue R250
5 % (v/v) Acetic acid
50 % (v/v) Ethanol
De-staining solution
5 % (v/v) Acetic acid
28 % (v/v) Isopropanol

Table 4.15: Stock solutions used during SDS PAGE

#### 4.4.3.3 Buffers for western blotting

Blotting buffer	Blocking liquid
25 mM Tris/Hcl pH 8	5 g/l Blocking reagent
150 mM Glycine	1 X Blocking buffer
10 % (v/v)Methanol	0.1 % (v/v) Tween-20
Fixing reagent	Ink staining solution
20 g/l K <sub>2</sub> S <sub>2</sub> O <sub>5</sub>	2 % (w/v) Ink (blue)
100 g/l Na <sub>2</sub> S <sub>2</sub> O <sub>3</sub> ·5H <sub>2</sub> O	1% (v/v) Acetic acid
TBS buffer	TBS-T buffer
50 mM Tris/HCl	20 mM Tris/HCl pH 7.5
138 mM NaCl	150 mM NaCl
2.7 mM KCl	0.1 % (v/v) Tween-20

*Table 4.16: Buffers for western blotting.  
The blocking buffer was commercially obtained  
as a 20 X concentrate.*

#### 4.4.3.4 Diverse stocks

Antibiotic stock solution
35 mg/mL Kanamycin
12.5 mg/mL Tetracycline in isopropanol
50 mg/mL Streptomycin
Other stocks
1 M IPTG

*Table 4.17: Diverse stocks.*

## 5 Methods

### 5.1 *Molecular biology*

All methods employed to manipulate, modify and obtain DNA will be described in the following sections.

#### 5.1.1 *DNA synthesis methods*

##### 5.1.1.1 *Polymerase chain reaction (PCR)*

The PCR method is based on the temperature-dependent activity of DNA polymerases from thermophilic organisms, and their dependence on specific primers for initiation of replication[159]. By repeating several cycles of DNA melting, primer annealing and elongation, the desired DNA fragment is exponentially amplified from a template, such as genomic DNA or a cDNA library.

In order to calculate the proper temperatures and times for the reaction, the following factors were taken into account, as described in the Phusion-polymerase user manual (*Invitrogen*).

1. The melting temperature for the primers was calculated by the nearest neighbor method[160].
2. The annealing temperature was calculated as the lowest melting temperature of the two primers plus 3 °C.
3. The elongation times were calculated by providing 15-30 seconds per kilobase of desired amplified product.

In table 5.1 a sample mix is given, while in table 5.2 a cycling program is described.



Component	Volume ( $\mu$ l)	Final concentration
5X Phusion HF-Buffer	10	1 X
10 mM dNTPs	1	0.2 mM
10 $\mu$ M Fwd.-Primer	1.25	250 nM
10 $\mu$ M Rev. Primer	1.25	250 nM
$\sim$ 0.3 g/l Template	1 $\mu$ l	6 mg/l
2 U/ $\mu$ l Phusion Polymerase	0.5 $\mu$ L	0.02 U/ $\mu$ l
Double distilled H <sub>2</sub> O	ad 50 $\mu$ l	

Table 5.1: Sample composition of a PCR mix.

N° of cycles	Step	Temperature ( $^{\circ}$ C)	Time (sec)
1	Denaturing	98	300
10	Denaturing	98	30
	Annealing	52.3	30
	Elongation	72	8
	Denaturing	98	30
25	Annealing	67.7	30
	Elongation	72	8
	Elongation	72	600

Table 5.2: Example of a PCR program. Two sets of cycles provide initially relaxed conditions to obtain large amounts of template and final stringency for increased specificity.

### 5.1.1.2 Site-directed mutagenesis

Site-directed mutagenesis was performed in a Biometra T personal thermal cycler. in which complementary primers were used to amplify an entire circular DNA molecule[161]. The PCR protocol was analogous to the one described in section 5.1.1.1, with the exception that the melting temperature of the primer was taken as annealing temperature, and the elongation time was always kept at 30 seconds per kilobase. The PCR was immediately followed by a DpnI restriction, which digests methylated, i.e. parental, DNA, but leaves unmethylated DNA intact. The digestion took

place by adding 1  $\mu$ l DpnI at 37 °C for 1 h. The digestion mix was then desalted by dialysis at RT for 1 hour against double distilled water and finally used to transform electrocompetent *E. coli* DH5 $\alpha$  cells.

## 5.1.2 Cloning

### 5.1.2.1 Digestion with restriction enzymes

Type II restriction enzymes are widely used at preparative scales to produce sticky ends in DNA fragments, which facilitate directional ligations between inserts and vectors. Analytical digestions are also common for quick identification of unsequenced DNA molecules.

Preparative digest mixtures (Table 5.3) were incubated at 37 °C for 4 hours and were either PCR purified or loaded onto a preparative 1 % agarose gel. Analytical digestions were only allowed to proceed for one hour before agarose gel electrophoresis.

Component (concentration of stock)	Preparative ( $\mu$ l)	Analytical ( $\mu$ l)
DNA fragment (53 ng/ $\mu$ l)	30	4
NcoI	2	0.25
XhoI	2	0.25
10 X NEB4	5	1
100 X BSA solution	0.5	0.1
Double distilled H <sub>2</sub> O	ad to 50	3.4

*Table 5.3: Exemplary digestion of DNA.*

### 5.1.2.2 Ligation

The active enzyme catalyzing DNA ligation was the ATP-dependent T4-DNA-ligase. Reactions (Table 5.4) took place at either 37 °C for 1 h, RT for 4 hours or overnight at 18 °C.

Component	Quantity
Digested vector	25 fmol
Digested insert	125 fmol
T4-DNA-ligase	1 $\mu$ l
Double distilled H <sub>2</sub> O	Add to 9 $\mu$ l
10 X T4-ligase buffer	1 $\mu$ l

Table 5.4: Composition of a ligation sample.

### 5.1.2.3 Transformation

During electro-transformation, cells are electroshocked using a very short pulse with high potential, which permeabilizes the cells and allows for DNA to enter the cytoplasm. 10  $\mu$ l of the ligation assay were dialyzed against double distilled water for 1 hour to avoid salt derived short-circuits during transformation. Two transformations were performed, one with 1  $\mu$ l of the sample, and one with the rest. The desired volume of the ligation assay was pipetted into a 50  $\mu$ l aliquot of chilled electrocompetent *E. coli* DH5 $\alpha$  and transferred into a 0.2 cm electroporation cell. The sample was then pulsed in an electroporator for 4.5 – 6 ms, 950  $\mu$ l LB-Medium were swiftly added and the cells were finally incubated for 1 h at 200 rpm in a 37 °C shaker. Subsequently, cells were plated on LB-agar plates containing the necessary selection markers.

For chemical transformation, either *E. coli* DH5 $\alpha$ , BL21(DE3)Gold, Origami 2 or Shuffle T7 cells were used. 50  $\mu$ l aliquots of these were incubated with 1  $\mu$ l DNA solution for 0.5 h on ice, then heat-shocked for 90 s at 42 °C, resuspended in 950  $\mu$ l LB -medium and subjected to the same growth conditions as the electroporated cells.

## 5.1.3 DNA Isolation methods

### 5.1.3.1 Qiagen methods

PCR purification, gel extraction and plasmid preparation kits were purchased from *Qiagen*, and their protocols were followed as described in the manufacturer's manual.

### 5.1.3.2 Plasmid preparation by the “quick & dirty” method

The quick & dirty method is based on the selective alkaline lysis[162].

A pelleted 5 ml LB medium overnight cell culture was treated with 250  $\mu$ l P1 buffer, which destabilizes cell walls and brakes down RNA molecules. Subsequently, 250  $\mu$ l P2 buffer were added to disrupt the cell membrane and lyse the cells. This step was followed by careful inversion of the vial, so as to ensure maximum mixture and avoid fragmentation of genomic DNA. 300  $\mu$ l P3 buffer were added, which produces insoluble potassium complexes of dodecyl-sulfate with large DNA molecules, while the small ones remain in solution. The suspension was then centrifuged at  $\sim$ 17000g for 30 min.

The supernatant was removed and mixed with 600  $\mu$ l isopropanol. A second 17000g centrifugation step for 30 min followed, the supernatant was discarded and the small pellet kept. The pellet was washed with 70 % ethanol and centrifuged again at 17000g for 10 minutes. The supernatant was carefully removed, and the remaining pellets were dried further at 37 °C for  $\sim$ 30 min. Finally, the dried pellets were redissolved into EB buffer and stored at -20 °C.

#### *5.1.4 Agarose gel electrophoresis*

DNA fragments were separated by the method of gel electrophoresis (1 % agarose gels in TBE buffer) performed at 100 - 120 V, 300 mA for 1 h. Bands were subsequently visualized by soaking in the ethidium bromide bath and exposure to UV light (254 nm).

#### *5.1.5 Cell preparation methods*

##### ***5.1.5.1 Preparation of chemically competent *E. coli* cells***

The following method was successfully performed with all strains discussed in this work.

50 ml liquid LB medium was inoculated with a single fresh colony from a new LB-agar plate. Cells were left to grow overnight at 37 °C and 200 rpm. 100 ml liquid LB medium were subsequently inoculated with 1 ml of the overnight culture. The culture was grown at 37 °C and 200 rpm until densitometric measurements showed  $OD_{600}=0.5 - 0.6$ .

The cells were then immediately cooled on ice and centrifuged (12300g, 10 min, 4 °C). The pellet was resuspended in 30 ml TFB I buffer. After 2 h incubation time on ice, the cells were again centrifuged (12300g, 5 min, 4 °C) and resuspended in 4 ml chilled TFBII. The suspension was then divided into 50  $\mu$ l aliquots in Eppendorf cups, shock frozen in liquid nitrogen and stored at -80 °C.

### ***5.1.5.2 Preparation of electrocompetent DH5a cells***

1 l liquid LB medium was inoculated with 1 ml of an overnight culture and incubated at 37 °C and 200 rpm until an  $OD_{600}=0.7 - 0.9$  was reached. The culture was chilled for 20 min on ice and centrifuged (12300g, 10 min, 4 °C). The pellet was resuspended in 450 ml chilled, sterile, double distilled water and centrifuged again as described, after which the process was repeated. The pellet was then resuspended in 450 ml sterile and chilled 10 % (v/v) glycerol and pelleted as above. Finally, the cells were resuspended in 2.5 ml sterile and chilled 10 % (v/v) glycerol and divided in 50  $\mu$ l aliquots in Eppendorf cups, shock frozen in liquid nitrogen and stored at -80 °C[163].

## 5.2 Biochemical methods

### 5.2.1 Recombinant gene expression

Expression cultures were used to produce recombinant protein for structural, biophysical and functional characterization.

#### 5.2.1.1 Analytical expression

Analytical expression was used to determine optimal expression conditions for each of the *de novo* cloned expression cassettes. Table 5.5 shows all conditions tested for each new construct. In all cases, culture volume was 50 ml in 250 ml Erlenmeyer flasks with an inoculation volume of 0.5 ml overnight pre-culture. All samples were shaken at 225 rpm.

Construct; expression strain	Medium	Induction type	Inductor concentration (mM)	Time/temperature (h/°C)
pET28a_SabA260/180/140; <i>E. coli</i> BL21 (DE3) Gold	LB	IPTG	0.1 / 1	4 h/37 °C; 36 h/18 °C; 16 h/25 °C
	TB	autoinduction	0	24 h /37 °C; 36 h/18 °C
pET28a_Epa1→6A/2A/3A; <i>E. coli</i> Shuffle T7	LB	IPTG	0.01 / 0.1 / 1	72 h/12 °C
	TB	autoinduction	0	72 h/ 12 °C; 24 h/ 37 °C
pET28a_Epa1→6A/2A/3A; <i>E. coli</i> Origami 2	LB	IPTG	0.01 / 0.1 / 1	72 h/ 12 °C
	TB	autoinduction	0	72 h/ 12 °C; 24 h/ 37 °C

Table 5.5: Overview of performed analytical expressions. All expressions were shaken at 225 rpm.

Cells were harvested by centrifugation at 4220g for 20 minutes at 4°C in falcon tubes. Supernatant was discarded, while the pellets were frozen at -20 °C.

#### 5.2.1.2 Preparative expression

Preparative expressions were either directly upscaled from the best analytical expression conditions, or taken as such from previous work. All expressions were scaled in 5 L Erlenmeyer flasks shaken at 160 rpm. Table 5.6 shows the chosen expression conditions for all constructs.

Construct; Strain	Medium	Type of induction	Time and Temperature (h/°C)	IPTG concentration (mM)	Volume of culture per flask (l)
pET28a_SabA260; <i>E. coli</i> BL21 (DE3) Gold	LB	IPTG	36 h/ 18 °C	1	1
pET28a_BabA210; <i>E. coli</i> BL21 (DE3) Gold	TB	IPTG	36 h/ 37 °C	1	3
pET28a_BabA235 <i>E. coli</i> BL21 (DE3) Gold	TB	IPTG	36 h/ 37 °C	1	3
pET28a_BabA 255; <i>E. coli</i> BL21 (DE3) Gold	TB	IPTG	36 h/ 37 °C	1	3
BabA2.2_5; <i>E. coli</i> BL21 (DE3) Gold	TB	IPTG	36 h/ 37 °C	1	3[164]
pET28a_Epa1_2wt; <i>E. coli</i> Origami 2	LB	IPTG	72 h/ 12 °C	0.01	2[155]
pET28a_Epa1_2wt; <i>E. coli</i> T7 Shuffle	LB	IPTG	72 h/ 12 °C	0.01	2
pET28a_Epa1→2A; <i>E. coli</i> T7 Shuffle	LB	IPTG	72 h/ 12 °C	0.01	2
pET28a_Epa1→3A; <i>E. coli</i> T7 Shuffle	LB	IPTG	72 h/ 12 °C	0.01	2
pET28a_Epa1→6A; <i>E. coli</i> T7 Shuffle	LB	IPTG	72 h/ 12 °C	0.01	2

Table 5.6: Overview of performed preparative expressions. Shaker speeds were set to 160 rpm.

After expression was complete, cells were pelleted for 10 minutes at 4 °C and 8000 rpm in a JA-10 rotor. Pellets were resuspended in 20 ml AM-buffer per liter of culture, shock frozen in liquid nitrogen and stored either at -20 °C for short term or at -80 °C for long term storage.

## 5.2.2 Cell lysis

Depending on the intended use of the cell extract, different methods of cell lysis were employed.

### 5.2.2.1 Analytical cell lysis through sonication

Analytical expressions had to be tested not only for protein production, but also for proper protein folding. Unfolded or mis-folded proteins usually form insoluble inclusion bodies, therefore, separating and analyzing the soluble and insoluble fractions of cell extracts is a good criterion for the identification of conditions leading to the expression of correctly folded proteins.

Pelleted cells were thawed and resuspended in 3 ml cold AM-Buffer. 1 ml of the suspension and 2

$\mu$ l 50 mg/ml lysozyme were then pipetted into a 1.5 ml Eppendorf cup and incubated for 10 minutes on ice. Cells were lysed by sonication for a total of 90 pulses divided evenly into 3 rounds separated by ~30 seconds pauses. Soluble and insoluble fractions were separated by centrifugation (17000g, 30 minutes, 4 °C). The supernatant was then decanted into a new Eppendorf cup, while the pellet was resuspended in 1 ml AM-buffer. 10  $\mu$ l of each were mixed with 10  $\mu$ l protein loading buffer and incubated at 95 °C for 5 minutes. 5 to 10  $\mu$ l of each were then loaded on a 12 % SDS-PAGE-gel.

#### **5.2.2.2 Preparative cell lysis to obtain inclusion bodies**

20 ml of a thawed aliquot from a preparative culture (5.2.1.2) were treated with 200  $\mu$ l 0.1 M EDTA (1 mM final concentration), 120  $\mu$ l 0.1 M PMSF (0.6 mM final concentration) and 480  $\mu$ l of a 100 mg/ml lysozyme solution (1.2 mg/ml final concentration). The assay was incubated at room temperature for 20 minutes with occasional shaking.

Then 2 ml 30 % LDAO (3 % final concentration) were added, and the sample was incubated for further 30 minutes at 37 °C with mixing at 225 rpm. Finally a spatula tip of freeze dried DNaseI was dissolved in the cell extract, followed by a further 30 minutes incubation at RT. This last step was repeated for as long as no DNA (identifiable by its high viscosity and its milky and mucilaginous consistency) was observable in the sample. Once the suspension had been properly clarified, a centrifugation step followed for 30 minutes (15000 rpm, 4 °C, JA-20), after which the supernatant was discarded and the pellet was either stored at -20 °C or further processed to isolate inclusion bodies.

#### **5.2.2.3 Preparative cell lysis to obtain soluble protein using the French press**

The heat sensitivity of soluble proteins often makes their cellular extraction challenging. The use of the French press (*Sim Amico*) offers a fast and efficient alternative for cell wall disruption, since the pre-cooling of the steel cell provides protection against heat denaturation to the cytoplasmic lysate.

An aliquot of the cell suspension was thawed and treated with 50  $\mu$ l 50 mg/ml lysozyme and a spatula tip of freeze dried DNaseI. The suspension was loaded on the pre-cooled pressure cell immediately afterwards, and the cell was operated at a maximum pressure of 6895 kPa (1000 psi). After two rounds, the cell lysate was centrifuged for 1 h in a JA-20 rotor (18000 rpm, 4 °C), supernatant was kept for purification, while the pellet was discarded.



#### 5.2.2.4 Preparative cell lysis to obtain soluble protein using the emulsifier C5

The emulsifier C5 (*Avestin*) works similarly to the French press, but allows for the processing of much larger volumes. However, it is more prone to clotting, and requires much more diluted suspensions. As such, it was mainly used to lyse cell extracts containing SabA gene products from long auto-inductions, which had to be diluted to ~300 ml with AM buffer before effective cell lysis could take place. The sample was treated analogously as with the French press, with maximum pressures between 50000 and 100000 kPa (7252 – 14503 psi) and a centrifugation step which lasted for 3 h.

### 5.2.3 Protein purification

#### 5.2.3.1 Isolation of inclusion bodies

To isolate inclusion bodies, the insoluble fraction of the cell extracts were homogenized with a potter homogenizer using 6 wash steps.

Wash step	Composition	Final concentration
1 to 3	20 ml AM buffer 200 $\mu$ l 0.1 M EDTA 60 $\mu$ l 0.1 M PMSF 500 $\mu$ l 30% (w/v) LDAO	1 mM 0.3 mM 0.75%
4	20 ml AM buffer 200 $\mu$ l 0.1 M EDTA 60 $\mu$ l 0.1 M PMSF	1 mM 0.3 mM
5 and 6	20 ml double distilled H <sub>2</sub> O	

*Table 5.7: Composition of washing buffers for the isolation of inclusion bodies*

After wash steps 1 to 5 the suspension was centrifuged for 20 minutes (4 °C, JA-20, 18000 rpm) and re-homogenized for 5 minutes in the next buffer. After the sixth wash-step, the suspension was divided into 10 2 ml Eppendorf cups and pelleted for 30 minutes at 13000 rpm. The supernatant was discarded and the pellets were kept at -20 °C for further use.

### 5.2.3.2 *Immobilized metal affinity chromatography: NiNTA*

Cloning of adhesin-encoding inserts into the pET28a vector allowed for the fusion of all gene products with either a C- or N-terminal His-Tag compatible with affinity purification by means of NiNTA chromatography[165]. Soluble proteins were purified under native conditions and insoluble proteins had to be purified under denaturing conditions; a 3 ml Ni-NTA column was used in all cases. Elution was monitored with a UV absorption detector (280 nm detection).

For native proteins, the column was equilibrated with 5 column volumes (CV) of water and 5 CV of AM buffer. The cell extract was then sterile-filtered and loaded onto the column via a peristaltic pump, operating overnight with a speed of 0.5 ml/min. Next day, the column was attached to either the Äkta Prime or the Äkta Purifier FPLC systems and first washed with a minimum of 10 CV of AM-buffer. Washing continued with AM buffer containing 5 % AMI buffer. In case any impurities eluted from the column, the washing procedure was continued until impurities were completely eliminated. Finally, a 5 CV gradient to 100 % AMI buffer was used to elute the purified protein, which was fractionated in 1 – 2 ml fractions. The column was rinsed with 5 CV AMI buffer, 5 CV water and 5 CV 20 % EtOH (v/v) for storage. All steps took place at 4 °C.

For protein stabilization from inclusion bodies cell pellets were thawed and resuspended with a potter homogenizer in 35 ml denaturing buffer 1. The suspension was then heated to 75 °C for 20 minutes in a water bath, then left to cool down and sterile filtered. 6 ml 50 % (w/v) NiNTA material suspension in 20 % (v/v) ethanol was centrifuged for 5 minutes at 750g and RT with the supernatant being discarded. The pelleted NiNTA material was centrifuged twice after being washed first with 15 ml water, and then 15 ml denaturing buffer 1. The NiNTA material was then mixed in the protein solution. The mixture was incubated with light shaking overnight at RT. The next day, the material was packed into a chromatographic column, attached to an FPLC system at RT and purified in the same way as described for native proteins, replacing AMI buffer with denaturing buffer 2.

### 5.2.3.3 *Size exclusion chromatography*

Size exclusion chromatography was the method of choice to polish pre-purified protein after the NiNTA step. The gel filtration system was composed of Superdex 200 material, which is composed of a dextran and cross-linked agarose matrix (figure 5.1)[166], packed in an XK 16/70 130 ml column. Using either the Äkta prime, Äkta purifier (*GE*) or an FPLC (*Pharmacia Biotech*), the column was equilibrated first with 1.5 CV water and then with 1.5 CV of the corresponding buffer.

Elution was detected by absorption at 280 nm.

Desalting columns are a special form of gel filtration in which the matrix is very tightly packed, only allowing small molecules to enter the beads, thus proteins are eluted in the exclusion volume. Buffers could be exchanged by using PD-10 desalting columns (*GE Healthcare*) and illustra MicroSpin G50 columns (*Amersham biosciences*). PD-10 columns were equilibrated as described in the manual [167], but were then loaded with 500  $\mu$ l protein solution and eluted in fractions of 500  $\mu$ l. Protein rich fractions were detected at the nano-drop spectrophotometer and pooled. Illustra MicroSpin G50 columns were equilibrated with 2 times 200  $\mu$ l of the target buffer and then packed, loaded and eluted as described in the manual [168].

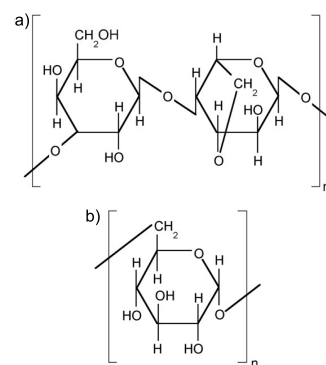


Figure 5.1: Structures of dextran (a) and agarose (b). Source: [166]

#### 5.2.4 Protein concentration

Amicon Ultra-4 (*Millipore*) concentrators were used to concentrate pooled fractions from the purification and refolding steps. The concentrators were chosen to have at least a 10 kDa smaller MWCO than the target protein to be concentrated. Concentrators were first rinsed for 5 minutes at 4220 g and 4 °C for native or refolded proteins and at RT for denatured proteins, with the same buffer in which the protein had been pooled. The protein solution was then loaded and concentrated to the desired concentration by the same procedure. Concentration was measured by absorption at 280 nm.

#### 5.2.5 Protein refolding

Protein refolding is always a two component system, comprising on the one hand the solubilized, denatured and purified protein in denaturing buffer, on the other, the refolding buffer which forces the protein to undergo proper folding. The two components can be mixed in two ways: 1) either through dialysis, which consists of two separate compartments joined by a semipermeable membrane with a MWCO that only allows for small molecules to pass through, and 2) by rapid dilution, where a small amount of concentrated protein solution is mixed with a large volume of refolding buffer and incubated for a certain amount of time. Refolding buffers are not always optimal for further biochemical and or biophysical characterization, as they may interfere with analytical measurements or crystallization, therefore it is sometimes necessary to introduce a third

step in which the protein is loaded onto a desalting column.

### 5.2.5.1 Analytical protein refolding by microdialysis

Different refolding and storage buffers were screened for optimal conditions by analytical microdialysis. The setup for microdialysis consisted of a microdialysis unit with a sample compartment filled with 10  $\mu$ l of 5 mg/ml protein solution in denaturing buffer 1 sealed with a small

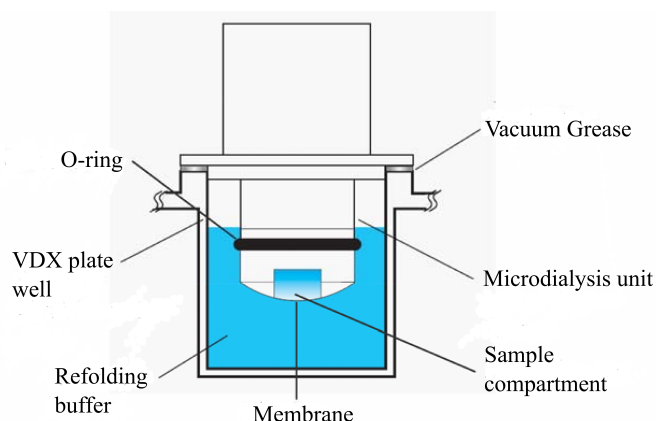


Figure 5.2: Microdialysis setup  
(Source: [www.hamptonresearch.com](http://www.hamptonresearch.com))

piece of snakeskin dialysis membrane (*Thermo Scientific*) with a MWCO of 3.5 kDa. The loaded microdialysis units were sunk into VDX 24 well crystallization plates filled with 1.5 ml refolding buffer in each well. The setup is depicted in figure 5.2.

All assays were tested at 4 °C and incubated for at least 12 hours, before the unit was removed from the well. The dialysis membrane was carefully set aside and the

solution or suspension, depending on the amount of precipitate, was transferred into an Eppendorf cup. The samples were then centrifuged at 17000g for 1 hour, the supernatant was transferred to a fresh Eppendorf cup and mixed with 10  $\mu$ l of protein loading buffer. The pellets were resuspended in 20  $\mu$ l protein loading buffer, with both samples loaded on the large *PerfectBlue two-gel system Twin M (PEQLAB)*. Alternatively, protein concentrations of the supernatant and pellet were spectrophotometrically determined by the BCA test, and refolding yields were determined from those.

Table 5.8 shows the tested storage buffers. Refolding buffers tested by this method were based on the original BabA full length refolding protocol and all comprised the same buffer system: 200 mM NaCl and 10 mM phosphate buffer with pHs ranging from 7.8 to 3.8.

Buffer system	pH	Additives tested
50 mM Tris/HCl	6, 7	200 mM NaCl 5 % Glycerol 5 % PEG 400
50 mM CHES	8, 9	200 mM NaCl 5 % Glycerol 5 % PEG 400
50 mM Acetate buffer, 200 mM NaCl	3.4.5	20 mM CaCl <sub>2</sub> 20 mM MgCl <sub>2</sub> 20 mM MnCl <sub>2</sub> 5 % Glycerol

Table 5.8: Tested storage buffer conditions for refolding trials of BabA fragments. All conditions were tested with and without 1 mM β-mercaptoethanol as a reducing agent.

Promising conditions were upscaled to 1 ml by using Slide-A-Lyzer cassettes. Supernatant concentrations were tested via absorption measurements and sample quality via circular dichroism (CD) spectroscopy.

### 5.2.5.2 Analytical protein refolding by rapid dilution

Analytical rapid dilution took place at the micro-scale through the commercially available iFold screen (Merck), as depicted in Figure 5.3.

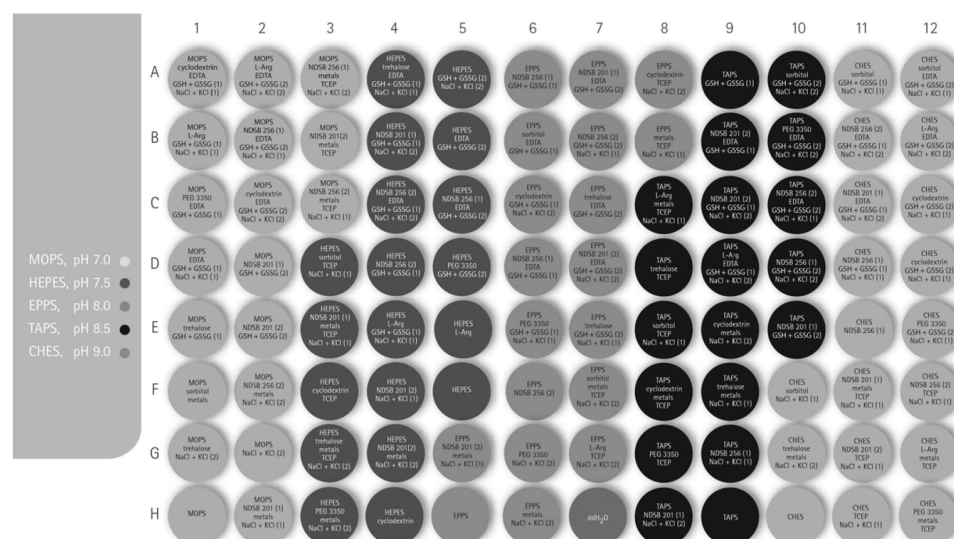


Figure 5.3: iFold Protein refolding system 2. (Source:www.merck-chemicals.de)

Folding efficiency was controlled as described [169], but it was also further confirmed by reproducing folding conditions in a 1 ml scale followed by analytical desalting either through microdialysis or PD-10 (*GE Healthcare*) desalting columns. Effects from the addition of RedOx buffers were also tested, by reproducing promising conditions with either 6 mM GSH to 4 mM GSSG or 9 mM GSH to 1 mM GSSG.

### 5.2.5.3 *Preparative protein refolding through dialysis*

The BabA full length gene product was refolded by dialyzing a sample of purified protein at ~1.5 mg/ml against 10 mM phosphate buffer pH 7.8, 300 mM NaCl, 0.05 % LDAO at a 1:300 ratio overnight. The resulting suspension was centrifuged at 100000g for 30 minutes, concentrated, sterile filtered and kept at 4 °C for further analysis[164].

### 5.2.5.4 *Preparative protein refolding through rapid dilution*

BabA fragments 210, 235 and 255 were refolded by adding a 5 to 8 mg/ml aliquot of purified protein in denaturing buffer 1 to refolding buffer in a 1:25 ratio and incubating for at least 24 h. The sample was then concentrated to ~0.5 ml, desalted with PD-10 columns to phosphate buffer pH 2.5, sterile filtered and kept at 4 °C for further analysis.

### 5.2.6 *Estimation of protein concentration*

Protein concentration was calculated by absorption at 280 nm and usage of the Lambert-Beer equation (eq. 5.1).

$$E = \varepsilon \cdot c \cdot d \quad \text{eq. 5.1}$$

*E: absorbance,  $\varepsilon$ : molar extinction coefficient, d: pathlength*

Using the appropriate protein sequence, the theoretical molar extinction coefficient was calculated by the expasy prot-param server[170], and used for protein concentration estimation.

Alternatively, protein concentrations were measured by the BCA test as described in [171].

### 5.2.7 Protein precipitation

Proteins in denaturing buffers could not be loaded directly onto SDS gels, as the guanidinium ions formed insoluble complexes with SDS, co-precipitating the proteins *in-situ*. Samples from refolding assays and purification under denaturing conditions had to be precipitated and re-solubilized in a compatible buffer prior to SDS-PAGE analysis.

One tenth of the sample volume of 72 % (v/v) trichloroacetic acid (TCA) was pipetted into the sample to a final concentration of 6.5 %. Samples were kept on ice for 1 hour, then centrifuged at 17000g for another hour. Supernatant was discarded, pellets resuspended in protein loading buffer and heated for 5 minutes at 95 °C. If the buffer turned yellow upon contact with the pellet, indicating excessive acidification, drops of 1 M Tris/HCl pH 8 solution were pipetted into the sample, until the color of the indicator turned blue.

### 5.3 Analytical Methods

The standard methods of SDS-PAGE and subsequent staining with Coomassie Brilliant Blue were used as described by Laemmli[172] and Bennett[173].

#### 5.3.1 Western blotting

Western blotting is an analytical method by which proteins immobilized on a nitrocellulose membrane are immunostained with enzyme-coupled antibodies[174]. The advent of protein tagging facilitated enormously the usage of monoclonal antibodies, as they could be synthesized in large scale to react against common tags, such as the His-tag. As all investigated proteins in this work were his-tagged, commercially available monoclonal anti-histag antibodies conjugated with horseradish peroxidase could be used to detect them.

Protein concentration was either estimated by absorption at 280 nm or via a previous Coomassie stained SDS-gel. In both cases, around 5 µg of total protein for each sample were loaded on an SDS-gel, and an SDS-PAGE was run. The gel was not stained, but set in contact with a nitrocellulose membrane in a sandwich setup with four Whatman papers in a semi-dry western transfer system (*PEQLAB*). The membrane and the papers had been previously activated by incubating for 30 minutes in blotting buffer at RT. Transfer took place for 1 h and 15 minutes at 5 V and 62 mA per gel. The membrane was then stained with ink; protein marker and lanes were carefully noted, and the membrane was destained with 0.1 M NaOH and water.

Two washes with TBS and once with TBS-T each for at least 10 minutes followed, before the membrane was immersed in blocking liquid (overnight at 4 °C or for 1 h at RT). The next day, the membrane underwent three further washing steps, before addition of blocking liquid, supplemented with 1:3000 anti-PENTA-His-HRP-conjugate antibody (*Qiagen*). The membrane was then incubated in the antibody solution for 1 h at RT. Antibody solution was kept for further Western blots, while the membrane was slightly dried and prepared for detection by carefully pipetting luminol solution on it and incubating for 2 minutes at RT.

The luminescent membrane was encased in a chamber in which auto-radiographic film was exposed to it for different times, ranging from 1 to 10 min. The exposed films were then developed in a dark room using Kodak developer solution D19, diluted acetic acid and fixing solution, and finally bathed in distilled water to wash off excess fixer.



### 5.3.2 CD spectroscopy

CD spectroscopy is based on the optical isomers' property of interacting differentially with light. When an enantiomeric chromophore interacts with linearly polarized light, the two circularly polarized components of the wavelength will be differently absorbed, thus giving rise to circular dichroism. Although optical activity is normally given by the configuration of asymmetrical carbon centers, polymers composed of optically active monomers, such as proteins, also show distinct polarized-light-interaction behaviors that can be correlated with structural, i.e. conformational, features. The chromophores responsible for protein circular dichroism are the peptide bonds, which interact mainly with far UV light (180 – 260 nm), and aromatic side chains, which are responsible for absorption in the near UV range (260 – 320 nm). The asymmetric environment of peptide bonds is provided mainly by secondary structures, while the tertiary structure is responsible for the differential absorption of aromatic amino acids. Structural information derived from far UV spectra, as the ones presented in this work, are especially useful in determining the overall secondary structure of a protein, while near UV spectra offer more information on the general folding of the protein[175].

So called melting curves are able to provide information about the stability and dynamics of a protein fold. Melting curves are executed by measuring ellipticity, the traditional signal of circular dichroism, at a single wavelength versus temperature. As protein folding is temperature-dependent, and absorption behavior is fold-dependent, it is possible to track protein denaturation by this method. A melting curve provides then both information on the stability of the protein, through the melting point, and on the quality of folding, due to its curve shape[176].

Both, far UV spectra and melting curves, were measured on a JA-18 CD spectrometer (*Jasco*) for the assessment of fold quality under different conditions, and to assess environment-dependent conformational changes undergone by sample proteins. To assess folding quality, proteins were diluted into double distilled water to a final concentration of between 0.1 to 1 mg/ml, while for the evaluation of buffer dependent conformational changes, different methods were employed. Either the proteins in storage buffer were desalted over illustra MicroSpin G50 desalting columns (Amersham) into transition buffer, or they were first diluted to 2 mg/ml into the transition buffer and then incubated for at least 2 h. In either case they were finally diluted to 0.1 - 0.3 mg/ml in double distilled water. Spectra were measured with settings as shown in table 5.9. Melting curves were measured directly afterward (table 5.10), finally a spectrum was taken after the cooling

procedure at both the initial and final temperature.

Parameter	Value
Wavelength range	190 – 260 nm
Data pitch	0.2 nm
Band width	1 nm
Response	1 sec
Sensitivity	Standard
Scanning speed	50 nm/min
Accumulation	3
Temperature	10 °C / 95 °C

Table 5.9: Settings for far UV CD spectrum measurement

Parameter	Value
Range	10 – 95 °C
Data pitch	0.1 °C
Band width	2 nm
Response	1 sec
Sensitivity	Standard
Temperature slope	2 °C/min
Monitored wavelength	208 nm

Table 5.10: Settings for melting curve analyses, both heating and cooling

For data processing, observed ellipticities were converted to *mean residue ellipticities* :

$$[\theta]_{(mrw, \lambda)} = \frac{MRW \cdot \theta}{10 \cdot d \cdot c} \quad \text{eq. 5.2:}$$

$[\theta]_{(mrw, \lambda)}$ : mean residue ellipticity  
 MRW: mean residue weight ( $\frac{\text{protein molecular weight}}{\text{number of residues} - 1}$ )  
 d: pathway in cm  
 c: protein concentration in g/ml.

### 5.3.3 Analytical gel filtration

Gel filtration can provide information about large conformational changes within a protein, specially differences in oligomerization in response to buffer changes. Analytical gel filtration differed from the preparative one mainly in bed and sample volume. The Superdex 200 150/10GL (GE Healthcare) employed for these experiments has a total volume of ~3 ml, while the sample volume was kept at 50  $\mu$ l. 50  $\mu$ l of 1 – 3 mg/ml protein solutions where loaded into the pre-equilibrated column and run at 0.3 ml/min (table 5.11). All runs were performed at the Äkta purifier, as it was the only system compatible with the analytical column. Elution was followed by absorption measurements at 280 nm.

Protein	Buffer
Epa1A,	AM
Epa1→3A,	AML
Epa1→6A	AME
BabA fragments	Phosphate buffer pH 5.8
	Phosphate buffer pH 2.5

Table 5.11: Buffer employed for analytical gel filtrations for each protein.

#### 5.3.4 High throughput glycan binding studies at the Consortium for Functional Glycomics.

The Consortium for Functional Glycomics (The Scripps Research Institute, La Jolla, USA) provides extensive glycan array technologies which permit high throughput qualitative screening of protein-carbohydrate interactions for a wide variety of carbohydrates (Appendix I).

The arrays are printed on N-hydroxysuccinimide (NHS) activated glass slides by dripping ~0.6 nl amine coupled glycan solution in printing buffer (300 mM phosphate pH 8.5, 0.005% Tween-20) using automated equipment (Figure 5.4 shows the NHS amine coupling reaction). After 30 minutes reaction time under an 80% humidity atmosphere, a blocking step of 1 hour with 50 mM ethanolamine in 50 mM borate buffer pH 9.2 ended the reaction. Slides are then thoroughly rinsed with water, dried and stored in desiccators at room temperature before use[177].

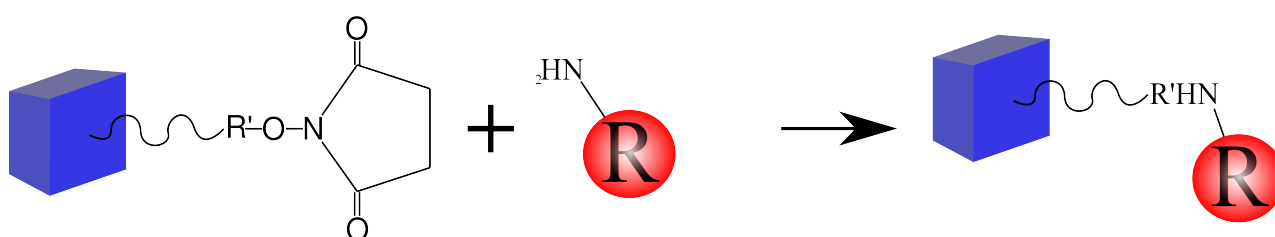
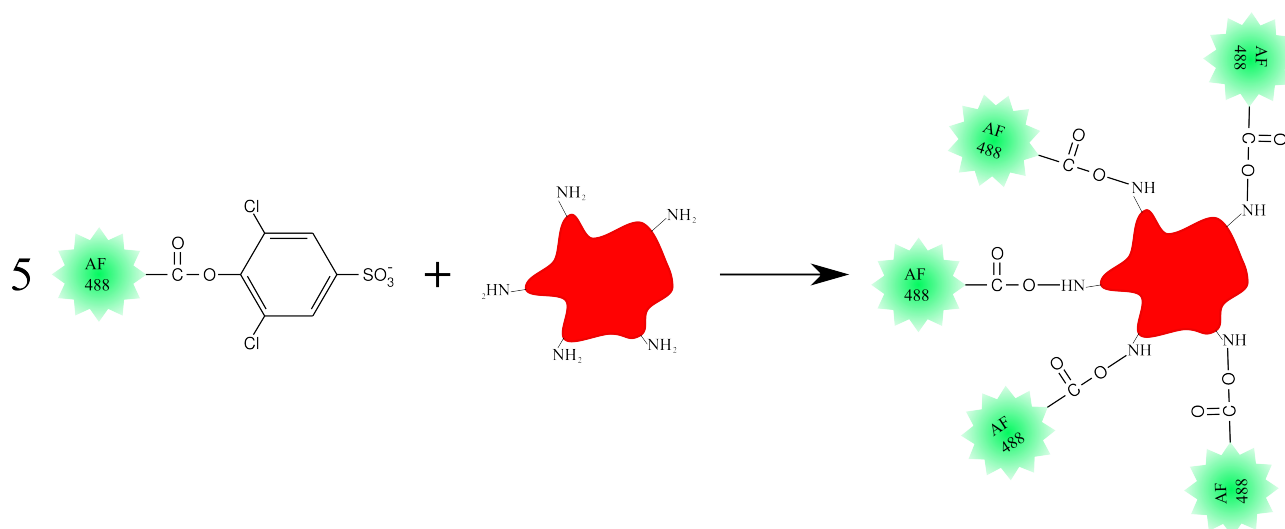


Figure 5.4: NHS activated glass surface reacts with amine coupled carbohydrate. N-hydroxysuccinimide (NHS) is shown coupled to a blue block of glass. A primary amine coupled to a carbohydrate R acts as a nucleophile, substituting NHS, and producing an amide bond to the glass surface.

BabA 235 was sent to the CFG for pH-dependent carbohydrate-binding screening in phosphate buffer, pH 2.5, at a concentration of 6 mg/ml. The protein was diluted directly before the assay to a final concentration of 200  $\mu$ g/ml either in phosphate buffer pH 2.5 or phosphate buffer pH 5.8 and incubated for 2 hours. Binding experiments were performed on the cfg glycan array chip V 3.2 and monitored using fluorescently labeled anti-6xHis antibodies.

Epa1A and variants were marked with the fluorescent dye Alexafluor 488 SPD (Invitrogen)[178] (Figure 5.5). 4 mg/ml protein solutions in 0.1 M sodium bicarbonate buffer pH 8.3 were mixed with 7  $\mu$ l of 10 mg/ml Alexafluor 488 SPD solution in DMF for 2 hours at RT. Residual dye was eliminated with MicroSpin G50 desalting columns (Amersham biosciences) pre-equilibrated with AME buffer. Measurements were performed on the cfg glycan array chip V4.1 at concentrations of 200  $\mu$ g/ml for all proteins and 50  $\mu$ g/ml and 1  $\mu$ g/ml additionally for Epa1A. Relative binding strength was measured directly over the signal from the fluorescent dye.



*Figure 5.5: Reaction of Alexa fluor 488 Sulphodichlorophenyl (SPD) ester with a sample protein. The SPD group acts as a good salient group in a reaction at RT in basic aqueous medium in which the primary amines of a protein (red) attack nucleophilically the carbonyl group of the ester. The result is a coupling of the Alexa Fluor fluorophore on lysins and the N-terminus of almost any protein[178].*

The experiment was performed by adding  $\sim$ 50  $\mu$ l of protein solution on the slide directly and covering it with a microscope cover slip. Samples were incubated for 30 to 60 minutes in a humidified chamber, after which they were washed frequently with the corresponding buffer, PBS buffer and PBS buffer containing 0.05% Tween-20 and deionized water. If the sample was to be set in contact with a labeled antibody. The antibody was given at this point and the procedure was repeated[177]. Detection of Alexafluor 488 took place at 495 nm excitation wavelength and 519 nm emission wavelength.

### 5.3.5 Fluorescence spectroscopy

Intrinsic fluorescence occurs in proteins lacking a cofactor when aromatic amino acids are excited by UV light (280 nm for tyrosines, 295 nm for tryptophanes) with emission maxima at around 320 - 345 nm. Whenever an aromatic molecule absorbs a photon, it may relax either radiantly or non-radiantly. The ratio of emitted photons per absorbed photon is the quantum yield (Q) of the fluorophore. When a molecule interacts with the excited fluorophore, the molecule may relax it non-radiantly, thus decreasing the quantum yield; this phenomenon is known as quenching, and such a molecule is called a quencher. [179]

Ligands are in general good quenchers of protein fluorescence, and since the quench intensity is proportional to the concentration of the protein ligand complex, it is possible to calculate affinity constants by fluorescence titration. Usually, fluorescence titration is monitored by changes in the quenching percentage displayed by the selected protein (equation 5.3).

$$q_{(c)} = \left(1 - \frac{FI_{(c)}}{FI_{(0)}}\right) \cdot 100 \quad \text{eq. 5.3}$$

*q<sub>(c)</sub>*: percentage of quench for ligand concentration *c*  
*FI<sub>(c)</sub>*: fluorescence intensity for concentration *c*  
*FI<sub>(0)</sub>*: fluorescence intensity at concentration 0

The quench is normally calculated for the wavelength with maximum emission, which is usually between 343 and 347 nm.

In order to perform a titration assay, two protein solutions were prepared in fluorescence buffer. While one of them was carbohydrate free, the other contained the carbohydrate against which the titration was to be performed (tables 5.12 and 5.13).

Lactose titration		
Component	Protein solution	Ligand solution
Protein	0.3 mg/ml	0.3 mg/ml
Glycan	0 mM	0.5/5/50 mM
CaCl <sub>2</sub>	5 mM	5 mM
EDTA	1 mM	1 mM

Table 5.12: Typical composition of the solutions necessary for lactose titration

T-antigen titration		
Component	Protein solution	Ligand solution
Protein	0.01 mg/ml	0.01 mg/ml
Glycan	0 mM	0.1/1/10 mM
CaCl <sub>2</sub>	5 mM	5 mM
EDTA	1 mM	1 mM

Table 5.13: Typical composition of the solutions necessary for T-antigen titration

The titrations took place at an FP-6500 fluorescence spectrometer (*Jasco*) with settings as per table 5.14. 500  $\mu$ l of carbohydrate-free protein solution were pipetted into a round fluorescence cell (*Hellma*) containing a magnetic stirrer. Carbohydrate containing solutions were kept on ice, and pipetted into the round cell (table 5.15). The stirrer was activated for a total of 90 sec, followed by 30 sec incubation and the spectrum measurement.

Dissociation constants ( $K_d$ ) were calculated by plotting carbohydrate concentration against percentage of quench with either the total binding equation (equation 5.4).

$$q(c) = \frac{q_{max} \cdot c}{K_d + c} + NS \cdot c \quad eq. 5.4$$

$K_d$ : Dissociation constant NS: non-specific binding constant.

The total binding equation accounts for specific binding, but also for non-specific, linear binding that might occur after saturation of specific sites. The non specific binding constant (NS) is a

measure of non-specific binding strength. High standard errors for the NS constant may indicate that the non-specific binding is negligible.

Scatchard plots were performed as a quality control for the fitting systems. A Scatchard plot is performed by opposing quench on the x-axis against quench divided by concentration on the y-axis. The resulting representation has been demonstrated to present a straight line defined by a  $-1/K_d$  slope and the value  $q_{max}/K_d$  at the X-axis intersection. Non-linear Scatchard plots are indicative either of cooperativity, or of the presence of different types of binding events in the solution. Although it is theoretically possible to perform a proper fit of a non-linear Scatchard plot by approximating it to two separate linear plots and some relatively sophisticated manual adjustments[180], a non-linear regression of the raw data is more reliable, and statistically sound. In order to perform a non-linear regression for a data series that produced a non linear Scatchard plot, equation 5.5 is used.

$$q_{(c)} = \frac{q_{max1} \cdot c}{K_{d1} + c} + \frac{q_{max2} \cdot c}{K_{d2} + c} \quad eq. 5.5$$

Equation 5.5 assumes that the system presents two different kinds of binding sites, which could either represent a monodisperse protein solution where the protein presents a primary and a secondary binding site, or a non-monodisperse solution, where a protein with a single binding site is in two different states[181]. The results is that a single system is defined by two dissociation constants,  $K_{d1}$  and  $K_{d2}$ . The challenge while using such an equation for non-linear regressions is that there are many possible solutions, most of which do not have any biological relevance. In order to improve the fitting process, it is important to obtain good starting values for all unknown parameters. Such values were calculated using the preliminary constants derived from linear regressions of the data as presented in the Scatchard plot.

During the process of evaluating the experimental data, the simpler model offered by equation 5.4 was preferred, as long as the  $R^2$  value was above 0.95 and the resulting fit did not show strong systematic residuals.

Parameter	Value
Excitation wavelength	295 nm
Emission range	300 – 400 nm
Band width excitation	3 nm
Band width emission	5 nm
Response	0.5 sec
Sensitivity	Medium or high
Data pitch	0.5 nm
Scanning speed	500 nm/min
Accumulation	3
Temperature	10 °C

*Table 5.15: Fluorescence spectrometer settings*

carbohydrate stock (mM)	Volume of stock ( $\mu$ l)	Final concentration ( $\mu$ M)
0	0	0
0.5	5	4.99
0.5	5	9.88
0.5	10	19.4
0.5	15	33
0.5	20	49.9
0.5	30	73.1
5	4	107
5	6	157
5	8	221
5	12	315
5	16	435
5	24	603
50	20	2080

*Table 5.14: Sample titration of epa1A against lactose. Protein concentration was kept constant, while the carbohydrate ligand was varied.*



## 5.4 Crystallographic Methods

### 5.4.1 Crystallization screening.

Reproducibility of crystals and diffraction quality can be important limiting factors when determining the crystal structure of a protein. Initial screening conditions were purchased from *Qiagen*, and utilized the sparse matrix screening strategy [182], each of the screens tested (see reference [183]) contained 96 different conditions. The sparse matrix methodology demands the use of aqueous solutions of different buffers, precipitants and salts that have been empirically shown to lead to successful crystal growth. This method has proven its usefulness time and again, as most commercial and initial screens follow the sparse matrix strategy nowadays, but they can be fundamentally flawed. Similar proteins might crystallize in similar yet markedly different conditions, it is therefore possible to miss crystallization conditions when using a sparse matrix strategy, because no systematic approach is taken to explore changes in pH, precipitant and or additive concentration, within a certain condition. Initial crystallization trials took place under a sitting drop, vapor diffusion setup and were pipetted with a *Microsys*<sup>TM</sup> SQ4000 (*Genomic Solutions*) crystallization robot.

80  $\mu$ l of each of the 96 crystallization conditions composing a screen were pipetted into their corresponding reservoir of an *Innova Plate* (*Hampton Research*) with an *Eppendorf multipette*, which was then loaded on the crystallization robot. The device pipetted 300 nl drops of the reservoir into each of the two respective wells on the plate. Finally, the robot pipetted 300 nl of either protein or buffer solution into the freshly pipetted drops. The setup was then isolated by sealing the plate with *VIEWseal*<sup>TM</sup> (*Greiner Bio-one*) adhesive transparent plastic sheets. The finished plate was then stored at either 18 °C or 4 °C, as temperature does affect crystal growth. Table 5.16 shows the standard setups for the different proteins tested for crystallization during this work.

Parameter	Epa1 A domain and variants	SabA 260	BabA fragments
Upper drop	Protein	Protein	Protein
Lower drop	Protein (half concentrated)	Buffer	Buffer
Concentration (mg/ml)	12	20	3; 15*
Protein buffer	AML, AME or AM	AM buffer	Phosphate buffer pH 5.8 and 2.5
Additives	20 mM lactoNbiose I for co-crystallization and 20 mM CaCl <sub>2</sub> if AME was used.		
Screens tested	Classic, Classic Lite, Cryos, Ammonium sulfate.	JCSG+, PACT, MBCI, MBCII, Anions,	

*Table 5.16: Screens and setups tested for crystallization of the different proteins. \*BabA had a maximum solubility at pH 5.8 of 3 mg/ml, thus crystallization at pH 2.5 was tested at 15 mg/ml and at 3 mg/ml for pH 5.8.*

Prior to storage, plates were documented by microscopic observation for crystals, and then at exponentially increasing times for up to 2 years after dispensing.

#### 5.4.2 Optimizing and reproducing crystallization

Conditions that lead to crystals during initial screening were firstly reproduced and then optimized. To reproduce the crystals, 0.75 – 1 ml of the same screen solutions from, if possible, the same production batch, were dispensed in 24 well VDX plates (*Hampton Research or Nextal*). Different drop sizes and protein-to-buffer-to-reservoir solution ratios were tested, as these can vastly improve crystal quality. Drops of protein-reservoir mix were then produced on cover-slides, table 5.17 shows the different types of drops prepared during reproduction.

Volume (in $\mu$ l, Protein/Buffer/Reservoir)
1/0/1
2/0/2
2/1/2

*Table 5.17: Drop sizes and ratios used during reproduction.*

The conditions yielding the most promising initial crystals within 1 week after pipetting were optimized. Optimization consists of performing a screen surrounding the initial hit in which such parameters as pH, precipitant, salt concentration or additive composition are tested. Tables 5.18 and

5.19 show the two main optimizations performed for the Epa1 A domain and variants.

Composition	Original	Optimization steps
0.1 M Mes (pH)	6.5	5.5/6/6.5/6.8
Ammonium sulfate concentration (M)	0.2	0.1/0.2/0.3/0.4
PEG 5000 MME (% in v/v)	30	22.5/25/27.5/30/32.5

Table 5.18: Optimization based on condition 92 from the EasyXtal Classics suite (Qiagen).

Composition	Original	Optimization steps
0.1 M Sodium acetate (pH)	4.6	4.3/4.6/4.9/5.2
Ammonium sulfate concentration (M)	0.2	0.1/0.2/0.3/0.4
PEG 4000 (% in v/v)	25	17.5/20/22.5/25/27.5/30

Table 5.19: Optimization based on condition 85 from the EasyXtal Classics suite (Qiagen).

#### 5.4.3 Diffractometric measurements and sample preparation.

Protein crystals are measured at very low temperatures ( $\sim 100$  K) to avoid radiation damage during the diffraction experiment. If ice crystals are formed during the freezing step, the structural integrity of the protein crystal might be compromised. It is therefore necessary to cryoprotect the crystals with the help of either such anti-freezing agents as glycerol or PEG 400, or high salt concentrations. Table 5.20 shows all the cryoprotectant solutions employed for diffraction experiments.

In order to prepare the sample, the crystals were first fished out of the mother liquor in either Cryo-Loops or MicroMounts (*Hampton*) and then soaked for no more than 30 seconds in cryoprotectant solution. Finally, the samples were frozen in liquid nitrogen and stored in such for measurement.

Cryoprotectant	concentration
Glycerol	15 – 30 % (v/v)
PEG 400	30 – 40 % (w/v)
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	1.5 M
N-acetyl-galactosamine	1 M

Table 5.20: Cryoprotectants and corresponding concentrations. All cryoprotectants were dissolved into solutions with as similar a composition to the original mother liquor of the crystal native condition as possible. The given concentration of ammonium sulfate is the final concentration.

The measurement of protein crystals took place either at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at beamlines ID14-2 or ID23-2, or in-house using the X-ray diffractometer. The crystal containing Cryo-Loop was transferred from the dewar container to the device's goniometer head and kept refrigerated with a 100 K nitrogen stream. The crystal was centered in both cases in such a way that its center remained within the incident radiation path for at least a  $180^\circ$  rotation of the goniometer head. Test images were taken either with the MXCubE contained DNA[184] (ESRF) or the in-house Mar-provided software at  $90^\circ$  from each other, and a strategy was developed using mosflm [185]. The crystal was centered for the corresponding angle range as indicated by the strategy, and images were collected with the recommended exposure time (Table 5.21).

Parameter	Epa1A	Epa1→6A	Epa1→3A	Epa1→2A
$\Delta\phi$ (deg)	0.3	0.2	0.5	0.3
$\phi$ range (deg)	110	120	98	85
Exposure time (sec)	20	0.5	5	480
Resolution (Å)	1.467 – 23.602	1.487 – 39.698	1.666 – 27.920	1.927 – 49.355
Location	ESRF/ID14-2	ESRF/ID14-4	ESRF/ID14-2	In House/Rotating anode

*Table 5.21: Data collection parameters for all datasets.*

## **5.5 Computer assisted methods**

A wide variety of computer assisted methods were used for very different purposes. For purposes of clarity, they will be divided into homology model building methods, structure elucidation and evaluation methods, and methods for calculating protein properties.

### **5.5.1 Homology model building methods**

Homology modeling is a very useful tool in both the solution of the phase problem, and in generating *in silico* models of proteins for which no experimental structure information is available. In both cases, a protein of similar sequence (at least 30% identity in most cases) and known structure is used as a template for the generation of model of the target protein. The resulting model can be used to determine initial phases for phase solution, or for the study of variants or paralogues for which there is no structural information. For the Epa1A project, homology models were fundamental in both aspects.

A homology model was built using the program Chainsaw from the ccp4 suite[186] to obtain initial phases for molecular replacement. Homology models of several paralogues of the Epa family were developed based on Ep1A using the program modeller[187] in order to study the plasticity and specificity of the Epa A domain carbohydrate binding pocket.

In any homology modeling effort work is divided into 5 steps[188][189];

1. Identification of homologue(s) that can be used as template(s) for modeling
2. Generating of a sequence alignment between target and template.
3. Building of the initial model
4. Evaluation of the model
5. Using the model in a research environment, either to interpret previous data, or to devise new experiments.

#### **5.5.1.1 Identification of homologues and choosing a template for modeling**

In order to produce an accurate model that will deliver reliable predictions, choosing the correct template is of capital importance. Assessing the *correctness* of a template is a difficult task, as no single clear criterion can be used [190]. The main consideration when choosing a template is

normally sequence identity, with at least 70% sequence identity needed between template and target for very accurate homology modeling, and at least 30% for a low confidence structure[190]. Other factors can also be taken into account, like biochemical information or secondary structure predictions. In this way, distant homologues within a family or a superfamily can also be used for successful homology building with the proper alignment[191], although with considerably greater effort.

In the case of the phase solution of Epa1A, the A domain of the Flo5 *Saccharomyces cerevisiae* flocculin[1] was chosen as template (Figure 5.6 PDB accession code 2XJP). Even though sequence identity was below 30%, both proteins had been characterized as PA14 domains[65][64], had similar far UV-CD spectra and showed only predicted  $\beta$  secondary elements in their structures[155][192]. The structure of Flo5A was known, as well as the one for the *B. anthracis* PA14 domain (PDB ID 1ACC)[141]. The core structures of Flo5A and the PA14 domain are quite similar (RMSD=2.98Å<sup>2</sup> over 67 common C $\alpha$  atoms), even though the sequence identity is so low that BLAST is not capable of finding any significant similarity. Finally, both Flo5 and Epa1 had been characterized as C-type lectins and contained the very characteristic DcisD motif[65][192][1]. The Epa1A and Flo5A modeling pair is one of the cases where, even with a very low sequence identity, structural and functional evidence overwhelmingly points to a common, conserved core structure upon which a homology model could be based.

The aim of the Epa1A variants project was to generate mutations within the Epa1A binding pocket to study the diversification of carbohydrate specificity within the Epa family. Since no prior structural information was available for any of the members of the Epa family, building homology models was the best tool with which to obtain reliable information on the specificity determinants within the Epa binding pocket. Under conditions where the template is already given, the next step is to obtain a set of target sequences for which reliable structure models can be calculated. As sequence identity is the most important criterion for assessing the feasibility of template based modeling, establishing developmental proximity to the template was the first step in obtaining suitable targets.

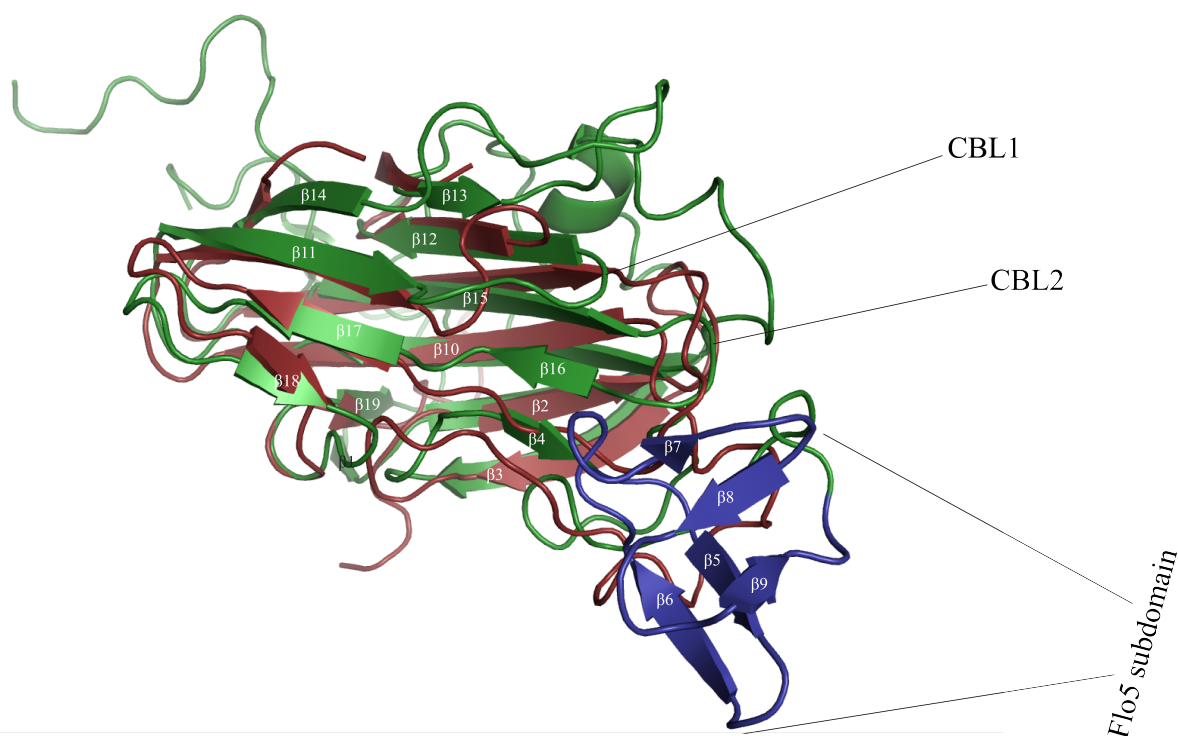


Figure 5.6: Structural alignment of Flo5A[1] and the *B. anthracis* PA14 domain[141]. The core of the Flo5A domain is shown in green, while the core of the *B. anthracis* PA14 domain is shown in red. Features of both structures that do not align well with each other, and that were therefore left out of the alignment, are shown in blue.  $\beta$  sheets nomenclature is based on Flo5A multiple alignments[1], as well as loops CBL1 and CBL2, that compose the calcium binding pocket and the Flo5 subdomain. Alignment was performed under pymol with the super command, which minimizes root mean square distances (RMSD) between the structures to be aligned without previously performing a sequence alignment, as the align command does. The core (represented respectively in green and red for Flo5A and *B. anthracis* PA14), composed of sheets 4 and 11 to 19 is structurally well conserved, with an RMSD of  $2.98\text{\AA}^2$  over 67 common Ca atoms, the Flo5A subdomain is a specific feature of this protein and is not predicted to be in any other PA14 containing structure except for Flo1A[1].

The Basic Local Alignment Search Tool (BLAST) from NCBI [143][144] was employed using the blastp algorithm and Epa1 A domain as query to search for suitable candidates within the protein sequences of *Candida glabrata* in the non-redundant (nr) protein sequence database. The nr database is composed of all non-redundant GenBank CDS translations, RefSeq Proteins database, the Protein Data Bank, SwissProt database, the Protein Information Resource (PIR) data-bank and the Protein Research Foundation data-bank[144].

The blastp algorithm is based on the original Basic Local Alignment Tool as applied to proteins[193][143]. BLAST employs a heuristic method to simulate dynamic programming algorithms[194] at a vastly diminished processing cost. In the first step of a search, blastp compares

all possible 3 amino acid long peptides with the query and produces list of homologous peptides. Each peptide is scored according to PAM120[195] or BLOSUM 62 [196]based substitution matrices. Each peptide obtains a score  $t$ , which is highest for peptides found within the sequence, and decreases proportionally to similarity. The  $t$  score is then compared with a  $T$  cutoff of 11, if the  $t$  score is 11 or more, the peptide is considered to be homologous to the query, if it is below, it is discarded. Homologous peptides are then used as queries within the chosen data-bank for subjects. Whenever an exact peptide is found within a subject, it is considered a hit. Blastp then determines whether it is at most within 40 amino acids of a second hit. If it is, it compares the amino acids surrounding the hits in the query and the subject sequentially, trying to maximize the sequence score through elongation. The scoring parameters are the same as for the generation of the initial peptide list, with the exception that gaps within the compared sequence are allowed, but penalized by the scoring matrix with -11 for opening and -1 for extension by one amino acid Each of the regions where query and subject align is called a maximal segment pair (MSP) and receives an MSP score, as previously described. If MSP scores are equal or greater than a cutoff value  $S$  (55 as default), the maximal segment pair is considered significantly homologous, and therefore a part of a correct alignment.

Four scores are then assigned to each query to subject pair:

1. Bit Score: the bit score is the addition of all MSP scores within the alignment that surpass the cutoff value  $S$ .
2. Identity score: the number of identical amino acids in the alignment divided by the number of total amino acids This value is given either as a quotient, or as a percentage.
3. Positives score: the number of homologous and identical amino acids divided by the number of total amino acids This value is also given either as a quotient, or as a percentage.
4. Gap score: the number of gaps introduced in the shorter chain of an alignment divided between the length of the longer chain. This value is also given either as a quotient, or as a percentage.

Although the bit score is the internal score employed for performing the pairwise alignment, it is difficult to compare this value between different alignments without a deep knowledge of the mathematical processing of the scoring matrix. The identity or positive percentages are much more widely used scores to give the degree of goodness of an alignment performed with BLAST.



Blast works extremely well under these parameters when the objective of the search is to find sequences where conserved amino acids are spread evenly, even if they are few. Due to the nature of the scoring process, however, if the conserved amino acids between query and subject are concentrated in short regions within the entire sequence, many of these regions can be passed over (up to 11% of the total search), and some real targets can be missed[143].

For the purposes of choosing homology candidates, a 45% identity cutoff in the A domain was adopted for any molecule of *C. glabrata* within the Epa family, as this is an average value for high quality homology structures[197].

#### ***5.5.1.2 Generating of a sequence alignment and a phylogenetic tree between target(s) and template(s)***

Once Flo5A was selected as template, sequence alignment was the starting point in generating a homology model. The sequence alignment is key to the reliability of the entire process, as small differences within the alignment can lead to shifts within the overall structure.

The program Chainsaw[186] does not provide a built-in alignment system, therefore, an external program had to be used. CLUSTALW[198][199] was the program of choice, as it is widely used and presents a very user-friendly GUI in the form of CLUSTALX[200]. The CLUSTALW algorithm is similar to the blastp algorithm[190], but the scoring matrix is more sophisticated. The most important difference in a pairwise alignment, where two sequences or alignments are aligned against each other, is that gap penalties are not only dependent on whether the gap is being opened or extended, but also on the amino acids involved in opening the gap and those which will pair with the gap[201]. The amino acid dependent gap penalty modulation reduces the penalty for amino acids frequently found on loops, while it increases the penalty for amino acids frequently found on secondary structures (equation 5.6). The modulation factors were originally calculated from known structures, and they cluster hydrophilic amino acids as those more prone to allow the opening and extension of gaps, while the hydrophobic ones tend to prevent it.

$$P = \begin{cases} 0 & \text{if } i=0, i=M \\ Ag[i] + Bg[j+1] + \sum_{l=j+1}^{l=j+k} Bh[l] & \text{if } 0 < i < M \end{cases} \quad \text{eq. 5.6:}$$

*P*: penalty assigned for a given gap at position *i* of sequence *A* corresponding to position *j* in sequence *B*.

*Ag*[*i*]: gap opening penalty on position *i* of sequence *A*.

*Bg*[*j*+1]: gap opening penalty on position *j*+1 on sequence *B*.

*Bh*[*l*]: gap extension penalty for all positions of sequence *B* within the gap.

*M*= last position of sequence *A*.

CLUSTALW is capable of doing multiple alignments, where up to thousands of molecules can be paired to increase reliability. CLUSTALW aligns multiple sequences by performing exhaustive pairwise alignments. It determines the most similar pairs, and orders them in a guide tree (Fig. 5.7) in order of similarity, then performs pairwise alignments of the paired sequences (step 1 Fig. 5.7) and recursively performs pairwise alignments with the alignments of the previous step (steps 2 to *n*, Fig. 5.7). In this manner, redundant evaluation of sequences is averted, and model bias towards the most similar molecules prevented[190].

Both features increase the reliability of the CLUSTALW method above the BLAST method, at the cost of increased CPU time. CLUSTALW is a very good algorithm for performing alignments of already handpicked sequences, but it cannot perform data-bank searches as quickly as BLAST can. Another important advantage of CLUSTALW above BLAST is that the former allows for human interaction during the

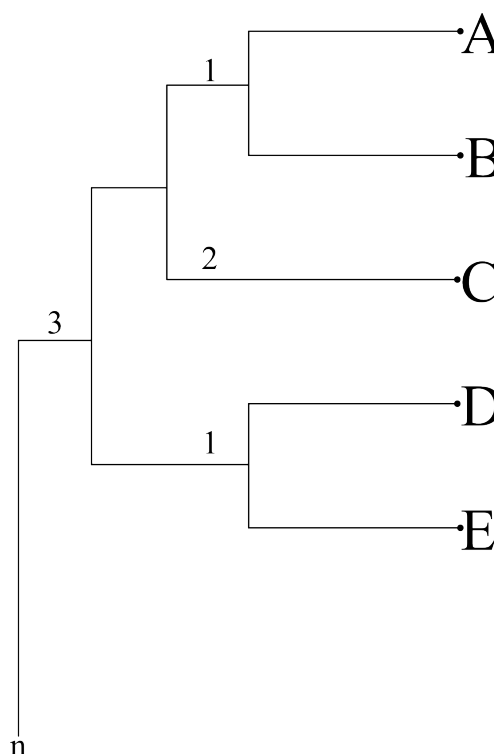


Figure 5.7: Guide tree as employed by CLUSTALW during multiple alignment. Sequences A, B, C, D and E are first ordered in pairs by overall similitude. Then, a pairwise alignment is performed for each pair (step 1). The resulting alignments of step 1 are then aligned against each other or against other sequences in order of similitude (steps 2 and 3) until the multiple alignment is complete. Diagram based on [190]

alignment generation phase, which allows the user to introduce chemical, biochemical or structural information that would otherwise be missed in a fully automated setup.

The A Domain sequences of all Epa paralogues were also processed in this manner, so as to obtain an Epa family multiple alignment.

At his point, the desired alignment of Epa1A and Flo5A for CHAINSAW was finished, but the evaluation of the paralogues of the Epa family was not. The reason was that, to obtain a full model of the evolutionary relationship within the selected members of the Epa family, it was necessary to perform a phylogenetic tree.

The CLUSTALX interface is also implemented with the capacity to build phylogenetic trees based on the multiple alignments generated in previous steps. The first step in generating a phylogenetic tree is to obtain a distance matrix of the sequences. CLUSTALX uses the percent of divergence between all sequences in the multiple alignment as the values within the distant matrix. The percent of divergence is calculated as the number of substitutions and deletions between a pair of sequences divided by the length of the longest sequence. Next, the phylogenetic tree is built with the Neighbor Joining (NJ) method[202][203]. The NJ method is an algorithm designed to select two sequences or groups of sequences (taxon) within a distance matrix, build a new subtree  $u$  and agglomerate the two selected taxa within the new subtree  $u$ , thus reducing the taxon set by 1 (Fig. 5.8)[203]. NJ selects the two taxa that minimize the  $Q$  criterion (equation 5.7) which takes its minimum value when two taxa are very near each other, and very distant from all other taxa within the taxon set.

$$Q(i, j) = (r - 2)d(i, j) - \sum_{k=1}^r d(i, k) - \sum_{k=1}^r d(j, k) \quad \text{eq. 5.7}$$

$Q(i, j)$ :  $Q$  score for taxa  $i$  and  $j$ .       $r$ : current size of the taxon set.

$d(i, j)$ : distance between taxa  $i$  and  $j$ .

$d(i, k)$ : distance between taxon  $i$  and any one taxon  $k$  from the taxon set of size  $r$ .

$d(j, k)$ : distance between taxon  $j$  and any one taxon  $k$  from the taxon set of size  $r$ .

After a taxon pair is selected and agglomerated within a new subtree  $u$ , the new distances between the joined taxa  $i$  and  $j$  and subtree  $u$ , and between subtree  $u$  and all other taxa,  $d(i, u)$ ,  $d(j, u)$  and  $d(u, k)$  need to be calculated (equations 5.8 and 5.9).

$$d(i, u) = \frac{1}{2}d(i, j) + \frac{1}{2(r-2)} \left[ \sum_{k=1}^r d(i, k) - \sum_{k=1}^r d(j, k) \right] \quad \text{eq. 5.8}$$

*d(i,u)* distance between new subtree *u* and taxon *i*

As can be seen in equation 5.8, distance between the new subtree and one of the taxa within it depends on the former distance between the joined taxa and the difference of their corresponding distances to all the rest of taxa within the taxon set. This is a good measure of the joint difference between the two taxa. The distance  $d(j,u)$  is calculated symmetrically.

$$d(u, k) = \frac{1}{2} [d(i, k) - d(i, u)] + \frac{1}{2} [d(j, k) - d(j, u)] \quad \text{eq. 5.9}$$

The distance  $d(u,k)$  is simply calculated as the average of the difference between the distances of the joined taxa towards the new subtree *u* and towards all other taxa within the taxon set.

Although it is not stated within the algorithm explicitly, it could be shown that the NJ method minimizes whole tree length *l*, as computed using the generalized Pauplin formula (equation 5.10) [204]. An example of the calculation of the generalized Pauplin formula weighting term  $w(i,j)$  can be seen in Fig. 5.8. The biological relevance of the principle of whole tree length minimization has been partly corroborated [205], but as with every phylogeny method, there is no guarantee that the generated tree is the actual best tree.

$$R = \frac{\sum \|F_{obs} - F_{calc}\|}{\sum |F_{obs}|} \quad \text{eq. 5.10}$$

*w(i,j)*: weighting term for the distance between two taxa *i* and *j* within a phylogenetic tree.  
*{ij}*: all possible combinations of taxa within the taxon set

Generalized Pauplin weighting term

$$w(i'j') = \frac{1}{n_i \cdot \dots \cdot n_j}$$

eq. 5.11

Distance	Weighting formula	Numerical result
$d(i'j')$	$w(i'j') = \frac{1}{n_i \cdot n_u \cdot n_j}$	$Z = \frac{LLG_o - LLG_{ro}}{RMS_{LLGrS}}$
$d(e''j'')$	$w(e''j'') = \frac{1}{n_e \cdot n_a \cdot n_k \cdot n_u \cdot n_j}$	$w(e''j'') = \frac{1}{2 \cdot 2 \cdot 5 \cdot 2 \cdot 2} = \frac{1}{80}$

Table 5.22: Calculation of the generalized Pauplin weighting term. The term  $w(i'j')$  is calculated as one over the multiplication of the number of branches  $n$  in each of the nodes within the distance. Two examples are given, as shown in green and red dotted lines on figure 5.8.

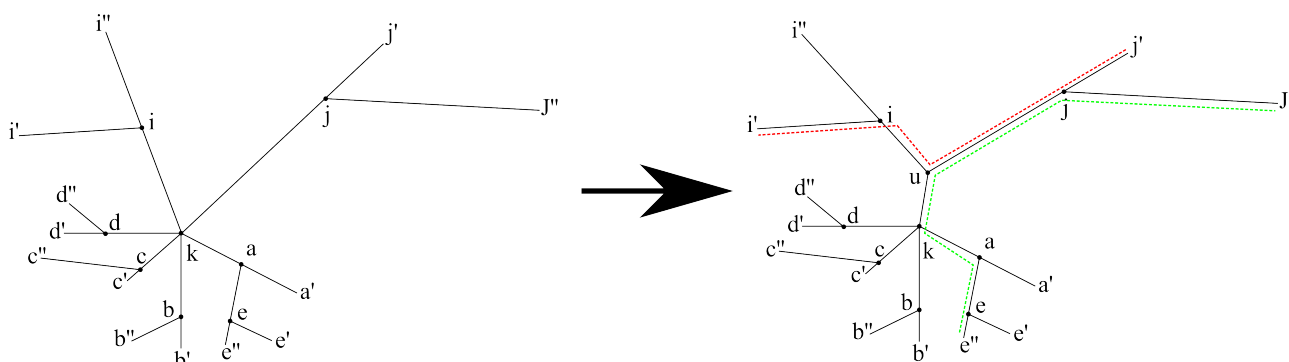


Figure 5.8: The neighbor joining method. a) The NJ algorithm selects the two taxa  $i$  and  $j$  that minimize the  $Q$  criterion (equation 5.7) by being near each other and distant to the rest of the taxon set  $a$ - $e$ . It then joins the two into subtree  $u$  and calculates resulting distances according to equations 5.8 and 5.9. b) the generalized Pauplin weighting term is calculated for distance  $i'j'$  (red dotted line) and  $e''J''$  (green dotted line). The weighting term is calculated by counting the number of branches at each of the nodes along the distance, multiplying them together and dividing 1 by that number (equation 5.11).

### 5.5.1.3 Building a model for molecular replacement

A multiple alignment of several PA14 domain containing sequences, including the Epa1A and Flo5A pair, and generated with CLUSTALW was used to produce a Chainsaw model. Only the aforementioned pair was used as a sequence guide for the Chainsaw program. Chainsaw uses a very simple algorithm to generate a homology model. It will use the alignment and the template structure

as input, and crop the template structure to eliminate features that will probably not be found within the target[186]. In its default mode, also called Mixed S mode, chainsaw eliminates any residues within the template structure, which are represented by gaps in the target. Any gaps within the template sequence are edited as gaps within the target structure. Amino acids that are not identical in the alignment are cropped to their  $\gamma$ -atom, which is then changed to either carbon, oxygen, nitrogen or sulfur to match the targets sequence. The resulting model is a so-called 'mixed model'[206] and can be considered as a hybrid between two limiting cases, an all-conserved model as the limit of high sequence identity and a polyserine model as the limit of low sequence identity. Precisely because the method of generating the structural model is so simple, the success of the structural model as a key to obtain a molecular replacement solution is even more dependent of a proper alignment.

Even with a structurally highly conserved central PA14 domain, it is possible for such structures to vary strongly within the loop regions, as was the case between Flo5A and the *B. anthracis* PA14 domain (Fig. 5.6). Therefore, after the Epa1A chainsaw model had been generated, several steps of manual retouching were performed to reduce the structure to its essential core.

#### **5.5.1.4 Building models for selected homologues.**

Once the final candidates had been chosen with the help of the phylogenetic tree, models of the Epa variants were generated with the program modeller[207], which uses comparative protein structure modeling by satisfaction of spatial restraints[187] to produce models based on sequence alignment of the desired protein against an experimentally obtained template model structure. For this, the putative A domain regions of the Epa variants, plus a 30 amino acid long C-terminal margin, were employed as target sequences. Modeller was used with the free modeller GUI, programmed by Kuntal Bushan[208], which simplifies the modeling process to 3 steps. First, an internal pairwise alignment is generated by methods similar to the CLUSTALW algorithm, with the difference that the gap penalty scores are not only dependent on statistical considerations, but on empirical data as represented by the template structure itself[207]. The resulting alignment was inspected for all alignments, and at no time could differences be seen between this alignment and the one performed by CLUSTALW.

Subsequently, modeller extracts spatial restraints from both an internal database on protein features and from the template structure as aligned on the target sequence (Table 3.21 gives an exhaustive list on spatial restraints used by modeller). From the spatial restraints, basic probability

density functions are calculated (pdf). A pdf is a function that gives the range and value of probability for the different states a system can adopt. For example, a peptide bond pdf will show one very large peak at the trans state for almost all amino acids, while proline will have two peaks, one in the cis and one in the trans state. As all probability functions, a pdf cannot adopt negative values, and, when integrated from  $-\infty$  to  $+\infty$ , it will always be 1. The basic pdfs are then combined to describe features. A feature is defined as a certain aspect within a structure that affects more than one atom. For example, the angle formed by three atoms, or the distance between two given atoms. A feature can be described as a weighted multiplication of several basic pdfs. Finally, a molecular pdf is defined as the multiplication of all feature pdfs[206].

Finally, the molecular pdf  $P$  is converted to the target function  $F$  (equation 5.12), which has the computational advantage of being less floating coma intensive and substituting multiplications for additions, making cpu overflow less likely[187]. The entire molecular pdf, and therefore target function, is not generated immediately, but iteratively. After each round of minimization, new feature pdfs are added to the  $P$  function. The addition of new features is governed by a growing amino acid window, which means that the sequence is read N to C terminally and incrementally minimized in this direction. Size of the window, and size of initial standard deviations for the basic pdfs can change strongly the initial conditions for folding. It is therefore recommended to generate an ensemble of models for difficult candidates.

$$F = -LnP \quad \text{eq. 5.12}$$

*F: target function*

*P: molecular probability density function (pdf)*

The program then tries to minimize the value of  $F$  iteratively. It is to be noted that the value of  $F$  is not an absolute value, and can only be used to compare different models performed with the exact same alignment. To asses model quality, modeller calculates discrete optimized protein energy (DOPE) potentials[209] for each position of the sequence as an average of a 15 amino acid window. DOPE potentials are essentially a normalized score per amino acid which shows the degree of chemical instability of the amino acid in its current environment. It mimics true potential energy in having highly positive scores for unstable locations, which are therefore poorly modeled, while showing negative scores for amino acids with stable states.

Models were then aligned with the template Epa1A structure using Pymol's align command[210], in

order to observe the relative position of the ligand-binding motif and the actual binding pocket and decide the mutagenesis strategy.

## 5.5.2 Structure elucidation methods

### 5.5.2.1 Data reduction

All steps in crystallographic data reduction were performed with the `xds_par` and `xscale` programs of the XDS program package[211], and the `SCALA` program of the `ccp4` suite[212]. The first step after obtaining diffraction images from a well diffracting crystal is to integrate the reflexes. During this process, the intensity  $I_{hkl}$  for each  $hkl$  reflex is derived from the images, furthermore, XDS divides each image in resolution shells, and calculates the quotient of the mean intensities  $I_{hkl}$  and their standard deviation  $\sigma_{I_{hkl}}$  for each shell. The quotient represents the signal to noise ratio, and maximum resolutions were chosen for values of  $I_{hkl}/\sigma_{I_{hkl}}$  of 2 or higher.

After the data had been integrated, it needed to be scaled with the program `xscale`. During data scaling, intensities are normed to a common scale so that systematic errors and uncertainties like radiation damage or absorption effects are accounted for. Evaluation of the integration and scaling process took place through the parameters of *completeness*,  $R_{merge}$ ,  $R_{meas}$  and  $I_{hkl}/\sigma_{I_{hkl}}$ .

Completeness is defined as the percentage of the expected unique reflections that are found within the dataset.  $R_{merge}$  is a quality parameter that measures the divergence of all symmetrically related reflexes, it is given by equation 5.13 and is normally expressed in a percentage.

$$R_{merge} = \frac{\sum_{hkl} \sum_i |I_i - \langle I \rangle|}{\sum_{hkl} \sum_i \langle I \rangle} \quad eq. 5.13$$

$I_i$ : intensity for reflection  $hkl$  at measurement  $i$ ,  
 $\langle I \rangle$ : mean intensity for reflection  $hkl$

$R_{meas}$  is another quality factor which is multiplicity independent, and therefore does not increase with it, as  $R_{merge}$  does [213].



$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{N}{N-1}} \sum_i |I_i - \langle I \rangle|}{\sum_{hkl} \sum | \langle I \rangle |} \quad \text{eq. 5.14}$$

*N*: multiplicity = total number of reflexes/unique number of reflexes

The absolute values of these parameters were used to evaluate the quality of the dataset, and the qualitative weighing of all parameters against each other lead to decide the maximum resolution at which the dataset would be further processed.

Unmerged data in .hkl format were loaded into the ccp4 package by converting them with the combat program into an .mtz file. In the case of Epa1A, the SCALA[212] task was used to merge the dataset and to set the FreeR flag, for use during phase solution and data refinement (see section 5.5.2.3), to ~1000 reflections. For the variants, the FreeR flag was imported using the import task from the Epa1A model structure during merging with SCALA.

### 5.5.2.2 Phase solution via molecular replacement

The solution of the phase problem is the central task in protein crystallography. Due to the nature of the data collection, only amplitudes can be directly integrated from the images, while the initial phases need to be determined by indirect methods and then refined.

In principle, there are four methods to solve the phase problem[214], of which only molecular replacement was needed to solve the structures of the Epa A domain. The molecular replacement method is based on the use of either a previously known structure or a model of the protein of interest based on an experimentally determined structure, which is then processed through reverse Fourier transform to determine initial phases, which are subsequently refined.

Phase solution by molecular replacement is a six-dimensional problem. The orientation of the searched molecule along the unit cell depends on both rotation along 3 angles ( $\alpha$ ,  $\beta$  and  $\gamma$ ) and translation along three axes (x, y and z) (Fig 5.9). Traditionally, cross-Patterson function based methods have been widely employed for phase solving[215], but recent developments in maximum likelihood probability algorithms have delivered better results[216]. The underlying assumption in maximum likelihood inferences is that '*the best model on the evidence of the data is the one that explains what has in fact been observed with the highest probability*'[217]. For crystallography, the

data are the individual reflection intensities, while the model is the calculated output including orientation, packing and the corresponding phases. Although not strictly true, it has been shown that the individual reflection intensities can be treated as independent variables without large errors being introduced into the system.[216]

The program Phaser[216] of the ccp4 package[212] is capable of producing maximum likelihood based rotation and translation functions. In these, the Logarithmic Likelihood Gain (LLG), the Z-score and the Clash Test are the fundamental quality parameters based on which the output solutions are evaluated. The LLG score is defined as the difference between the likelihood of the model and the likelihood calculated from a Wilson distribution, which is a random distribution of the atoms composing the model. Therefore, LLG measures how much better the model is as compared with random atoms. The Z-score is defined by equation 5.15, and it gives the number of standard deviations above the mean for a particular LLG score. A Z-score above 5 usually indicates a correct model, although a value as low as 4 can still indicate a correct solution.

$$Z = \frac{LLG_o - LLG_{ro}}{RMS_{LLGr_s}} \quad \text{eq. 5.15}$$

*Z*: Z-score,  $LLG_o$ : LLG for the current orientation

$LLG_{ro}$ : mean LLG for a random set of orientations

$RMS_{LLGr_s}$ : standard deviation for a random sample of LLG values.

During the clash test, accepted models are tested for molecular overlap of the C $\alpha$  atoms within the unit cell; some overlap is acceptable for initial phases, and in this work the threshold was set at a maximum of 10 clashes. A high number of overlaps tends to indicate either a false positive model, or that the wrong space-group has been used. A clash is defined as any symmetry-related molecule or components of the asymmetric unit that are found within less than 3 Å of any C $\alpha$  of a given model.

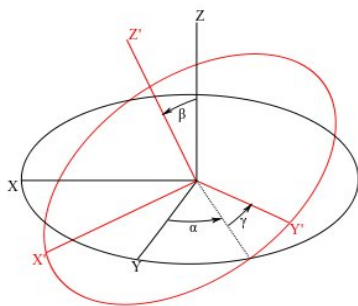


Figure 5.9: Rotation parameters during molecular replacement phase solution[214].

The Epa1A dataset was solved using the phaser program and a model of the Flo5 A domain (kindly provided by Maik Veelders) processed with

Chainsaw [186] and coot [218] (section 5.5.1). Maximum resolution was set to 3 Å and space group to C222<sub>1</sub>. The mutants presented >99% identity with the solved Epa1A, they were all processed directly using refmac5[219] and the Epa1A model stripped of water and ligands (see section 5.5.2.3 for more information on both programs). Relevant mutations were generated with coot[218] after the first refinement cycle, so as to see that the electron density showed the mutated amino acids without them having been introduced. The Epa1Ato3 structure could only be solved through phaser, as, although the unit cell seemed to be the same, the orientation of the protein model within was not.

### 5.5.2.3 Model building and data refinement

During the model building process, the amino-acid sequence of the target protein is modeled into the calculated electron density. Models for high resolution datasets can be quickly built by automated algorithms. Model building was accomplished automatically for the Epa1A domain with the program ARP/wARP and its integrated warpNtrace[220] function. Before running ARP/wARP, Model building was not necessary for the Epa1A variants, as the structures had >99% sequence identity with the Epa1A model. The sequence still applied except for the mutations, which were generated later with coot[218].

Models were then refined using a combination of automated maximum likelihood methods and manual refinement. For the automated methods, refmac5[219][212] was employed, while for the manual steps the program coot was used.[218] During automated refinement, the R and R<sub>free</sub> (eq. 5.16 and 3.17) factors were minimized, while the root mean square deviations (RMSD) for the bond length and angle were kept at all times below 0.01Å and 1.5°, respectively. Both R and R<sub>free</sub> use the same equation, but while R compares all the calculated data against all the observed data, R<sub>free</sub> compares the subset of the calculated structure factors corresponding to the FreeR flag, against the actual observed reflections marked by the FreeR flag, which are left out by refmac5 and phaser during model refinement and phase solution. Therefore, the R<sub>free</sub> factor is an internal quality parameter impervious to model bias, since it compares calculated reflections against the observed reflections, which have not been used for the calculation of the model.

$$R = \frac{\sum \| |F_{obs}| - |F_{calc}| \|}{\sum |F_{obs}|} \quad \text{eq. 5.16}$$

$$R_{free} = \frac{\sum \| |F_{obs, free}| - |F_{calc}| \|}{\sum |F_{obs, free}|} \quad eq. 5.17$$

$F_{obs}$ : observed structure factors       $F_{calc}$ : calculated structure factors  
 $F_{obs, free}$ : observed structure factors marked with the FreeR flag

The automated refinement of Epa1A was started by loading the scaled, integrated and merged dataset with FreeR flags and the best output model of the ARP/wARP program into the ccp4i refmac5 interface. Parameters of the GUI were set to perform a restrained refinement with no prior phase information, isotropic B factors and 10 maximum likelihood refinement cycles.

After automated refinement, the output electron densities and structure models were loaded into coot, where the standard settings were 1-2 $\sigma$  for the electron density and 3.5 $\sigma$  for the difference electron density. The main chain of the protein model underwent both real space refinement and regularization. Both waters and ligands were added and saved.

The saved output from manual refinement was loaded again for a second round of automated refinement cycles with refmac5. The process was repeated until the R and R<sub>free</sub> factors could not be lowered any further. Different strategies were employed during advanced automated refinement, like TLS refinement[221], rigid body refinement, Babinet scaling[222], adding computed hydrogens, and changing the geometric weighting factor[219]. These strategies were only pursued if, after the refinement job, the R<sub>free</sub> had been reduced by at least 1 point. TLS groups for TLS refinement were calculated using the TLS Motion Determination (TLSMD) server [223][224], which produces different ensembles of TLS groups that may be selected.

### 5.5.3 Protein properties calculation methods

#### 5.5.3.1 Isoelectric point (pI), extinction coefficient ( $\epsilon$ ) and molecular weight (MW)

The program ProtParam[225] from the ExpASy server calculated the theoretical values of the pI,  $\epsilon$  and MW for a given protein sequence.

### 5.5.3.2 Unit conversion, melting curve fitting and deconvolution of spectra for CD measurements

A simple macro for the OpenOffice.org Calc converted ellipticity in mean molar ellipticity per residue using eq. 5.2. Data had to be converted from the Jasco proprietary format to ASCII .txt files using the in-built Jasco File Converter™.

Melting curves were fitted using the QtiPlot[226] and equation 5.18 as described in [227].

$$\theta_T = (\theta_i - \theta_f) \cdot \frac{e^{\frac{\Delta H}{1.987 \cdot T} \cdot (\frac{T}{T_m} - 1)}}{1 + e^{\frac{\Delta H}{1.987 \cdot T} \cdot (\frac{T}{T_m} - 1)}} + \theta_f \quad \text{eq. 5.18:}$$

$\theta_T$ : mean molar ellipticity per residue at temperature  $T$ .  $\theta_i$ : mean molar ellipticity per residue for initial temperature.  $\theta_f$ : mean molar ellipticity per residue at final temperature.

$\Delta H$ : initial enthalpy in calories.

$T_m$ : melting temperature, temperature for which  $\theta_T = 0.5\theta_f$ .  $T$ : temperature, in Kelvin.

$$\theta_T = (\theta_i^a - \theta_f^a) \cdot \left( u_1 \cdot \frac{e^{\frac{\Delta H_1}{1.987 \cdot T} \cdot (\frac{T}{T_{m1}} - 1)}}{1 + e^{\frac{\Delta H_1}{1.987 \cdot T} \cdot (\frac{T}{T_{m1}} - 1)}} + u_2 \cdot \frac{e^{\frac{\Delta H_2}{1.987 \cdot T} \cdot (\frac{T}{T_{m2}} - 1)}}{1 + e^{\frac{\Delta H_2}{1.987 \cdot T} \cdot (\frac{T}{T_{m2}} - 1)}} \right) + \theta_f^a \quad \text{eq. 5.19:}$$

$\theta_T$ : mean molar ellipticity per residue at temperature  $T$ .  $\theta_i^a$ : apparent mean molar ellipticity per residue for initial temperature.  $\theta_f^a$ : apparent mean molar ellipticity per residue at final temperature.

$\Delta H_1$ : initial enthalpy in calories for first transition.  $\Delta H_2$ : initial enthalpy in calories for second transition.  $u$ : initial fractional extinction coefficient for each transition.

$T_{m1}$ : melting temperature for first transition,  $T_{m2}$ : melting temperature for second transition.

Equation 5.18 is the simplest conceivable model for the fitting of a two state melting curve. One must assume that the heat capacity of the system does not change with temperature, and that no linear denaturation occurs before or after the sigmoidal one[227]. An adapted model for a three state melting curve is given with equation 5.19, but the assumptions made are the same ones. It is important to note that initial values for all constants are of extreme importance when fitting melting curves. The suggested values for initial enthalpies given by Greenfield in [227] have proven invaluable in the fitting of these complex systems.

Deconvolution of spectra was performed using the K2d server[228]. The K2d server employs Kohonen's self-organizing map (SOM)[229] neural network algorithm. Briefly, the K2d server emulates the growth of sensory neurons, where only the fastest converging neuron survives upon connection with the growth factor emitting region, while all other neurons enter apoptosis. In the case of K2d, 41 input parameters, corresponding to each whole-number wavelength between 200 and 240 nm, are confronted by a set of 100 “neurons”. Each neuron is represented by a random set of 41 values for each wavelength derived from a weighted combination of the training set, i.e. although the values are random, it is known what amount of  $\alpha$ -helix,  $\beta$ -sheet and random coil would produce such a spectrum. The neuron which most closely resembles the experimental spectrum “survives” while all others “die”. The process is iterated with new neurons whose values are near the surviving neuron from last cycle. The final surviving neuron in the last cycle is a good approximation of the average secondary structure quantities that would produce the spectrum.

The neural network produces then 2 quality parameters, the secondary structure proportions and a corresponding calculated spectrum as output. The first quality parameter is the square distance between the calculated spectrum and the experimental spectrum. It is a value which holds the same significance as the RMSD between a set of points and a corresponding function in a linear regression, for example. The square distance is calculated as the euclidean distance between two vectors. A square distance value of 0 would mean that the calculated and experimental values fully overlap, and are therefore identical. The second quality parameter is the estimated error, which is an estimation performed by comparing the actual calculated and experimental data with the calculated and experimental data employed during training of the neural network, for which real errors could be calculated, as secondary structure percentages were known. The estimated error is given in percentage, and the program has an in-built cutoff of 22.7% after which it determines that the calculated and experimental spectra are too different, and therefore the deconvolution failed.

### ***5.5.3.3 Imaging and exploring structural models***

Structural models were studied using pymol v1.2r1[210]. Estimations of secondary structure on the model were performed with the stride plug-in as distributed with the VMD package[230], while structural alignments were performed using the CEAlign plug-in [231]. All snapshots presented in this work were performed with pymol, unless otherwise stated.

#### **5.5.3.4 *Secondary structure predictions***

Secondary structure predictions were performed using the Jpred3 server[232]. The Jpred server uses the Jnet[233] neural network algorithm to predict the secondary structure of the query. As the first step, Jnet generates a multiple alignment using PSIBLAST[193], which is a similar algorithm to BLASTp as described in section 3.5.1.1, but less stringent, therefore allowing for more distant homologues to be identified. It then removes any gaps within the query sequence, and with it the corresponding portions of the subjects. Next, Jnet compares the resulting array with its trained database, based on 480 unique, cross-validated proteins and generates the secondary structure prediction.

## 6 Results

### 6.1 Epithelial Adhesins from *Candida glabrata* (Epa)

*C. glabrata* is, with *C. albicans*, the most common fungal pathogen in *Homo sapiens*. Even though most cases of candidiasis are mild and occur on a mucosal level, systemic infections in immunocompromised patients[44] can have up to 52% median mortality rates[234], inducting *C. glabrata* into the category of hospital pathogens with such other, mainly bacterial, pathogens as *Staphylococcus aureus*. The challenges faced when treating a systemic candidiasis are quite different than a multiresistant bacterial infection. The eukaryotic nature of the pathogen implies a much more similar cellular structure between host and pathogen, which in most cases in strong adverse reactions to drugs and difficult and long-winded therapies[235]. It is therefore of capital importance to develop alternative therapies based on differential characteristics between host and pathogen, which will improve both survival rate and living quality of the patient. One of the most interesting targets in *C. glabrata* is the adhesion mechanism, which allows it to colonize both the external mucosae and the internal epi- and endothelia during pathogenesis[234]. The adhesion system, as especially exemplified by the epithelial adhesin family (Epa), is completely different to the way the human organism interacts with its own carbohydrates, and is absolutely necessary for primary colonization. Even though this system has been shown to be necessary for infection, *in vitro* and biochemical characterization have yet to be undertaken, barring some semi-quantitative binding assays[41]. Especially structural knowledge of the glycan binding domain of members of the Epa family and their specificity would be of tremendous use in the development of new, alternative therapies against candidiasis.

For the study of the Epa1 adhesive domain (Epa1A), several constructs had been previously generated by producing fragments from the Epa1 full length gene from *C. glabrata* (gene ID 34809538)[155]. Out of these, the pET28a Epa1\_2 construct was the one adopted for this work. The gene product Epa1\_2 (Epa1A from now on) was 262 amino acids long, from which amino acids 32 to 259 correspond to the sequence fragment 40 to 267 of full length Epa1. The short N-terminal fragment in Epa1A contained a His-tag, which was used during protein purification. Epa1A has a molecular weight of 29.4 kDa and a predicted pI of 6.20.



### 6.1.1 Purification of the Epa1A domain

Epa1A could be readily purified from raw cell lysates following expression and using NiNTA-affinity columns. However subsequent gel filtrations yielded very poor results (Fig. 6.1 a))[155]. When purified in AM buffer, Epa1A tended to either aggregate in the size exclusion column, eluting near the exclusion volume (Fig. 6.1, Peak I), or interacted with the Superdex 200 material, i.e. eluting far behind the theoretical maximal column volume (Fig 6.1 Peak III). Only a small fraction would appear approximately at the calculated elution volume for the protein (Fig. 6.1 Peak II). It was from this fraction that preliminary crystals could be obtained[155]. As the Superdex 200 material is a polysaccharide matrix (Fig. 5.1), it was proposed that specific protein-carbohydrate interactions might be causing the separation problems observed for this column. Lactose and EDTA were both tested as binding inhibitors on an analytical scale with a pre-packed analytical gel filtration column (Superdex 200 150/10GL, GE Healthcare)(Fig.6.1 b)). Both lactose, at a concentration of 50 mM, and EDTA, at 10 mM, were effective inhibitors of the interactions with the column, and large quantities of the Peak II fraction could be obtained. The procedure was successfully upscaled to the preparative level, and high yields could be obtained (~7 mg purified

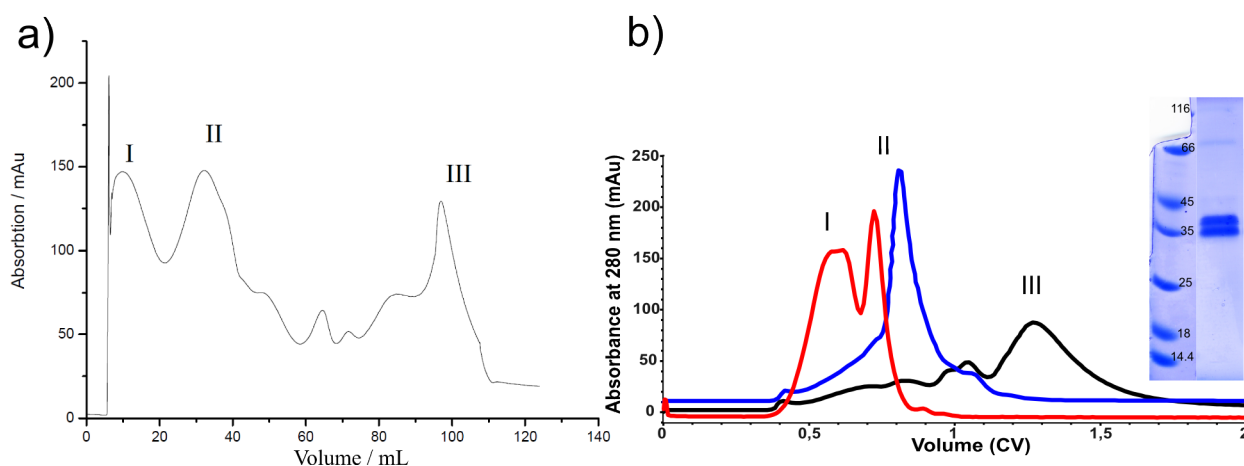


Figure 6.1: a) Sample size exclusion chromatography as run in [155]. Peaks I, II and III correspond all to Epa1A, but only peak II yielded preliminary crystals. b) Analytical gel filtration for Epa1A. Purification under standard conditions using AM buffer [155] in black. Purification with AML as running buffer in blue and with AME in red. The main peak under standard conditions is to be found after one column volume (CV) has run through the column and is analogous to peak III in a), while the main peaks with AML and AME are within 0.6 and 0.8 of the total column volume, and correlate well with peak II in a). SDS-PAGE shows pure protein after gel filtration. Although there appears to be a contamination or degradation of the protein in the form of a double band, it could be shown through mass spectrometry that both bands correspond to Epa1A and that it was not a product of limited proteolysis[155]. Molecular weight marker weights are given in kDa.

protein per liter of culture).

### 6.1.1.1 Crystallization of *Epa1A* under optimized conditions

Reproduction trials started with protein purified in AML at a concentration of 12 mg/ml. The classic and MBC screens were tested, but also optimizations based on the preliminary hits presented in [155] were planned and employed (section 5.4.2 and [183]). Finally, crystals could be obtained in the optimization screens (Table 5.16) although, interestingly, there was no crystal growth in the original conditions themselves, showing the importance of combining the sparse matrix strategy with a systematic screening of promising hits. The optimizations considerably increased crystal size and quality (Fig. 6.2).

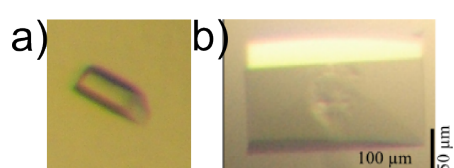


Figure 6.2 *Epa1A* crystals. a) Preliminary crystals produced under condition 92 of the Qiagen Classic screen (0.1 M MES pH 6.5, 0.2 M ammonium sulfate, 30% (w/v) PEG 5000 MME) plus 0.2 M lactose.. b) Optimized crystals based on condition 92.

### 6.1.2 Structure determination of the *Epa1A* domain.

*Epa1A* structure solution took place via molecular replacement using the Flo5A structure (PDBID: 2XJP)[1] as template for a chainsaw model, even though sequence identity was at ~20%. Fig. 6.3 shows the alignment used to generate a successful chainsaw model.

Table 6.1 gives an overview of the entire structure solution process and statistics.

Epa1	:: : : : : * *	* * : : : * * : : : * : : : * : : : * : : * * * : * * * : * : * * * : : * :	80
Flo5	D I S L A S K D P T T P L G C S P D I T T P K K G L S M E L Y S I D F R K K G S Y P C D A A M D P N F P R T G Y K S H R L L A K V D G V T G N I N F Y H	80	
ruler	1 ..... 10 ..... 20 ..... 30 ..... 40 ..... 50 ..... 60 ..... 70 ..... 80		
Epa1	..... * : : * * : * :	* * * * : * * * * * * * * * * * : * * * : * * * :	132
Flo5	A T K G C T P Q I G H L P A --- S V - N Y F - K P I --- --- T M T N F T M L L C Y F R P K V T G F H T F I - S A D D I	132	
ruler	..... 90 ..... 100 ..... 110 ..... 120 ..... 130 ..... 140 ..... 150 ..... 160		
Epa1	:: : * . . * * : * * : * : * : *	* : * * * : * * * * * : * * * * * : * * * * * : * * * : * * * : * * * : * * * :	209
Flo5	L V N E G A G N A D C C R E D S S A D H F G N L Q A V A I --- G S K T A N D E L T V H I D A G V Y P I R L F Y N N R E D G A L S F T E K T E S N E N	209	
ruler	..... 170 ..... 180 ..... 190 ..... 200 ..... 210 ..... 220 ..... 230 ..... 240		
Epa1	..... * * * * * : * : * * * * : * : * :	236	
Flo5	T V S D - F S E Y F F S L D D --- T E E G C P G L I S Y D S	236	
ruler	..... 250 ..... 260 ..... 270		

Figure 6.3: Alignment between Flo5 A domain sequence and Epa1 A domain sequence for the purpose of developing a Chainsaw model. Even though sequence identity is very low, key features such as the Flo5A subdomain, supposed to be absent in Epa1A, is predicted as a large gap in the alignment, and the CBL2 loop of Flo5A[1] corresponds with the specificity determining region of Epa1A[41].

### 6.1.2.1 Data collection, processing and phase solution.

The Epa1A crystal was measured at the ESRF (Grenoble, France) at beamline ID14-4 at a maximum resolution of 1.467 Å,  $\Delta\phi$  of 0.3° and a  $\phi$  range of 110°. Data was processed with the XDS suite[211], giving a final  $R_{\text{merge}}$  of 7.0%, completeness of 98.8% and  $I/\sigma$  of 14.6 for a resolution cutoff of 1.5 Å.

The Bravais lattice of the crystal was characterized as centered orthorhombic. Initially, it was assumed that the space group was C222 (number 21) but only datasets assigned to space group 20 (C222<sub>1</sub>) delivered significant molecular replacement solutions. Phase solution was achieved with the ccp4 program Phaser[216] by searching the dataset up to a resolution of 3 Å with a manually trimmed chainsaw[212] model (Fig 6.4) based on the weakly homologous Flo5A structure[1] for one monomer. The content of one molecule per asymmetric unit was indicated by the Matthews coefficient of 2.31 Å<sup>3</sup>/Da and an estimated solvent content of 46.75%. The selected solution presented final Z-scores of 5.2 for rotation and 6.5 for translation, an LLG score of 41.45 and no clashes.



*Figure 6.4: Models of Epa1\_wt used for phase solution of the Epa1\_2wt dataset. Grey: final model used for structure solution. Green: pruned regions of the original automated chainsaw model. Region A (amino acids 35 – 48 and 259 – 270) contains the C and N-terminal extensions, while B (amino acids 114 – 130) corresponds to the region modeled on base of the Flo5 sub-domain. The DcisD motif of calcium-binding site is included for orientation.*

### 6.1.2.2 Model building and refinement

Rigid body refinement followed by 20 cycles of automated ARP/wARP model building yielded a model with 4 partial chains for which 209 amino acids of the total 262 of the Epa1A gene product had been assigned. The process improved  $R_{\text{free}}$  from initial 50% to a final value of 25%. The model correctness was estimated to be of 99.8% and the R factor after ARP/wARP had a value of 20.6%. During refinement, both the  $\text{Ca}^{2+}$  and carbohydrate could be found and built into the electron density (Fig 6.5 c)). As expected, Epa1A is an all- $\beta$  protein composed of a  $\beta$ -sandwich motif in which 13 out of 15  $\beta$ -strands are involved (Figures 6.5 and 6.6). Carbohydrate and calcium binding occurs mainly on the CBL1 and CBL2 loops, while lid loops L1, L2 and L3 act as an outer shell isolating the binding pocket from solvent (Figure 6.5 a) and b)). L3 also interacts directly with the sugar over the galactose moiety and Trp197 (Fig. 6.5 b)), while L1 and L2 are quite rigid by covalent binding over a disulfide bridge between C77 in L1 and C118 in L2 (Fig. 6.6).

Interestingly, lactose could be modeled within the carbohydrate electron density, but a disaccharide composed of galactose and glucose bound over a  $\beta 1 \rightarrow 3$  glycosidic bond, this was surprising, as the protein had been co-crystallized in 100 mM lactose which had a purity of 98% according to the manufacturer's data (*Fluka*, lot code: 397309).

The final model could be refined to an R factor of 16.7 % and an  $R_{\text{free}}$  of 19.1 %. The asymmetric unit contains a total of 226 out of the 262 amino acids of the Epa1A gene product (amino acids D43 to S267 of the Epa1 sequence), 244 water molecules, two glycerol molecules, one  $\text{Ca}^{2+}$  ion, a Gal  $\beta 1 \rightarrow 3$ Glc disaccharide and one MES molecule. Fig. 6.5 shows the general fold of the Epa1 A domain and a detailed view on its binding pocket.

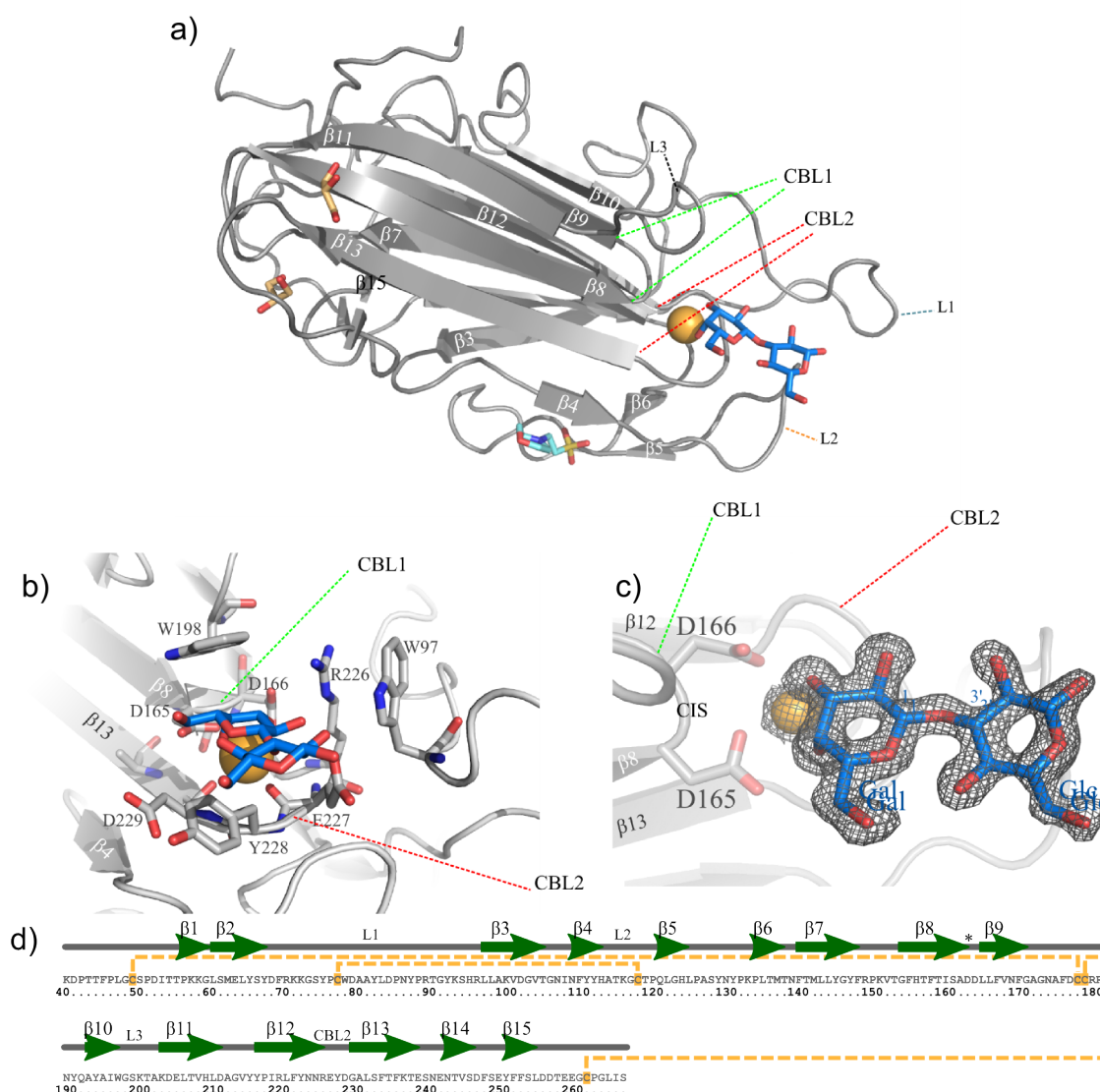


Figure 6.5: The Epa1A structure. a) General fold of the Epa1 A domain, showing the typical PA14  $\beta$ -sandwich fold[64] comprised by  $\beta$ -strands 2 to 4 and 7 to 13. Carbohydrate is portrayed in blue, MES in cyan, glycerol and  $\text{Ca}^{2+}$  in orange. Lid loops L1, L2 and L3 further isolate the binding pocket from the solvent, while loops CBL1 and CBL2 form the binding pocket itself. b) Detail on the binding pocket. The  $\text{Ca}^{2+}$  ion is coordinated by the peptide backbone of CBL2 and the D<sub>cis</sub>D motif [1] of CBL1 and Asn 227, while the carbohydrate interacts with the side-chains of the CBL2 amino acids, W198 and W97. c) Omit map showing the electron density at a contouring level of  $1.2\sigma$  and  $1.6 \text{ \AA}$  resolution of both the carbohydrate and the calcium ion. Only a  $\beta 1 \rightarrow 3$  glycosidic bond could be modeled in the electron density. The D-cis-D motive formed by D165 and D166 is shown for orientation. d) Secondary structure elements present in the Epa1A structure.  $\beta$  strands are marked in green from 1 to 15. Disulfide bridges between cysteins are marked in yellow dashed lines. The calcium binding loops 1 and 2 (CBL1, CBL2) are marked as an asterisk (\*) and CBL2, respectively, while the lid loops are marked as L1, L2 and L3.

Data collection		Refinement	
Space group	C222 <sub>1</sub>	Resolution (Å)	1.5
Cell parameters		Reflections (F>0)	41751
a (Å)	74.64	R (%)	16.7 (28.0)
b (Å)	104.33	R <sub>free</sub> (%)	19.1 (33.0)
c (Å)	69.79	Protein atoms	1955
Wavelength (Å)	0.933	Water molecules	244
Resolution (Å)	1.47– 19.89	Metal atoms	1
Completeness (%)	98.8 (98.7)	Average B-factor (Å <sup>2</sup> )	7.34
Total reflections	174640	RMSD	
Unique reflections	43249	Bond length (Å)	0.01
Multiplicity	4	Bond angle (°)	1.38
R <sub>merge</sub> /R <sub>meas</sub> (%)	7.0 (65.5)/ 7.9 (74.8)		
I/σ	14.6 (2.6)		
Phase solution			
Translation Z score	6.5		
Rotation Z score	5.2		
LLG	41.45		
Clashes	0		

Table 6.1: Overview of the statistics in data collection, phase solving and refinement of the Epa1A structure model.

### 6.1.3 Generation of Epa1A variants.

In order to explore possible changes in specificity within the Epa family, the amino acids responsible for the binding specificity in Epa1A had to be mutated. To define the key amino acids, the task was approached hierarchically. First, a BLAST[143] search was performed within the *C. glabrata* genome to find Epa1 homologues sharing at least 45% identity with the Epa1 A domain. Secondly, a multiple sequence alignment and a phylogenetic tree were generated with the chosen homologous A domains using Clustal X v2.0 [198] and COBALT[236] with identical results(Fig 6.6).

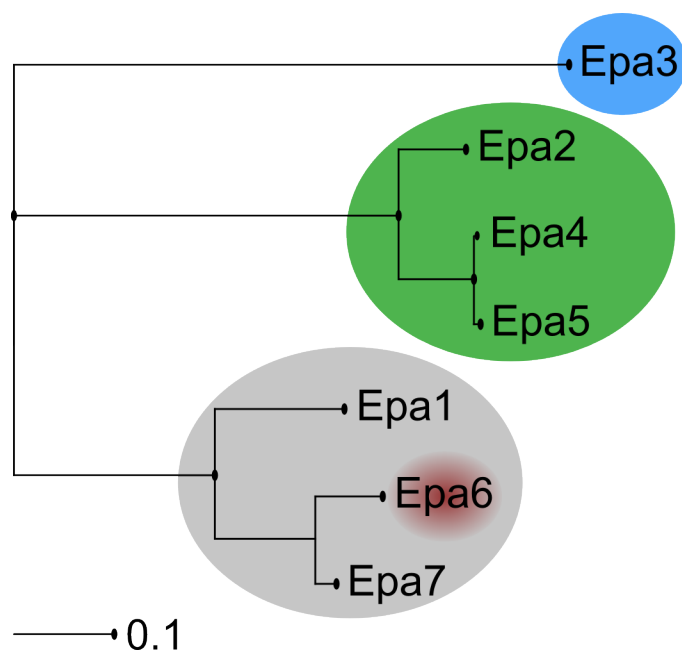


Figure 6.6 Phylogenetic tree analysis of the seven A domains from the Epa Family with at least 45% identity to Epa1. Three branches or clusters are clearly differentiated within the tree. Grey: Epa1A cluster. Green: Epa2A cluster. Blue: Epa3A cluster. Even though it is not recognized as a separate cluster, Epa6A (in red) shows a different CBL2 region. Therefore, it was considered to be a parallel subclass for variant selection. Evolutionary distance scale is given on the bottom left (section 5.5.1.2).

Next, the ligand-binding motif[41] was examined for each of the clusters and compared to the Epa1A binding pocket by homology modeling (Fig 6.7) as performed with modeller v9.7[207]. Finally, mutagenesis for conversion of Epa1A into the derived Epa subclasses was undertaken and confirmed through sequencing (Appendix II).



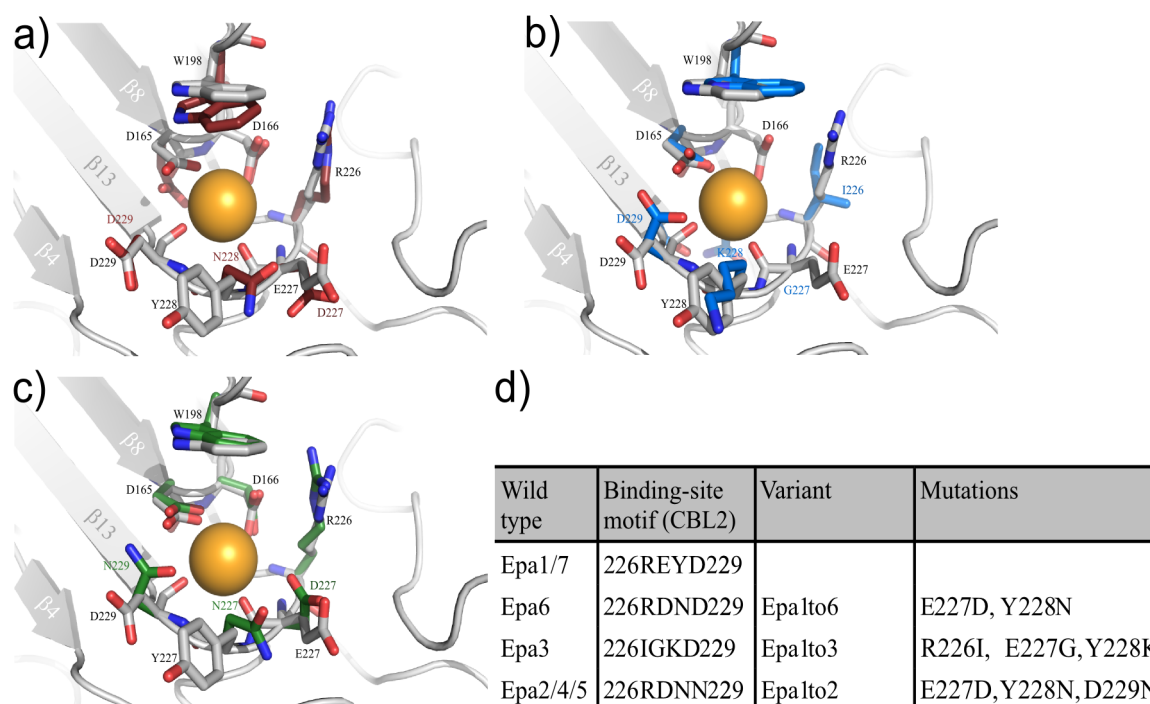


Figure 6.7: Binding pockets of the homology models of the members of the *Epa* family subclasses represented in Fig. 6.6. superimposed on the X-ray structure of *Epa1A*. Grey: *Epa1A* X-ray structure. a) red: *Epa6A* model, b) Blue: *Epa3A* model, c) Green: *Epa2A* model. d) Table showing the different binding-site motif sequences for each cluster and the undertaken mutations. Although *Epa1*, 6 and 7 pertained to the same cluster, *Epa6* showed a deviant binding-site motif for CBL2, and therefore was also taken into account for the mutagenesis strategy.

#### 6.1.4 Expression of *Epa* variants

Conditions around the optimal expression conditions for *Epa1A* [155] were tested as described in section 5.2.1 for all variants shown in Fig. 6.8 and optimized expression conditions could be promptly attained. Figure 6.8 shows an example of a solubility test for the *Epa1*→*6A* variant. Preparative expressions were performed and *E. coli* cells were harvested as described in section 5.2.1.

#### 6.1.5 Purification of *Epa* variants.

The *Epa1 A* domain variants were analogously purified using the optimized purification protocol for the *Epa1A* gene product. Interaction with the analytical Superdex 200 column varied among the different *Epa1A* variants (Fig. 6.9). Preparative purifications of the mutants in either AML or AME

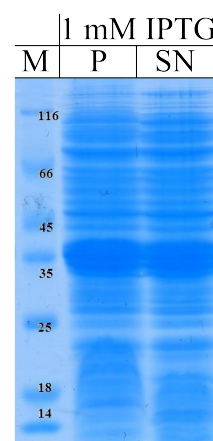


Figure 6.8: Example of solubility test. M: molecular weight marker, with molecular weights in kDa. P: Pellet. SN: Supernatant.



buffers yielded large quantities of pure protein (~10 mg per litre of culture).

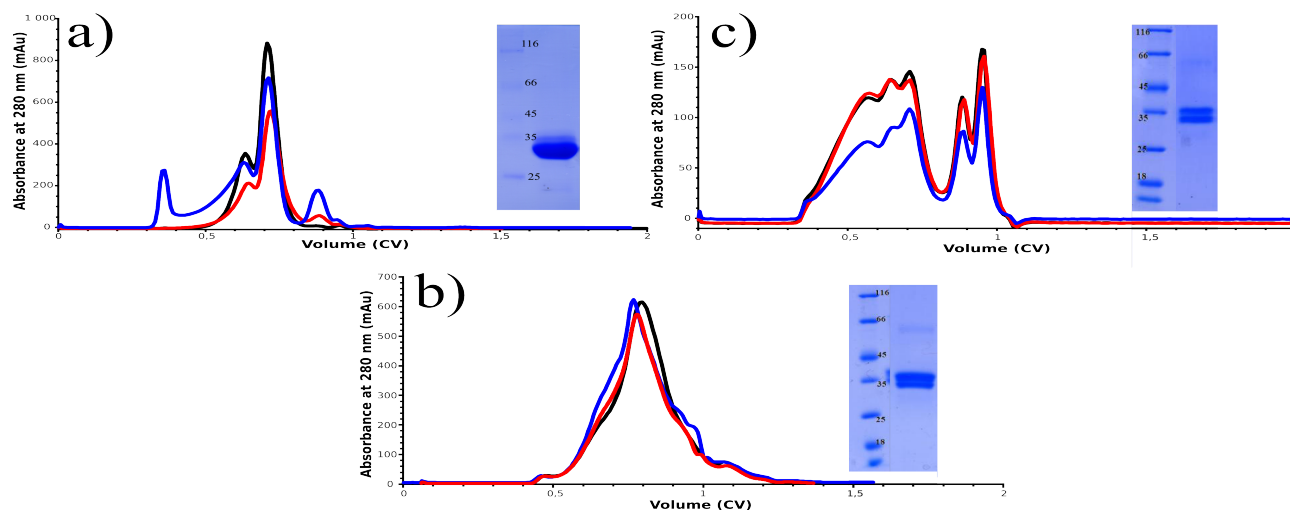


Figure 6.9: a) *Epa1*→3A b) *Epa1*→6A c) *Epa1*→2A. SDS-PAGEs show the corresponding samples prior to analytical gel filtration. Analytical gel filtrations of the *Epa1* A domain mutants. Black: gel filtration under standard conditions. Blue: gel filtration with AML buffer. Red: gel filtration with AME buffer. Although the different buffers do not seem to have any effect on *Epa1*→6A, it could be shown by fluorescence titrations that only the purification with AME buffer supplemented with calcium yielded active protein (section 4.7.1). All main peaks are between 0.6 and 0.8 CV, corresponding to the crystallizing fraction II of *Epa1*wt. Interestingly, *Epa1*→3A does show a different ratio in the typical double band that both *Epa1*A and all other variants present.

## 6.1.6 Structural analyses of subclass-converted *Epa1*A variants

### 6.1.6.1 Crystallization of the *Epa1*A variants

Crystallization conditions as described in sections 5.4.1 and 5.4.2 were set up for the *Epa1*A mutants at 18 and 4 °C in Innova plates (*Hampton Research*). Crystals grew at similar rates as the original *Epa1*A, with 4 to 7 days as an average. There was a general tendency to produce better diffracting crystals using the optimization of condition 92 of the classics screen, but no crystals could be produced in the original solution of condition 92. Crystals of *Epa1*A variants could only be grown in the presence of carbohydrate. Although it was possible to produce co-crystals of *Epa1*→6A and lactoNbiose I (Fig. 6.10), they diffracted poorly. Crystals grown by the sitting drop method grew out of a thick, transparent precipitate that hampered picking. Reproduction of target conditions in hanging drop format with a total volume of 2 µl generally solved this problem and increased crystal quality.

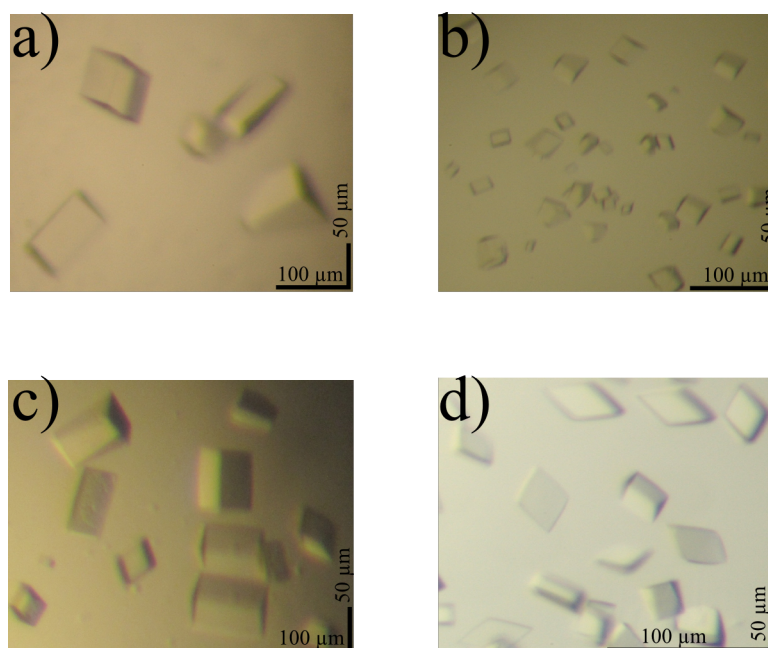


Figure 6.10: Diffracting crystals of the *Epa1* A domain mutant. a) Crystals from *Epa1*→3A·lactose grown in 0.1 M MES pH 6.5, 200 mM ammonium sulfate, 25 % (w/v) PEG 5000 MME. b) Crystals of *Epa1*→2A·lactose grown in 0.1 M MES pH 6, 200 mM ammonium sulfate, 22.5 % (w/v) PEG 5000 MME. c) Crystals of *Epa1*→6A·lactose grown in 0.1 M MES pH 6, 100 mM ammonium sulfate, 32.5 % (w/v) PEG 5000 MME. d) Crystals of *Epa1*→6A·lactoNbioseI grown in AME buffer + 10 mM lactoNbiose I + 20 mM CaCl<sub>2</sub>, 0.1 M MES pH 6.8, 100 mM ammonium sulfate, 22.5 % (w/v) PEG 5000 MME.

### 6.1.6.2 Data collection and improvement of diffraction quality of the variant crystals

Datasets could be collected for all 3 mutants, but the diffraction qualities of *Epa1*→3A·lactose and *Epa1*→6A·lactoNbioseI crystals were initially poor (8 – 10 Å highest resolution shell). At that point, crystals had been frozen in cryo-buffer composed of the original condition with 30 % (v/v) Glycerol added.

Since some of the crystals looked damaged after the freezing process,

Additive	30 % (v/v) Glycerol	30 % (w/v) PEG 400	30 % (v/v) MPD	1.5 M Ammonium Sulfate	1 M N-Acetylgalactosamine
<i>Epa1</i> to6 + LactoNbiose					
resolution	13.3 Å	no diffraction	10.7 Å	9.39 Å	
<i>Epa1</i> to3					
resolution	8.01 Å	3.26 Å	6.16 Å	1.8 Å	2.84 Å

Figure 6.11: Improving diffraction quality. Sample diffraction images from *Epa1*→6A lactoNbiose I co-crystals (upper row) and *Epa1*→3A crystals (lower row). Using Ammonium sulfate as a cryo-protectant dramatically improved *Epa1*→3A diffraction quality. Maximum resolutions were visually estimated using *imosflm*[185].

different cryo-conditions were tested (Fig 6.11). 1.5 M ammonium sulfate significantly improved the diffraction quality of Epa1→3A·lactose crystals, but no such effect could be achieved for the Epa1→6A·lactoNbiose crystals. Table 6.2 shows a summary of the collection strategies for the different datasets.

Variant	Range of $\varphi$ (deg)	$\Delta\varphi$ (deg)	$R_{\text{merge}}$ (%)	$R_{\text{meas}}$ (%)	$I/\sigma$ (average)	Completeness (%)	Space group	Maximum resolution (Å)
Epa1→3A·lactose	98	0.5	4.5	5.3	16.3	98.9	C222 <sub>1</sub>	1.63
Epa1→6A·lactose	120	0.2	4.1	4.6	26	99.6	C222 <sub>1</sub>	1.50
Epa1→2A·lactose	85	0.3	7.9	10.7	8.59	97.7	C222 <sub>1</sub>	2.00

*Table 6.2: Summary of the collection strategies for the Epa1 A domain variants. Epa1→3A·lactose and Epa1→6A·lactose datasets were collected at ESRF beamlines ID23-2 and ID14-2, respectively. Epa1→2A·lactose dataset was collected using an in-house X-ray diffractometer. Maximum resolutions were obtained as an output of Scala[212].*

### 6.1.6.3 Phase solution and refinement

Phase solution and refinement took place concurrently for the variants by molecular substitution, using the wild type Epa1A model structure without water molecules or ligands as starting point for refinement in refmac5[219]. In the case of Epa1→3A, indexing differences within the crystal seemed to prevent refmac5 to accept the model. Accordingly, molecular replacement was done with phaser using the Epa1A structure as template and gave  $Z$  scores of 27.2 and 40.1 for the rotation and translation functions, respectively. The mutated amino acids were rebuilt with coot[218], when preliminary electron densities had been obtained, along with ligands and water molecules. Although 50 mM lactose had been used for co-crystallization in all cases, Epa1→6A and Epa1→2A seem to have bound the  $\beta$ 1→3 disaccharide that was also found for the Epa1A structure. Even though there seems to be a definite preference for the  $\beta$ 1→3 disaccharide above lactose, the definition of the electron density for the glucose molecule varied widely a) amongst the variants (Fig. 6.12), and b) with the orientation of the pyranose ring of the glucose being either planar or perpendicular to the galactose (diagrams performed with Avogadro[237]). The main statistics and specifics of the structure solution for all Epa1A variants are summarized in table 6.3.

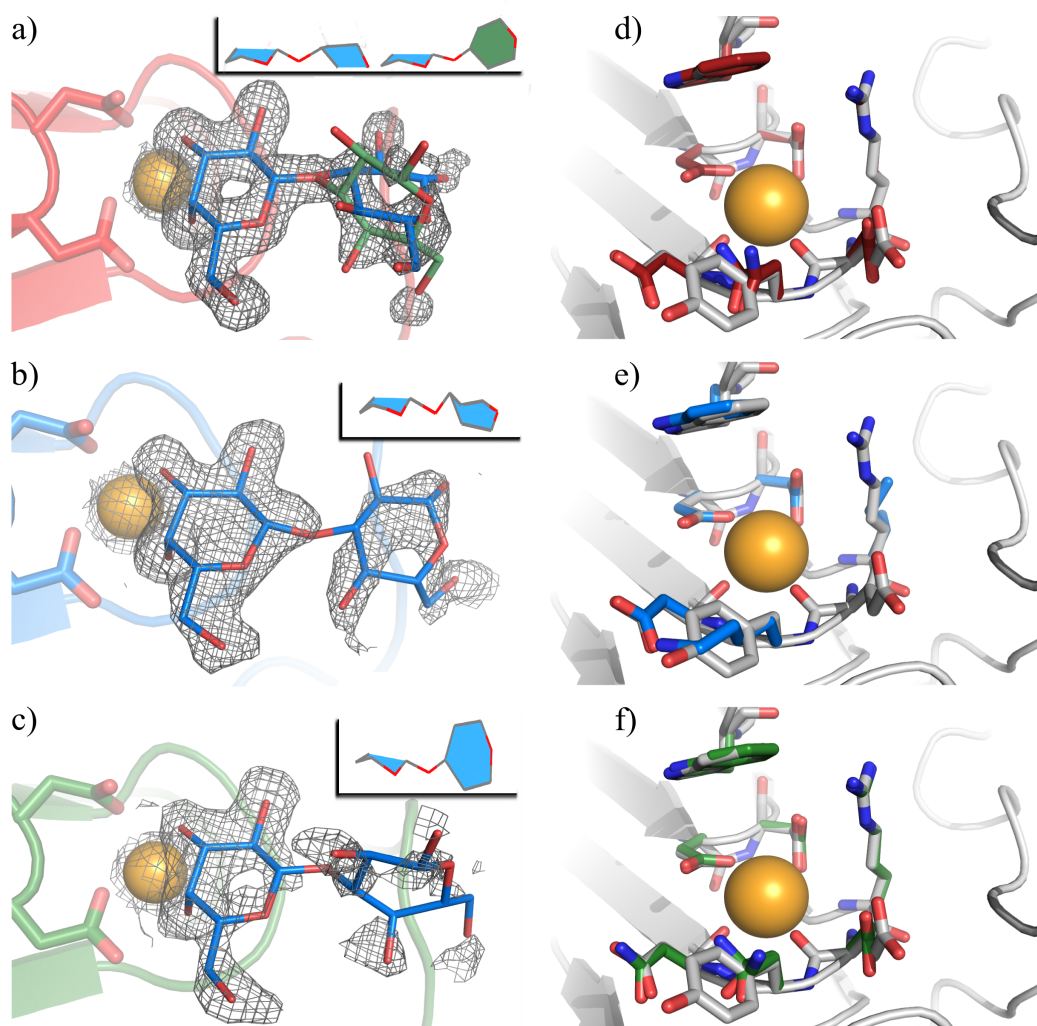


Figure 6.12: Structural details of the *Epa1 A* domain variants. Red, blue and green polypeptides stand for *Epa1*→6A, *Epa1*→3A and *Epa1*→2A, respectively. Calcium shown in orange, disaccharide in blue. a), b), c) top view of the electron density corresponding to the disaccharide for each of the mutants. The galactose maintains a constant orientation, while the glucose was rotated along the glycosydic bond in the different variants. In the case of *Epa1*→6A, shown in a), two conformations were available for the glucose, shown in blue and in dark green. Glucose orientation in comparison with galactose can be appreciated in the top right diagrams. d), e), f) detail of the binding pocket of the mutants (color) compared to the wild type structure (gray), all changes are concentrated on the CBL2 loop. Electron densities were calculated by *refmac 5* as omit maps and are presented at  $\sigma=0.8$  and *carve* = 1.6 Å. Occupancies were always 1.0 for the galactose except for *Epa1*→2A in c), where it was 0.7. Glucose occupancies were 0.3 (average B-factor: 24.62 Å<sup>2</sup>) for blue and 0.5 (average B-factor: 32.32 Å<sup>2</sup>) for green conformations in *Epa1*→6A in a). For the *Epa1*→3A the occupancy of glucose was 0.5 (average B-factor: 18.78 Å<sup>2</sup>) in b) and in the case of *Epa1*→2A it was 0.5 (average B-factor: 21.00 Å<sup>2</sup>) in c). Orientation diagrams calculated with *Avogadro*[237].

Data collection	Epa1→3A	Epa1→6A	Epa1→2A
Space group	C222 <sub>1</sub>	C222 <sub>1</sub>	C222 <sub>1</sub>
Cell parameters			
A (Å)	74.88	74.43	74.56
B (Å)	102.7	103.9	104
C (Å)	69.05	69.28	69.15
Wavelength (Å)	0.933	0.94	1.54
Resolution (Å)	1.63	1.55	2.00
Completeness (%)	98.9 (94.4)	99.6 (99.9)	97.7(99.7)
Total reflections	124967	186996	55079
Unique reflections	33171	39090	18024
Multiplicity	3.8	4.8	3.1
R <sub>merge</sub> /R <sub>meas</sub> (%)	4.5 (59.5)/ 5.3(70.4)	3.8 (47.4)/4.3(52.9)	12.3(51.9)/14.9(63.9)
I/σ	16.3(2.0)	26(3.5)	7.4(2.2)
Phase solution			
Translation Z score	27.2	n.d.	n.d.
Rotation Z score	40.1	n.d.	n.d.
LLG	2174	n.d.	n.d.
Clashes	0	n.d.	n.d.
Refinement			
Resolution (Å)	1.63	1.55	2
Reflections (F>0)	31967	37731	17402
R (%)	19.49	15.28 (24)	22.05(26)
R <sub>free</sub> (%)	20.8	17.52 (29)	25.17(27)
Protein atoms	1804	1924	1896
Water molecules	117	195	138
Metal atoms	1	1	1
Average B-factor (Å <sup>2</sup> )	11.2	11.6	14.4
RMSD			
Bond length (Å)	0.01	0.01	0.01
Bond angle (°)	1.37	1.37	1.34

Table 6.3: Overview of the data collection and refinement statistics for the Epa1 A domain mutants.

### 6.1.7 Carbohydrate binding of Epa A domains

Carbohydrate specificity and affinity of Epa1A for and variants was quantitatively characterized by fluorescence titration and semi-quantitatively by high-throughput binding assays to glycan arrays. These arrays were provided by the consortium for functional glycomics (CFG, [www.functionalglycomics.org](http://www.functionalglycomics.org)) at La Jolla, Ca, USA.

#### 6.1.7.1 High-throughput semi-quantitative binding assays of Epa1 and mutants.

The CFG glycan arrays provide a very powerful tool to determine the preferred ligands for all manner of glycan binding proteins (GBPs). By testing 465 different immobilized carbohydrates in its last iteration (glycan chip V4.1), it is possible to obtain very fast, semi-quantitative information on GBP binding preferences. In the case of Epa1A and variants, the information provided could be partially compared with previous experiments[41], in which some of the Epa family members had been tested for binding against CFG glycan arrays. The study of Epa1A and its variants could expand the information offered there and clarify whether carbohydrate binding was only affected by short range interactions within the binding pocket, or whether long range effects on other regions of the protein may play an important role in ligand recognition.

In order to characterize the binding specificity of Epa1A, binding experiments were performed at three different concentrations, 1  $\mu\text{g/ml}$ , 50  $\mu\text{g/ml}$  and 200  $\mu\text{g/ml}$  (Fig. 6.13 a, 6.13 b and 6.14 a, respectively). Binding to the 4.1 CFG printed glycan array chip showed similar binding profiles for Epa1A, and Epa1 $\rightarrow$ 6A (Fig. 6.14 a, b). Epa1 $\rightarrow$ 2A and Epa1 $\rightarrow$ 3A showed also similar profiles, but Epa1 $\rightarrow$ 3A was much less active (Fig. 6.15 a and b).

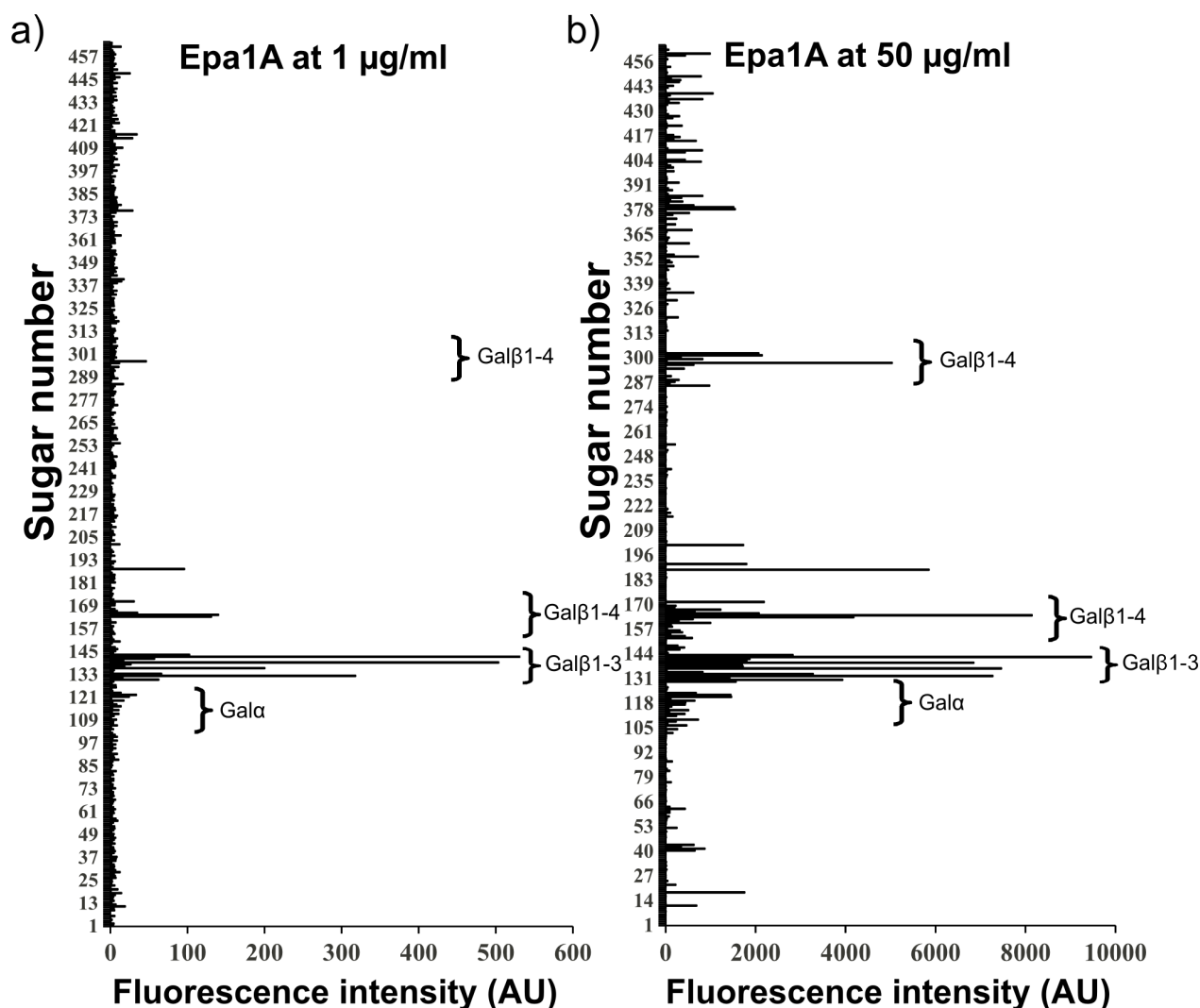


Figure 6.13: Binding of epithelial adhesin *Epa1A* to CFG glycan array V4.1 at different protein concentrations. a) Binding of *Epa1A* at 1 µg/ml. b) Binding of *Epa1A* at 50 µg/ml. Each carbohydrate has a unique identification number (carbohydrate number) which can be consulted in Appendix I Fluorescence intensities are given in arbitrary units. As binding conditions become more stringent due to limiting amounts of protein, the carbohydrates with the highest affinity are the ones sequestering most of it. As a result, at 1 µg/ml binding occurs almost only at the galactoseβ1-3 oligosaccharide containing region. Secondary specificities can be observed at higher concentrations, as is apparent at 50 µg/ml, where some binding can be seen at the galactose α region and especially at the Galβ1-4 regions.

A clear preference for mucin-type O-glycan core motifs could be attested for *Epa1A* (Figures 6.14 a) and b) and 6.16 a)), especially the core 1 Galβ1→3GalNAc motif or T-antigen[11][25]. This preference was also present for *Epa1*→6A, but the specificity was somewhat reduced, with many other glycans within the range of best binders (Fig 6.16 b)). In the case of *Epa1*→2A a preference for sulfogalactose containing carbohydrates is predominant (Fig. 6.16 c)).

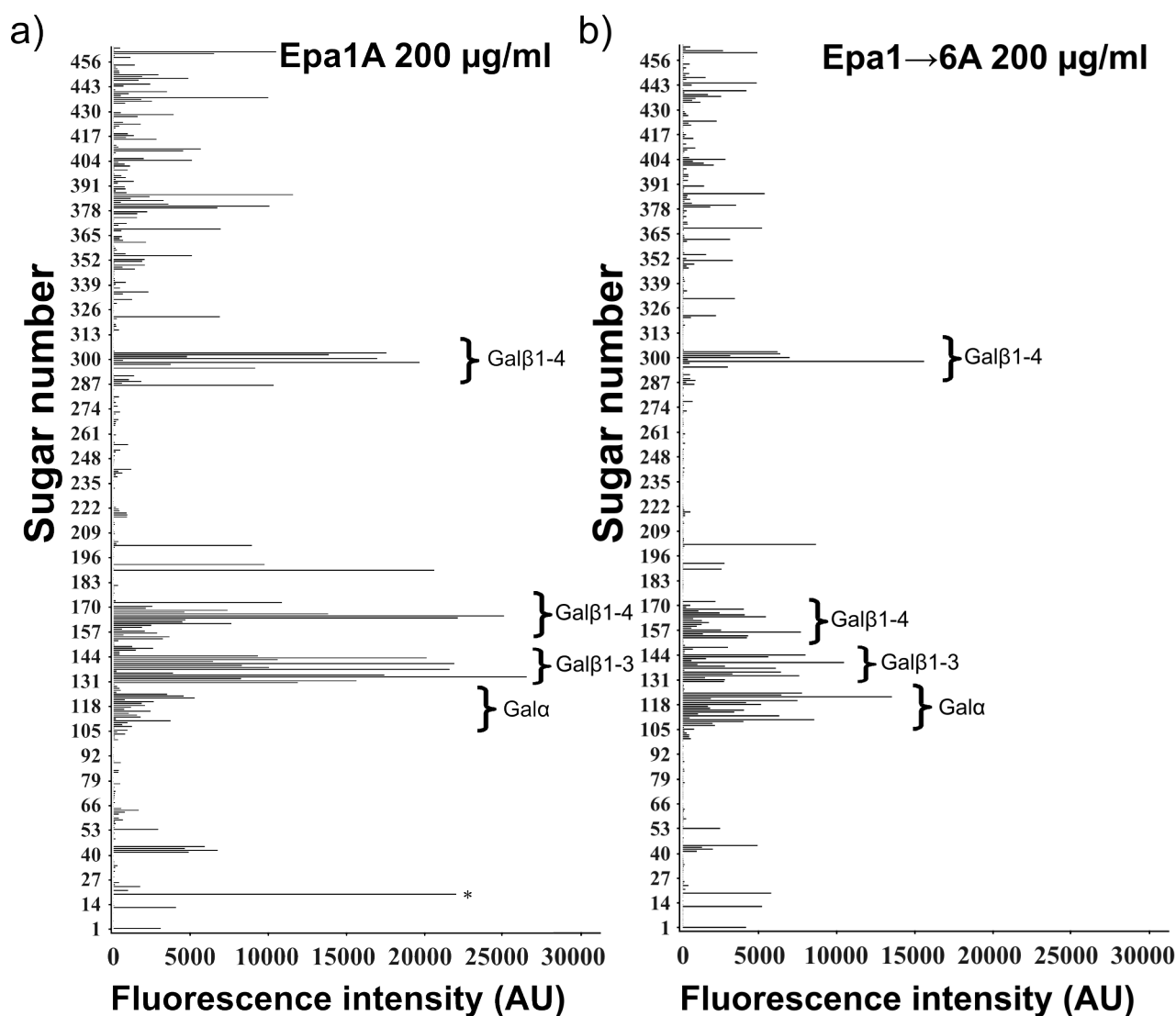


Figure 6.14: Binding of Epithelial adhesin to CFG glycan array V4.1 at a protein concentration of 200 µg/ml. a) Binding of Epa1A. b) Binding of Epa1→6A. Each carbohydrate has a unique identification number (carbohydrate number) which can be consulted in Appendix I Fluorescence intensities are given in arbitrary units. Both proteins show a clear preference for oligosaccharides containing non-reducing galactose. Epa1wt has a higher specificity for β1-3 bound galactose with a secondary specificity for β1-4 carbohydrates. In the case of Epa1→6A, no clear specificity between α1-3, α1-4, β1-3 and β1-4 can be differentiated, with the highest peaks evenly distributed amongst them, and a clearly lower general binding strength than Epa1A. \* marks an especially prominent differential peak in Epa1A, which corresponds to a branched core 3 O-GalNac oligosaccharide (Fig. 6.16).



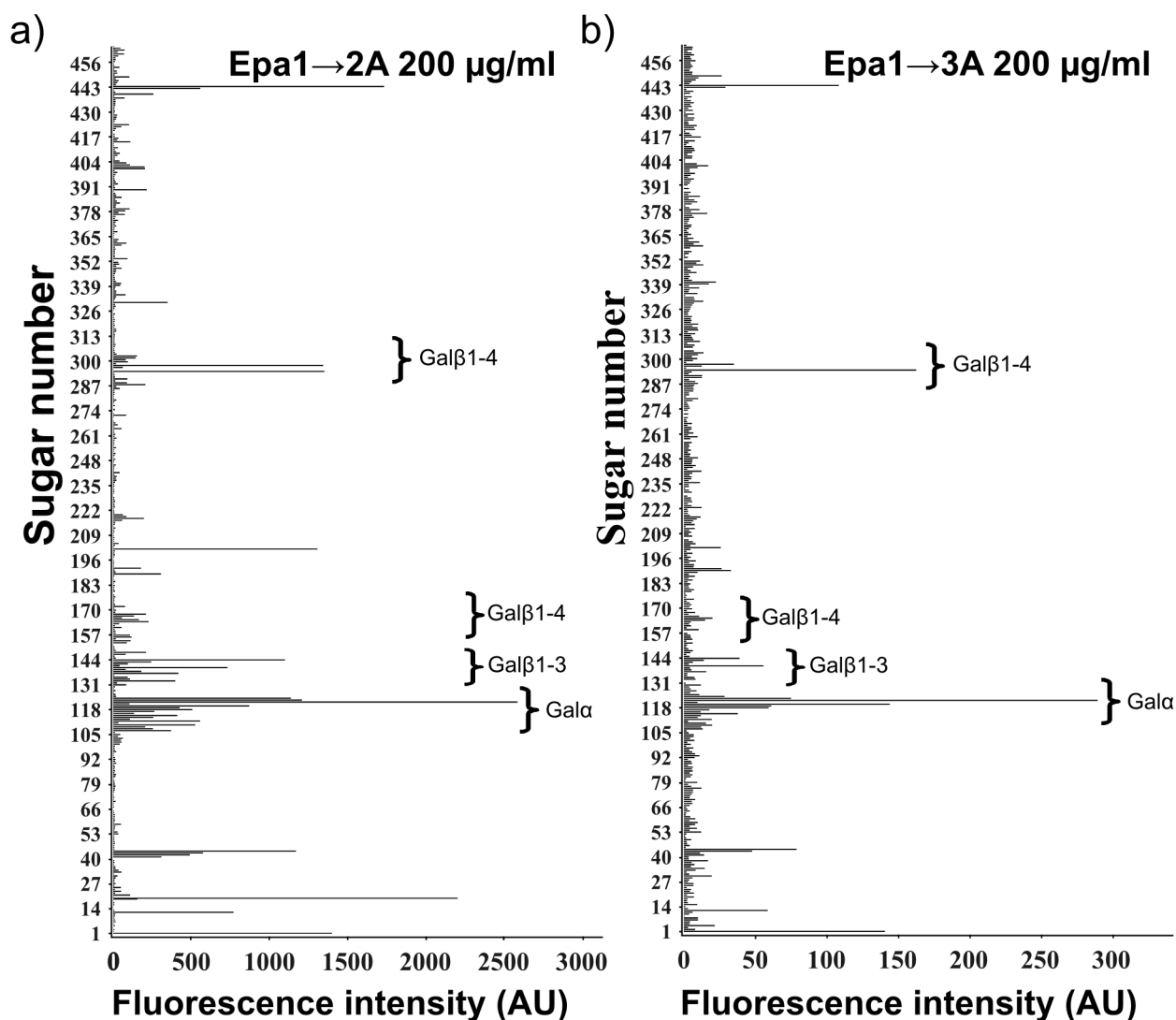


Figure 6.15: Binding of Epithelial adhesin to CFG glycan array V4.1 at a protein concentration of 200 µg/ml. a) Binding of Epa1→2A. b) Binding of Epa1→3A. Each carbohydrate has a unique identification number (carbohydrate number) which can be consulted in Appendix I. Fluorescence intensities are given in arbitrary units. Epa1→2A shows relatively weak binding, with a ten fold decrease in maximum fluorescence intensity compared to Epa1A. It does present a different specificity to Epa1A and Epa1→6A, with a preference for  $\alpha$ -Galactose containing carbohydrates and a secondary preference for Gal $\beta$ 1-3 and Gal $\beta$ 1-4 oligosaccharides. Epa1→3A is an even weaker binder, with 100 times less maximal binding intensity than Epa1A, and a clear preference for  $\alpha$ -Galactose containing carbohydrates and the only Gal $\beta$ 1-4 oligosaccharide 295, [6OSO3]Gal $\beta$ 1-4[6OSO3]GlcNAc.

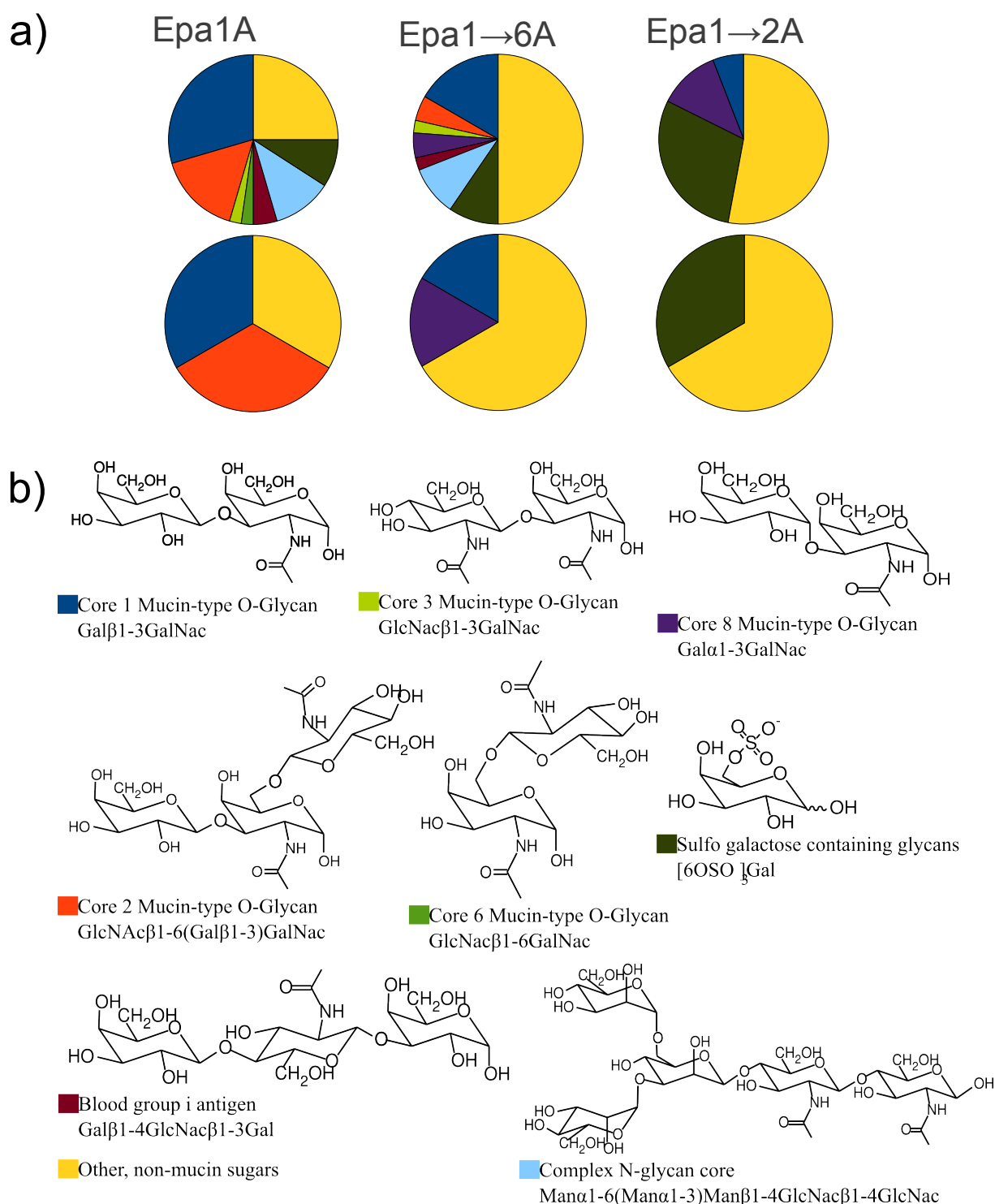


Figure 6.16: Distribution of carbohydrate ligands by type. a) Pie charts showing the distribution of ligands with fluorescence intensities up to 20% of maximum (top) or showing the distribution of ligands up to 50% of maximum (bottom). b) Structural formulas of the top binding carbohydrates, giving color codes for the pie charts. Carbohydrates present on mucins are accounted for individually, all other carbohydrates are collectively referred as “other”. The proportion of mucin related glycans decreases in both *Epa1*to6 and *Epa1*to2, while the proportion of other carbohydrates increases, showing an increasing promiscuity in the mutants. In the case of *Epa1A*, ~75% of all ligands with at least 20% of maximum fluorescence intensity are carbohydrates present on the surface of mucins, which correlates well with the ecological niche of *C. glabrata*, i.e. the colonization of mucous membranes. For numerical values consult table 6.4


















	Epa1wt		Epa1to6		Epa1to2,4,5	
	Up to 20%	Up to 50%	Up to 20%	Up to 50%	Up to 20%	Up to 50%
	13	5	7	1	1	
	7	5	2			
	1		1			
	1					
			2	1	2	
	2		1			
	5		4			
	4		4		5	2
	11	5	21	4	9	4
Total	44	15	42	6	17	6

Table 6.4: Distribution of carbohydrate ligands per type. Numerical values.

 *O*-glycan, core 1;  *O*-glycan, core 2;  *O*-glycan, core 3;  *O*-glycan, core 8;  Blood group I containing oligosaccharides;  Complex *N*-glycans;  Sulfo-Galactose containing carbohydrates;  Other carbohydrates.

For each of the proteins, the left column represents the distribution of ligands that bind with at least 20% of the fluorescence intensity of the top binder for that protein. On the right, ligands with at least 50% of the fluorescence intensity of the top binder are accounted for.

In order to further characterize the binding specificities of Epa1A and its variants, a closer look was taken at the binding data for the mono- and disaccharides present within the CFG glycan array (Figures 6.17 – 19). Monosaccharide affinity is informative about the binding specificity of the calcium – CBL1 – W197 sub-pocket (Fig. 6.4), while disaccharide binding can show the binding preference of the entire pocket in a much more defined way than the entire array. On the other hand, the array delivers a much more thorough profile and sets the glycan binding protein in a more biologically relevant context, as most oligosaccharides found within human tissues are larger than disaccharides, as figure 6.16 b shows. Another important point about the disaccharide representation is that both during the fluorescence titrations presented in this work and in previous ones [155] lactose (Galactose- $\beta$ 1-4Glucose) was chosen as a putative ligand. The disaccharide array shows very clearly that lactose is bound with very low affinity when compared to other sugars, and therefore cannot be a primary ligand of Epa1A.

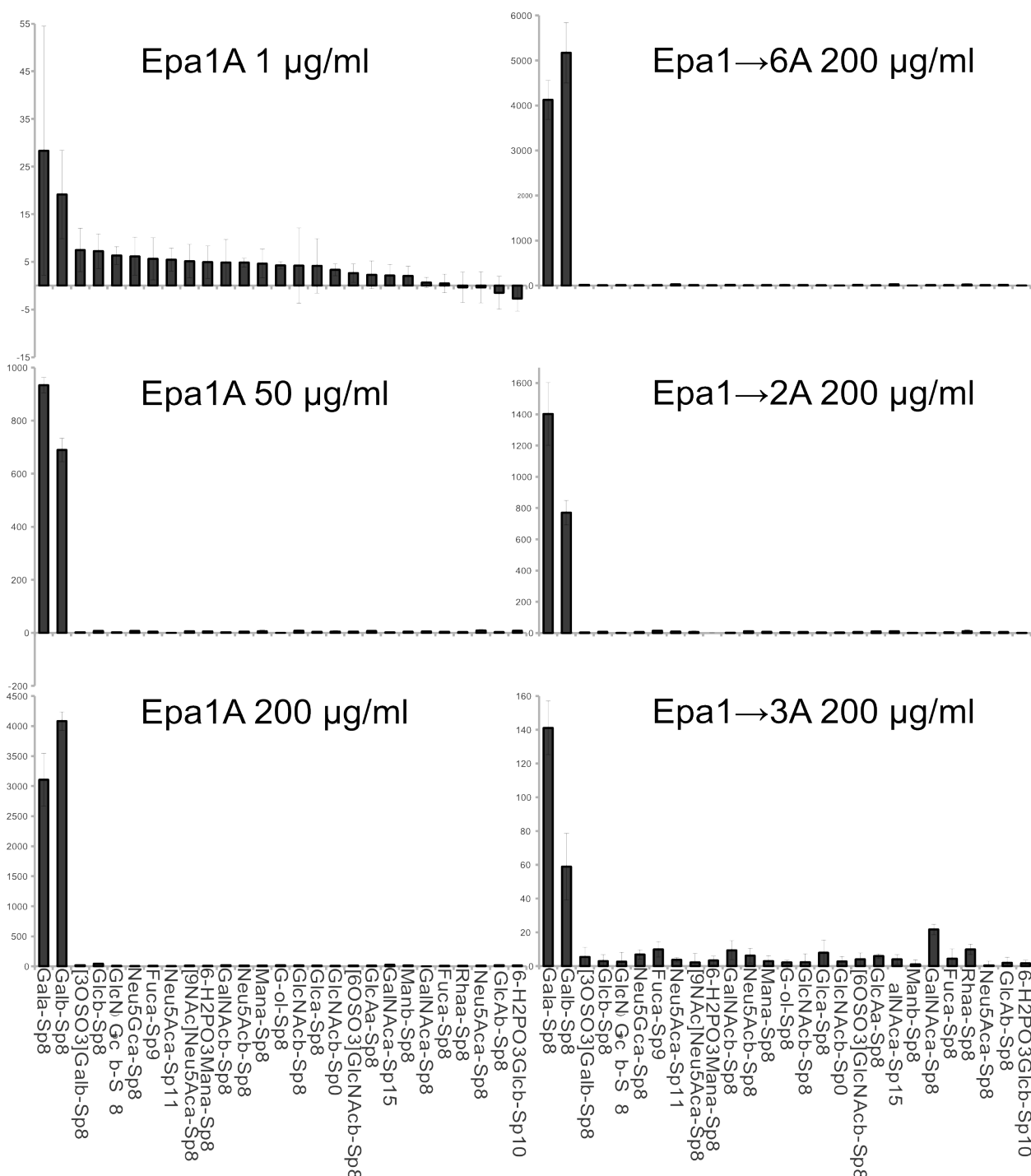


Figure 6.17: Array of all monosaccharides present in the CFG glycan array V4.1 with binding intensities for *Epa1A* and variants. Monosaccharides were ordered in decreasing binding intensity for *Epa1A* at 1  $\mu\text{g}/\text{ml}$ , and all others were maintained in the same order for comparison purposes. The preference of *Epa1A* and all variants for galactose is very clear, and further confirms a strongly binding and very specific CBL2- $\text{Ca}^{2+}$ -W198 subpocket.

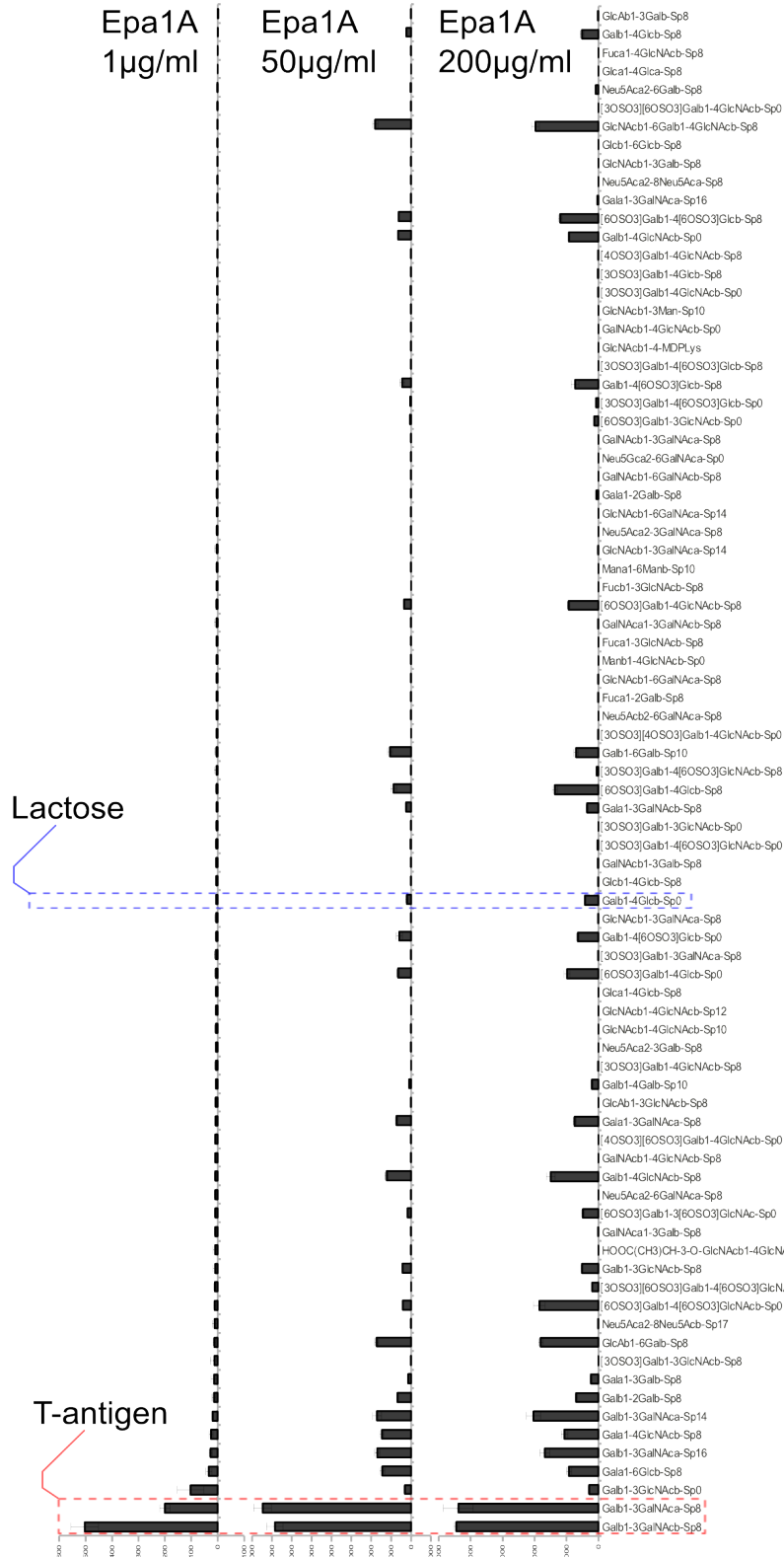


Figure 6.18: Array of all disaccharides present on the CFG glycan array V4.1 with binding intensities for Epa1A at different concentrations. Glycans were ordered by increasing binding intensity at 1 µg/ml, and kept like that at all other concentrations for comparison purposes. The high affinity for the T-antigen (boxed in red) is clear in this array, as represented by the two highest peaks. When protein availability becomes critical, at 1 µg/ml, Epa1A appears to prefer the T antigen which is β-o-linked to the surface rather than the α linked. However, at less limiting concentrations it does not show any strong preference. Lactose and its corresponding binding intensities are framed in blue, evidencing the weakness of Epa1A binding to lactose.

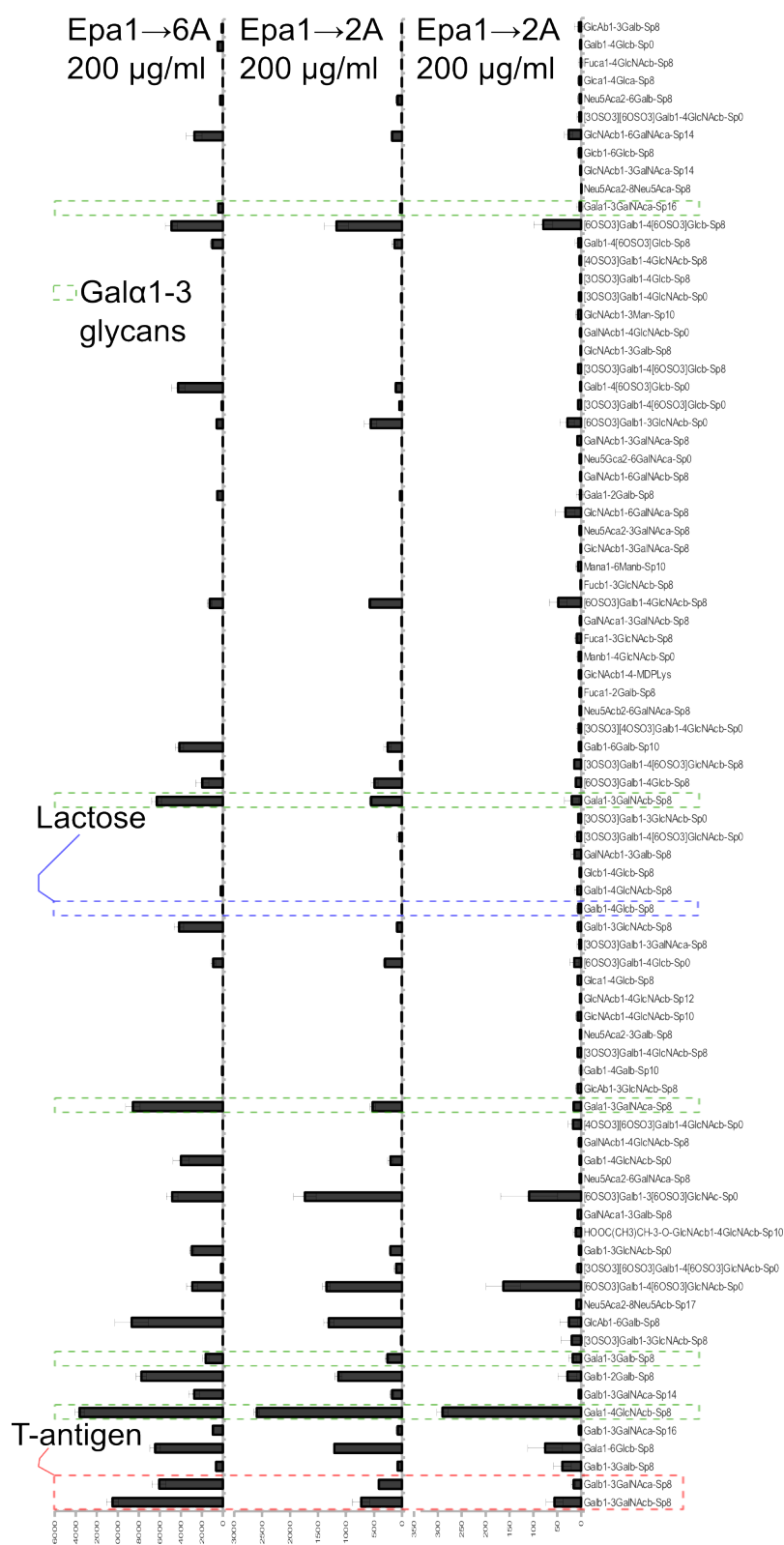


Figure 6.19: Array of all disaccharides present on the CFG glycan array V4.1 with binding intensities for Epa1A variants at 200 µg/ml. Glycans were arranged in the same order as in Fig. 6.19 for comparison purposes. In all cases, it can be clearly seen that through the change in CBL2, the preference for the T-antigen has been diminished (red box). Epa1→6A shows a high promiscuity with any disaccharide with a galactose at the non-reducing end, while in the case of Epa1→2A the preference has been reverted, and the T-antigen is not within the strongest binders anymore. There is also a marked increase for the affinity towards galactose-α1-3 containing disaccharides (green boxes). Epa1→3A shows very weak binding in all cases. Lactose (blue box), is very weakly bound, as was the case for Epa1A.

### 6.1.8 Quantitative carbohydrate binding assays via fluorescence titration

Titration curves were performed with the putative ligand of Epa1A, lactose (galactose $\beta$ 1-4glucose)[155], and the T-antigen (galactose $\beta$ 1-3N-acetyl-galactosamine). The T-antigen was one of the best binding disaccharides for Epa1A and variants (figures 6.18 and 6.19), and was very similar to the  $\beta$ 1-3 disaccharide found in the Epa1A and Epa1 $\rightarrow$ 6A co-crystal structures (Fig. 6.5).

All titrations were repeated three times, averaged and processed as described in section 5.3.5, yielding the titration curves as exemplified in Figure 6.20 (see Appendix III for a full list).

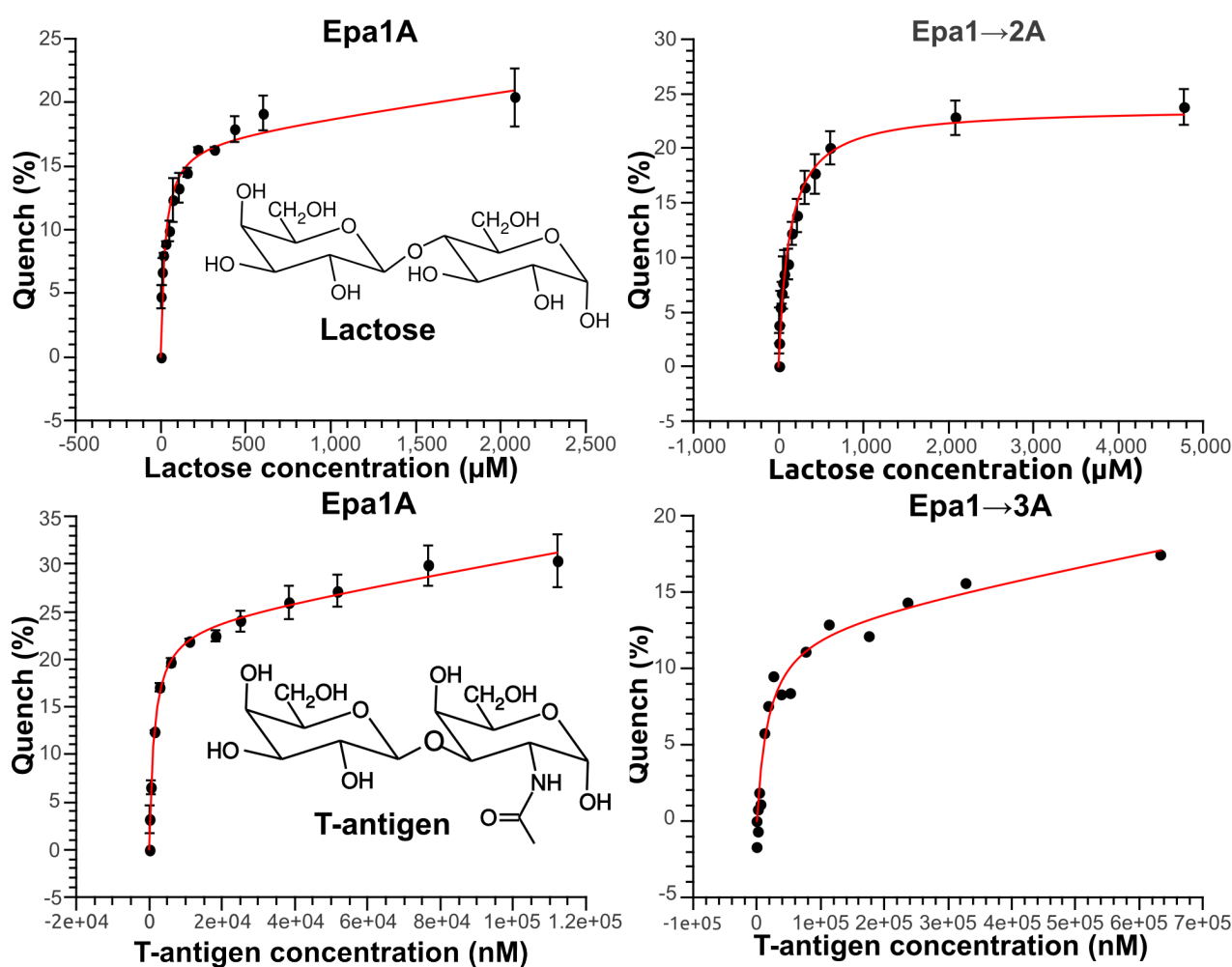


Figure 6.20: Examples of fluorescence titration curves against lactose and the T-antigen. Up: titrations against lactose for Epa1A and Epa1 $\rightarrow$ 2A. Bottom: titrations against the T-antigen for Epa1A and Epa1 $\rightarrow$ 3A. The ligands are illustrated under the corresponding Epa1A titration curves. All fits (red) were performed with equation 5.4.

Fits were performed using equation 5.5 for total (specific and non-specific) binding. Quality control

for the data by Scatchard plots[238] showed that some of the measurements presented apparently two different binding sites (Fig. 6.21). Fitting with equation 5.6 gave slightly better  $R^2$  values, but at least one of the two dissociation constants had very high standard errors (see Annex II). Since the total binding regressions did not present significantly worse fits and less error prone constants, the simplest possible model, as represented by equation 5.5, was adopted.

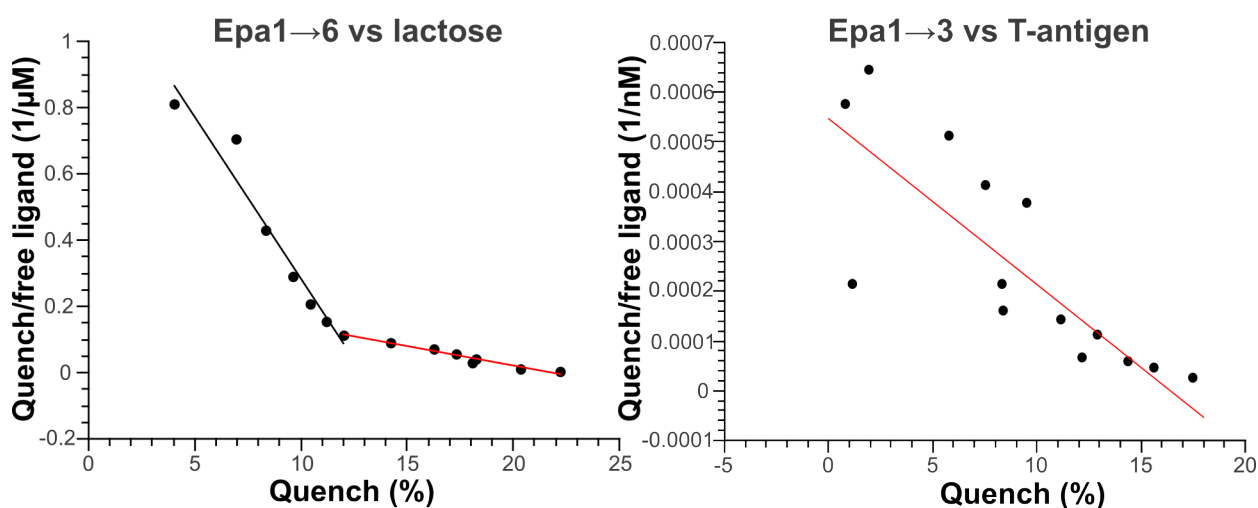
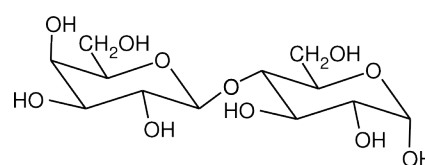


Figure 6.21: Examples of scatchard plots. Left: lactose titration against *Epa1*→6A. The fits show that there were two different binding processes during titration. Right: in the case of the T-antigen's titrations against *Epa1*→3A, data dispersion is larger, but the best fit only includes one binding process.

Protein	Lactose			
	$K_d$ ( $\mu\text{M}$ )	$Q_{\max}$ (%)	NS ( $10^{-3}/\mu\text{M}$ )	$R^2$
<i>Epa1A</i>	$24.9 \pm 5.1$	$17.1 \pm 0.9$	$1.9 \pm 0.8$	0.97
<i>Epa1</i> →2A	$200 \pm 43$	$23.1 \pm 1.1$	$-0.03 \pm 0.54$	0.97
<i>Epa1</i> →3A	$6850 \pm 3670$	$12.8 \pm 3.3$	$0.14 \pm 0.04$	0.97
<i>Epa1</i> →6A	$23.5 \pm 5.2$	$16.6 \pm 0.9$	$2.1 \pm 0.9$	0.96



Galactose- $\beta$ 1-4-glucose

Table 6.5 Summary of the constants obtained by fluorescence titration for *Epa1A* and its variants.  $K_d$ : dissociation constant.  $Q_{\max}$ : maximum quench. NS: non specific binding constant.  $R^2$ : correlation coefficient.



Protein	T-antigen			
	$K_d$ ( $\mu\text{M}$ )	$Q_{\text{max}}$ (%)	NS ( $10^{-2}/\mu\text{M}$ )	$R^2$
Epa1A	$1.4 \pm 0.1$	$23.8 \pm 0.6$	$6.8 \pm 0.8$	0.99
Epa1 $\rightarrow$ 2A	—	—	—	—
Epa1 $\rightarrow$ 3A	$17.9 \pm 5.2$	$12.9 \pm 1.3$	$8.1 \cdot 10^{-4} \pm 3.0 \cdot 10^{-4}$	0.97
Epa1 $\rightarrow$ 6A	$0.8 \pm 0.1$	$22.5 \pm 0.7$	$9.6 \pm 1.1$	0.99

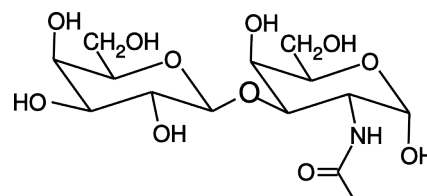
Galactose- $\beta$ 1-3-N-acetyl-galactosamine

Table 6.6 Summary of the constants obtained by fluorescence titration of Epa1A and variants against the T-antigen. At conditions where T-antigen dissociation constants could be measured, Epa1 $\rightarrow$ 2A was not stable, and measurements presented strong dispersion of data, which made analysis impossible.  $Q_{\text{max}}$ : maximum quench. NS: non specific binding constant.  $R^2$ : correlation coefficient.

When comparing the dissociation constants for lactose and the T-antigen, it becomes immediately apparent that the T-antigen is a much better binder than lactose. Epa1A and Epa1 $\rightarrow$ 6A displays at least 20-fold higher affinities for the T-antigen than for lactose. The case of Epa1 $\rightarrow$ 3A is especially interesting, since the binding of lactose can be considered marginal, while the binding for the T-antigen is approximately 400 times stronger and well within normal parameters for C-type lectins.

## 6.2 Lewis antigen binding adhesins from *H. pylori*

The Lewis antigen binding adhesins of *Helicobacter pylori*, mainly the blood group antigen binding adhesin (BabA) and the sialic acid binding adhesin (SabA), are important determinants in the colonization of the human gastric mucosa[79]. BabA has been proposed to be responsible for primary binding to fucosylated blood groups, while SabA acts as a secondary binder to sialylated Lewis antigens upon activation of the host's inflammatory response[92]. Both proteins are therefore crucial in the understanding of *H. pylori* colonization of the stomach, and represent for the development of bactericidal drugs targets alternative to the transpeptidases and the bacterial ribosome. Interestingly, though, almost no *in vitro* work has been performed on these two proteins[239]. Currently, there is no information on the bio- and physicochemical characteristics of these proteins, nor of their interactions with possible ligands, only information derived from genetic experiments or patient isolates. In this work, extracellular fragments of BabA (gene ID: AAC38081.1) could be expressed, refolded and characterized by CD and semi-quantitative glycan affinity measurements, while SabA could be recombinantly expressed for the first time. Table 6.7 gives a short summary of the employed fragments. And their nomenclature.

Name	Length	Relative length	Molecular weight (kDa)	pI
BabA full length	741	1 - 741	80.3	9.03
BabA 144	158	21-163	16.9	6.72
BabA 210	224	21-229	24.1	6.73
BabA 235	249	21-254	26.7	7.68

*Table 6.7: BabA full length characteristics and derived extracellular fragments. The relative lengths refer always to the region within BabA full length represented within the fragments. Fragments are 15 amino acids longer than the relative sequence length since they needed a start codon for translation and a C-terminal His-tag with a linker for affinity purification. All values were calculated using ProtParam[225].*

### 6.2.1 Bioinformatic studies

Since much of the biophysical characterization of BabA relied on the correct interpretation of changes in secondary and tertiary structure, a Jpred[232] *in silico* analysis of the BabA235 fragment was undertaken (Fig. 6.22). The resulting secondary structure prediction yielded preliminary information (Table 6.8) that could be used to calculate a theoretical CD-spectrum using the K2d server and the poly-Lys method[240] (Fig. 6.23).

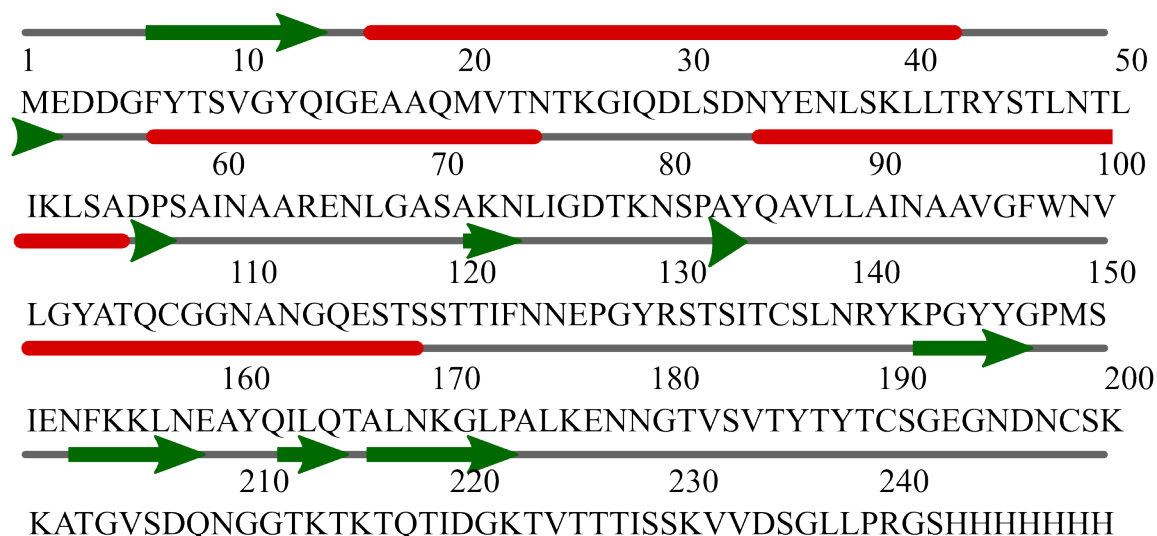


Figure 6.22: BabA235 jpred secondary structure prediction. The prediction is the output of a Jpred3 job[232].

Jpred3 prediction of secondary structure	
$\alpha$ - Helical portion (%)	34.1
$\beta$ - Sheet portion (%)	15.7
Random portion (%)	50.2

Table 6.8: Predicted secondary structure portions for the BabA235 fragment using the Jpred secondary structure prediction of Fig. 6.24.

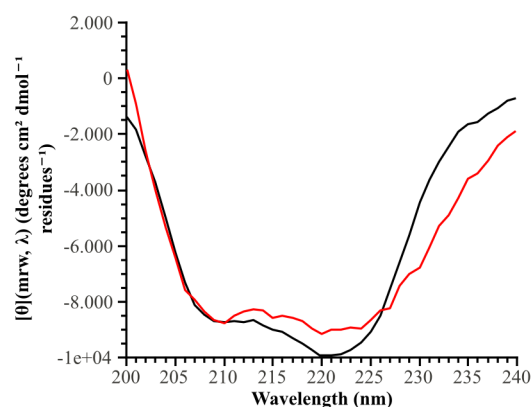


Figure 6.23: K2d prediction of a spectrum with secondary structure as per table 4.6 (black) versus BabA235 CD native spectrum at pH6 (red).

### 6.2.2 Refolding of BabA fragments

BabA extracellular fragments could only be produced in large quantities as inclusion bodies. Therefore, it was necessary to establish reproducible refolding methods. BabA fragments were first tested for refolding by dialysis. Concentrations of soluble BabA fragments measured by BCA test of pellet and supernatant fractions showed promising analytical results, with up to 92 % refolded protein. However, upscaling proved very challenging, and alternatives were tested. The iFold system 2 refolding screen (*Merck*) permitted high-throughput screening for effective and cheap refolding conditions (figure 6.24).

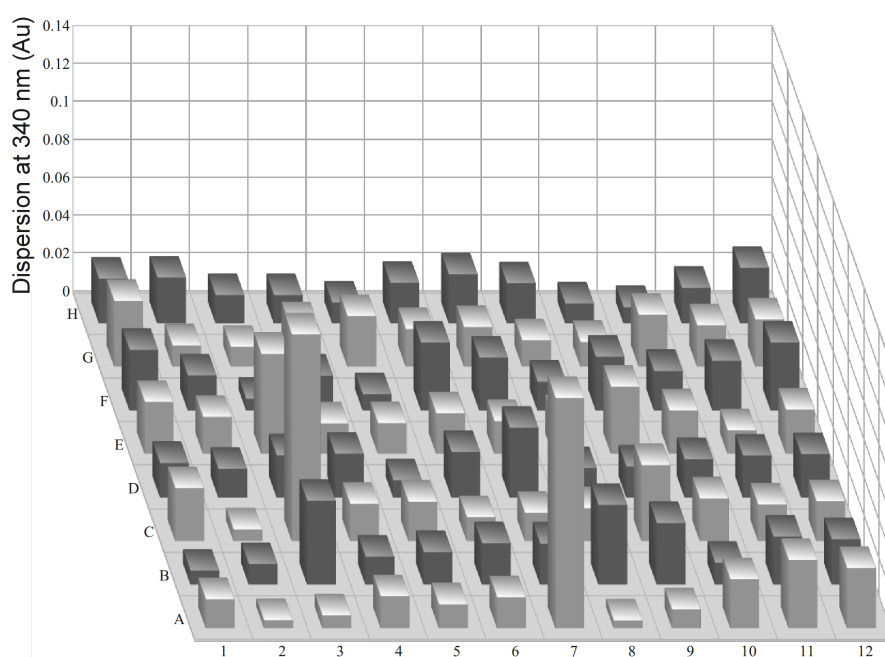


Figure 6.24: Graphical view of the iFold system 2 protein refolding screen for BabA 210. The lower the dispersion at 340 nm, the less particulate could be found in the condition, which was interpreted as BabA insoluble aggregates. Condition E4 was selected for refolding, as it had a very low signal, and had an inexpensive composition.

Condition E4 (0.1M HEPES pH 7.5, 0.5 M Arginine) was selected for further optimization, due to its good results in the screen and simple composition. Low reproducibility for condition E4 was counter-acted by supplementing it with 6 mM GSH and 4 mM GSSG. All other fragments of BabA could be maintained in solution under the same conditions. In order to characterize the refolded proteins, the high concentrations of arginine had to be reduced, as they are deleterious for analytical methods. A microdialysis screen against different buffers provided information on the best conditions for storage and characterization (Figure 6.25). Based on this, 50 mM Tris/HCl pH 6, 200 mM NaCl and 50 mM CHES pH 9, 200 mM NaCl were chosen as candidates. Although the pH 6

condition did not yield optimal results, it was considered to best represent *in vivo* conditions for BabA. On the other hand, conditions 7 and 8 in Fig. 6.22 initially seemed to be very promising, but presented poor reproducibility during upscaling for further tests.

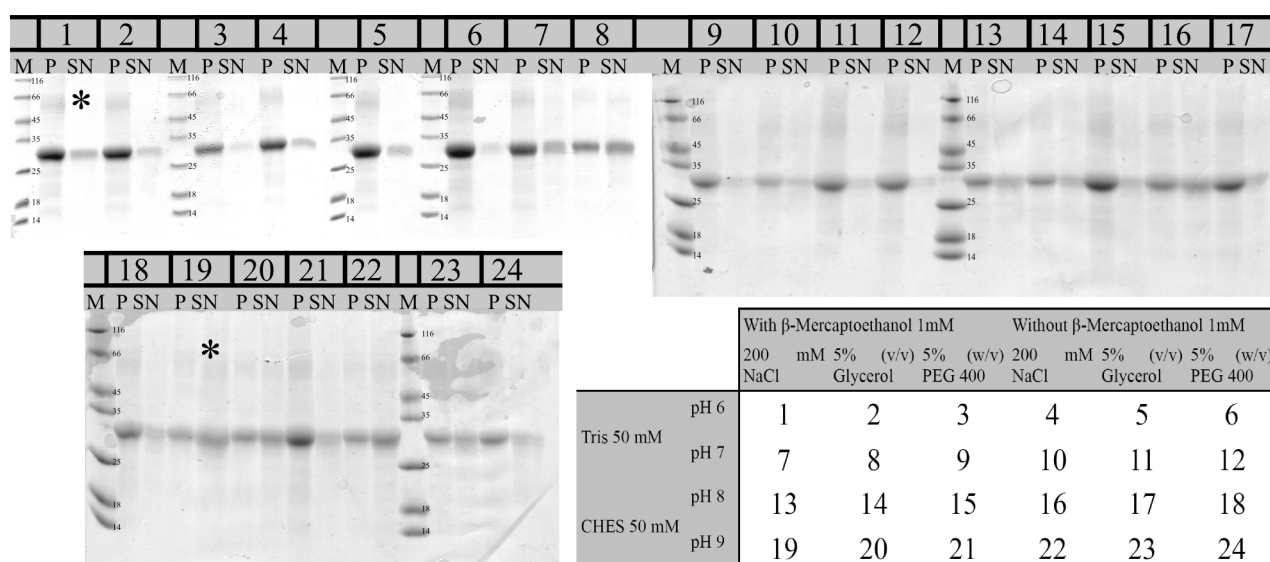


Figure 6.25: Microdialysis results of BabA 235 against different storage buffers. After at least 12 hour dialysis, samples were centrifuged, pellet and supernatant were separated and loaded on SDS-polyacrylamide gels. On the lower right side, samples are correlated to the screen. Protein is accumulated in the supernatant under very basic pHs (13 – 24), while it is mostly insoluble at neutral pHs (9 – 18). P: Pellet. SN: Supernatant. An asterisk (\*) marks selected supernatants for CD quality control.

Upscaling yielded maximal concentrations of ~15 mg/ml for BabA 210 and 235 at pH 9, while only 3 mg/ml could be achieved under pH 6. As for BabA 144, the protein was almost completely insoluble at pH 6, while maximal concentrations at pH 9 were around 5 mg/ml (Data not shown).

### 6.2.3 CD spectroscopy of BabA 235

Initial CD measurements at conditions obtained from 6.2.1 showed a marked pH dependency of the overall folding characteristics (Figure 6.26 b) and d)). To further characterize the dependency, the Tris/HCl and CHES buffers were substituted for the more broadly buffering potassium phosphate buffer, and an entire screen ranging from pH 2 until pH 10 was performed (Fig. 6.26). Interestingly, mildly acidic pHs seemed to induce a compact fold, which denatured cooperatively and irreversibly, while strongly acidic or basic pHs produced reversible denaturation with very little or no cooperativity when melting (Table 6.10). Neutral pHs were not tested, as protein solubility was negligible at or around pH 7. The behavior of BabA 210 was consistent with the one observed for BabA 235, but this was not the case for BabA 144, which could not be folded at all (data not

shown).

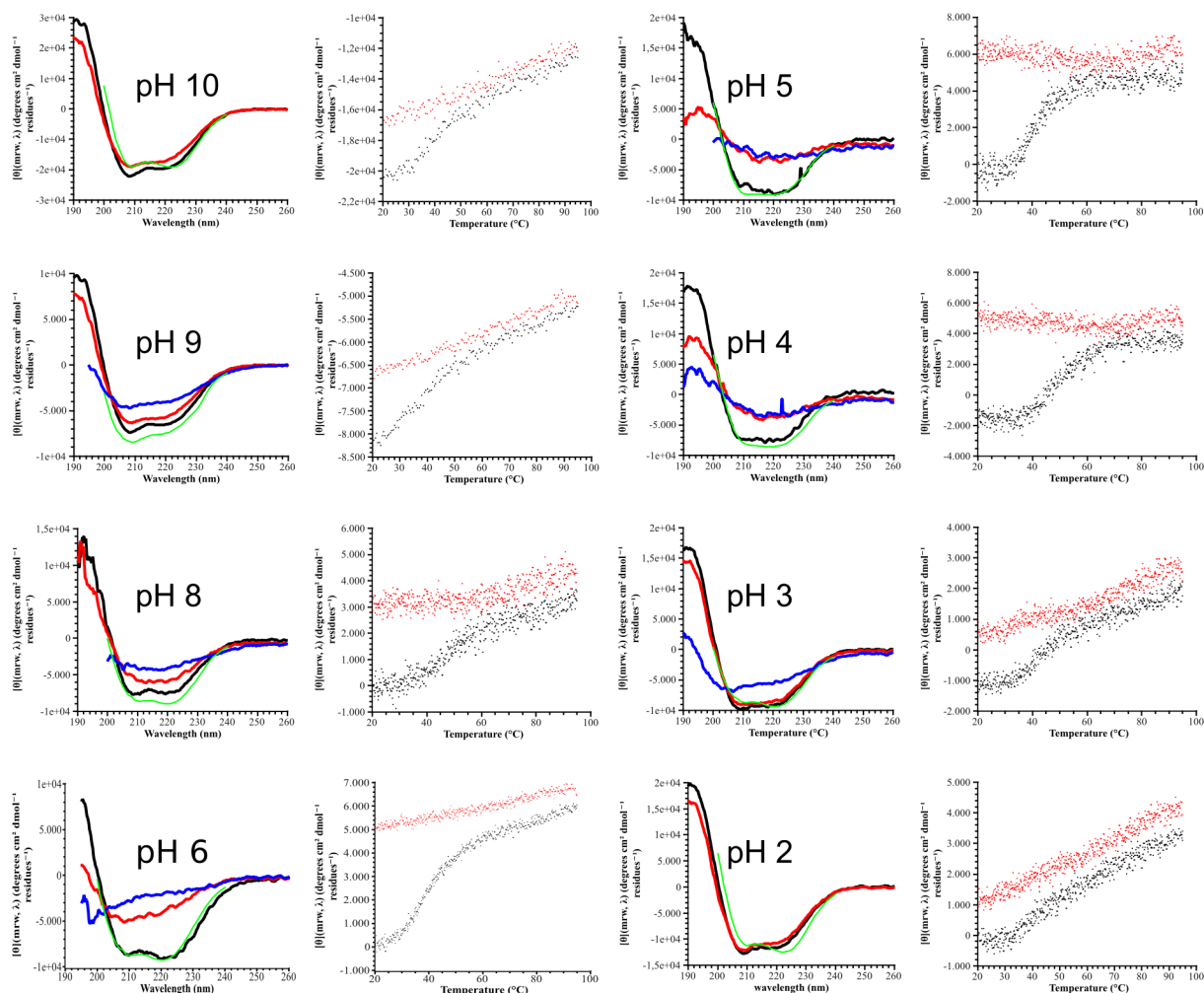


Figure 6.26: pH dependency of BabA 235 fragment as followed by CD spectroscopy. Spectra (left) and melting curves (right) under each epigraph show the unfolding process under different pHs. Spectra: black for native spectrum, red for spectrum after melting curve, blue for spectrum at 95 °C, green for deconvolution result with K2d[228]. Melting curves: black for melting, red for cooling. Protein thermal denaturation was followed at 220 nm.

Another interesting feature of the pH dependent spectra is that, regardless of the program employed, spectra within the mildly acidic range could be much better and more reliably deconvoluted than the ones in strongly acidic or basic range (Fig. 6.26 illustrates K2d[228] deconvolution results in green and Table 6.9 gives the results of these deconvolutions), up to the point where the deconvolutions at pH 10, 9, 4, 3 and 2 were considered unreliable by the software (Table 6.9).

pH 10		pH 5	
$\alpha$ - Helical portion (%)	63	$\alpha$ - Helical portion (%)	34
$\beta$ - Sheet portion (%)	6	$\beta$ - Sheet portion (%)	16
Random portion (%)	31	Random portion (%)	50
Square distance ( $\theta$ )	786	Square distance ( $\theta$ )	90
Maximum error (%)	>22.7	Maximum error (%)	18.2
pH 9		pH 4	
$\alpha$ - Helical portion (%)	23	$\alpha$ - Helical portion (%)	32
$\beta$ - Sheet portion (%)	23	$\beta$ - Sheet portion (%)	19
Random portion (%)	55	Random portion (%)	49
Square distance ( $\theta$ )	99.47	Square distance ( $\theta$ )	134.1
Maximum error (%)	>22.7	Maximum error (%)	>22.7
pH 8		pH 3	
$\alpha$ - Helical portion (%)	30	$\alpha$ - Helical portion (%)	31
$\beta$ - Sheet portion (%)	13	$\beta$ - Sheet portion (%)	10
Random portion (%)	56	Random portion (%)	58
Square distance ( $\theta$ )	77	Square distance ( $\theta$ )	398
Maximum error (%)	18.2	Maximum error (%)	>22.7
pH 6		pH 2	
$\alpha$ - Helical portion (%)	31	$\alpha$ - Helical portion (%)	41
$\beta$ - Sheet portion (%)	10	$\beta$ - Sheet portion (%)	16
Random portion (%)	59	Random portion (%)	43
Square distance ( $\theta$ )	16	Square distance ( $\theta$ )	545
Maximum error (%)	8	Maximum error (%)	>22.7

Table 6.9: Deconvolution results from K2d[228]. The program takes a maximum relative error of 22.7 as the uppermost cut-off for a reliable deconvolution. Only the deconvolution data within the mild pH range from 8 to 5 can be considered reliable, while the rest does not adjust well to the sample spectra employed during the deconvolution process.

pH	Initial $\Delta H$ (kCal/mol)	$T_m$ ( $^{\circ}C$ )	Quality of fit ( $R^2$ )
10	-9.2 $\pm$ 1.3	41.4 $\pm$ 3.1	0.98
9	-8.1 $\pm$ 1.0	31.7 $\pm$ 4.7	0.99
8	-22.2 $\pm$ 0.9	51.8 $\pm$ 0.4	0.94
6	-19.2 $\pm$ 0.3	41.4 $\pm$ 0.2	0.99
5	-38.9 $\pm$ 1.2	41.7 $\pm$ 0.2	0.97
4	-35.2 $\pm$ 1.1	50.4 $\pm$ 0.2	0.97
3	-18.7 $\pm$ 0.7	50.7 $\pm$ 0.4	0.95
2	-10.6 $\pm$ 0.7	56.6 $\pm$ 0.8	0.97
Final pH	Initial $\Delta H$ (kCal/mol)	$T_m$ ( $^{\circ}C$ )	Quality of fit ( $R^2$ )
2.5	-13.0 $\pm$ 0.3	56.8 $\pm$ 0.4	0.99
5.8	-31.2 $\pm$ 0.7	45.4 $\pm$ 0.3	1
	-49.2 $\pm$ 2.2	71.5 $\pm$ 0.3	

*Table 6.10: Initial enthalpies and melting temperatures of the melting process of BabA 235 against pH. Although melting temperatures ( $T_m$ ) are a measure of overall fold stability, initial enthalpies are a better measure of sigmoidicity in a melting curve (equation 5.19), and therefore fold compactness. A denaturation with an initial enthalpy of 0 will show a linear melting curve, while very negative initial enthalpies give very steep sigmoidal curves. All values calculated with QtiPlot and equations provided in [227]. The last two rows show the data for the spectra in Fig. 6.28 b and c, respectively. Data at pH 5.8 are calculated with a three state model based on the Greenfield equations, therefore it produces two melting temperatures and two initial enthalpies.*

Interconvertibility between the two pH dependent conformations was tested by preparing two sets of samples. The control set was composed of two samples that had been directly refolded and buffer exchanged to pH 6 and pH 2.5 in 10 mM potassium phosphate buffer (Fig. 6.28 a and b). The transition set was composed of 1 aliquot of the control sample which had been diluted 4 times in 0.1 M potassium phosphate buffer pH 6 for a final pH of 5.8 and incubated on ice for at least 2 hours (Fig. 6.27 c) and table 6.10, last row). The results show that the protein is capable of pH dependent interconversion, and that the resulting structure is very similar to the compact refolded state achieved when the protein is directly refolded at pH 6. Interconversion is not perfect, though, as some of the protein does not denature in the same manner, but presents much higher stability, resulting in two distinct transitions (Table 6.10 and Figure 6.27 c).



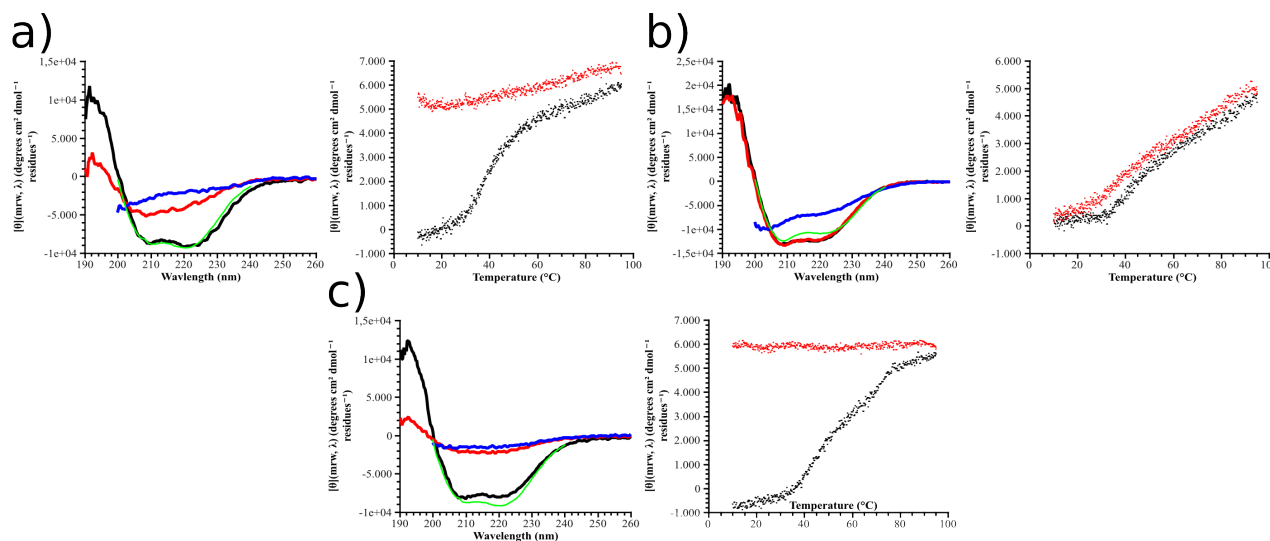


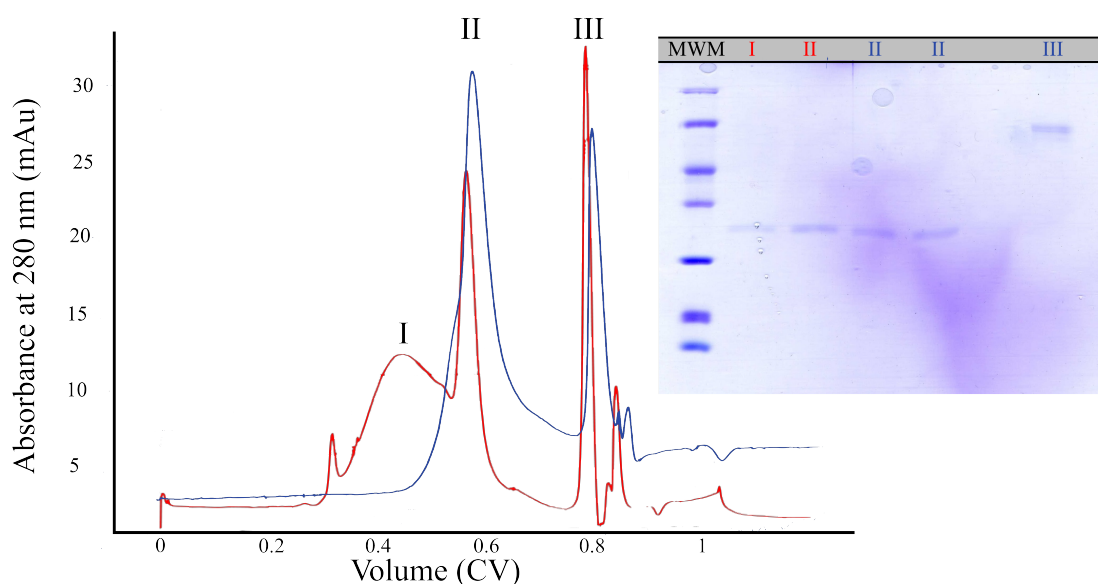
Figure 6.27: Transition of the BabA 235 relaxed conformation to the compact conformation as monitored via CD spectroscopy. Left, black: native spectrum, red: denatured spectrum, blue: spectrum at 95  $^{\circ}\text{C}$ . Right, black: melting curve, red: cooling curve, green: result of K2d deconvolution (Table 4.10). a) Sample refolded with a final pH of 6. It is the same spectrum as Fig. 6.22 d, presented here for comparison. b) Sample refolded with a final pH of 2.5. c) Sample refolded with a final pH of 2.5, which was then equilibrated with buffer at pH 6 for a final pH of 5.8. A 3-state denaturation can be observed in c

	pH 6	pH 2.5	pH 5.8
$\alpha$ - Helical portion (%)	31	32	31
$\beta$ - Sheet portion (%)	10	16	11
Random portion (%)	59	52	58
Square distance ( $\theta$ )	16	48	28
Maximum error (%)	8.0	9.7	8.0

Table 6.11: K2d deconvolution results for the transition spectra in Fig. 6.26. The values for pH 6 are written for comparison purposes. For pH 2.5 the CD curve could be fitted properly, but most importantly the secondary structure values of the transition spectrum and at pH 6 correlate very well, pointing to a common structure.

#### 6.2.4 pH dependent analytical gel filtration of BabA 235 fragment

Analytical gel filtration was performed to obtain a different, more structure-oriented perspective on the transition of the two BabA 235 conformations. Two gel filtrations were performed, one at pH 2.5 and another at pH 6. In both cases, protein at 2 mg/ml was loaded in refolding buffer, as the buffer was directly exchanged within the column (Fig 6.28). It is apparent from these results that the protein is tightly packed at pH 6, while it is at least partially aggregated at pH 2.5.



*Figure 6.28: Analytical gel filtrations of the BabA 235 fragment at different pHs. Gel-filtration at pH 2.5 (red) shows a total of 3 peaks (I,II and III) of which I and II are the BabA 235 fragment, as demonstrated by SDS-PAGE (right). Peak I runs at an apparent molecular weight above 200 KDa, and therefore either represents an aggregated or a lax state of the protein. Peak II runs at an apparent molecular weight of ~48 KDa, which is a good approximation to a dimeric protein. Peak III is not BabA 235, and was not further investigated. The gel-filtration at pH 5.8 (blue) shows only peak II and III, thus showing a more compact and defined fold under these conditions. The 12% SDS-PAGE gel (right) shows samples of all peaks, color-coded by curve of origin.*

### 6.2.5 High-throughput semi-quantitative binding assays of BabA 235

As with the Epa A domains, BabA 235 was tested for binding in cooperation with the Consortium for Functional Glycomics (CFG) on oligosaccharide-coated chips. In the case of BabA 235, the sample was stored at pH 2.5 and either exposed directly to the chip (Fig. 6.29 left), or after two hour equilibration in transition buffer at a final pH of 5.8 (Fig. 6.29 right).

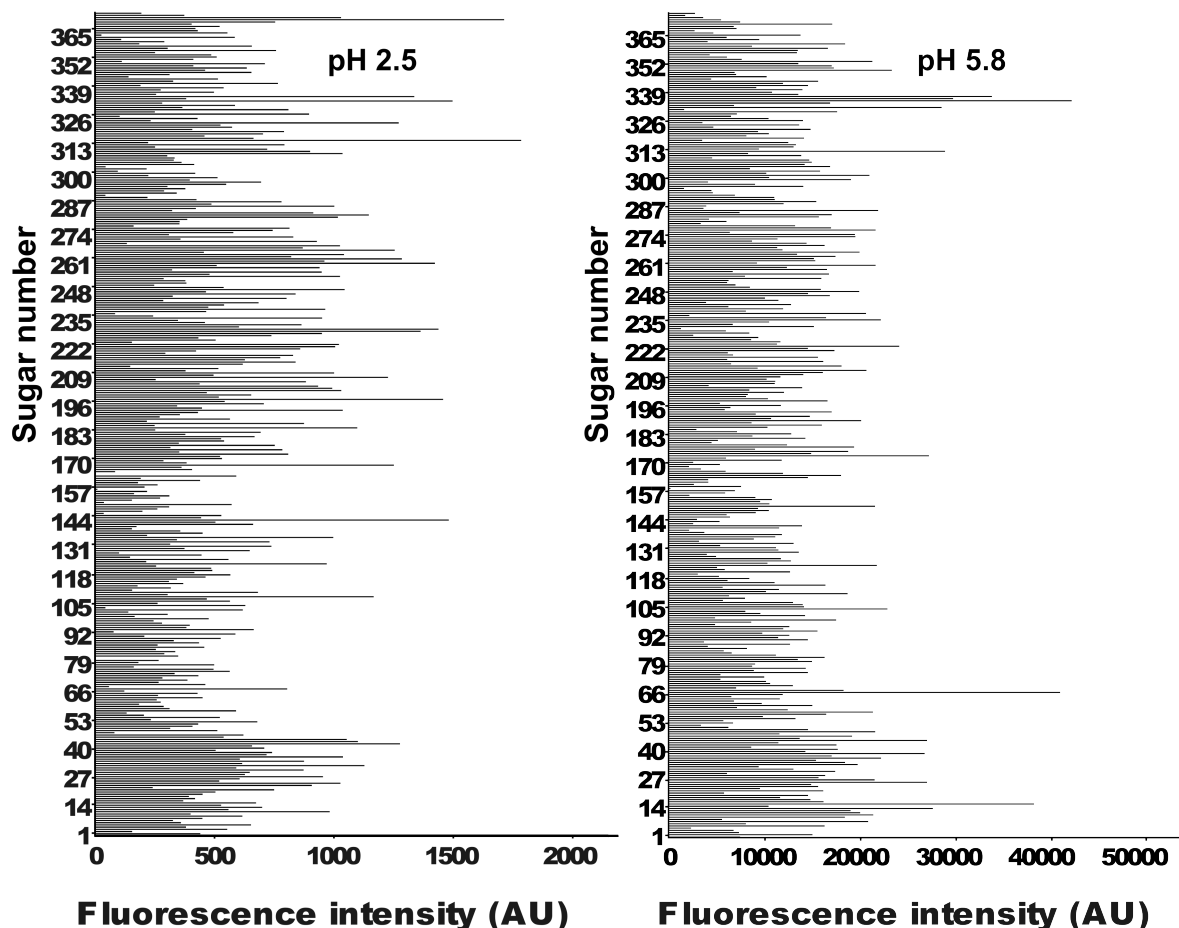


Figure 6.29: Binding of BabA 235 to the CFG V3.1 Chip at different pHs. Left: binding to the chip at pH 2.5. Right: binding to the chip at pH 5.8. Each number on the x-axis corresponds to a unique oligosaccharide (Appendix I). Fluorescence intensity is measured in arbitrary units (AU).

The data showed a strong difference between the binding capacity at pH 2.5 and pH 5.8, with a 25 fold stronger binding for the highest affinity carbohydrates at pH 5.8. That kind of difference could be also observed for the members of the Lewis antigen family (Fig. 6.30 and table 6.12). A correlation analysis was performed for the two conditions, resulting in a Pearson  $R^2$  value of 0.036% at a 95% confidence interval. These values indicate that there is no correlation between the measurements at pH 5.8 and at 2.5 as can be seen from the scatter plots presented at figure 6.31.

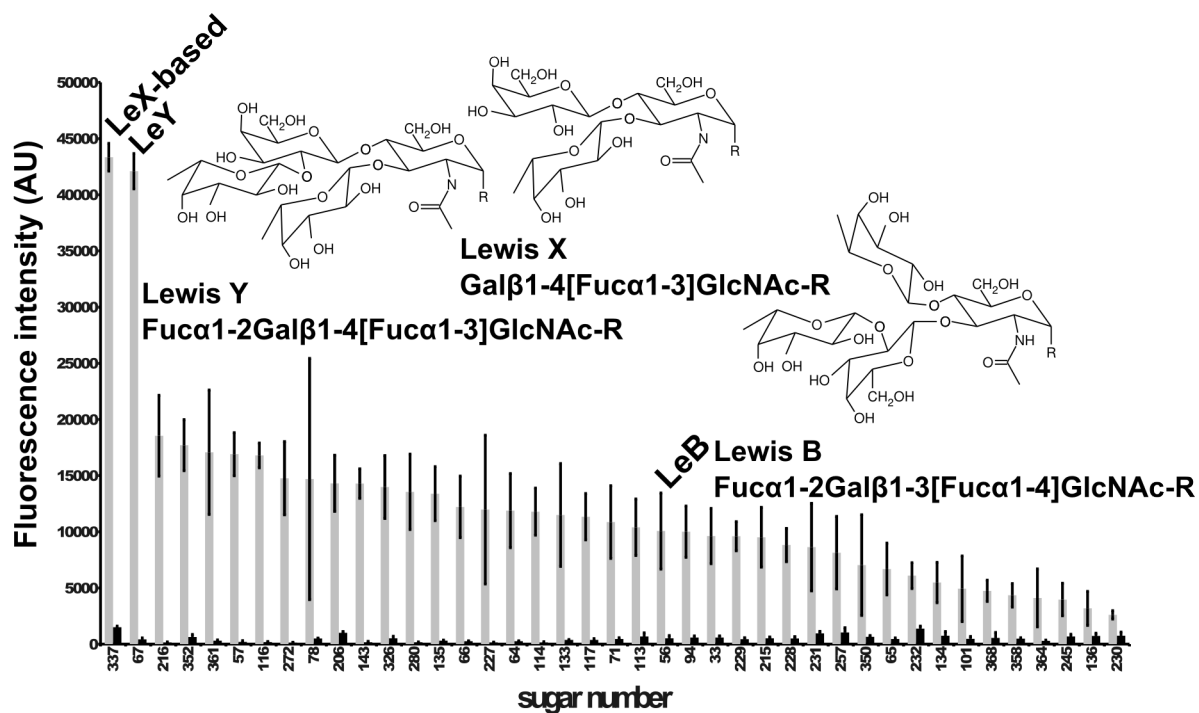


Figure 6.30: Binding of the BabA 235 fragment to the different Lewis antigens as presented in the Glycan Array V3.1. Gray: binding at pH 5.8, Black: binding at pH 2.5. carbohydrate numbers are given on the x-axis (table 6.9 or Appendix I), fluorescence intensity in arbitrary units on the y axis. Lewis oligosaccharides have been arranged in decreasing binding intensity at pH 5.8. The two strongest binders are a LeX dimer and LeY, while the putative LeB ligand for BabA is among the weak Lewis antigen ligands. Structures of the 3 ligands of interest are presented

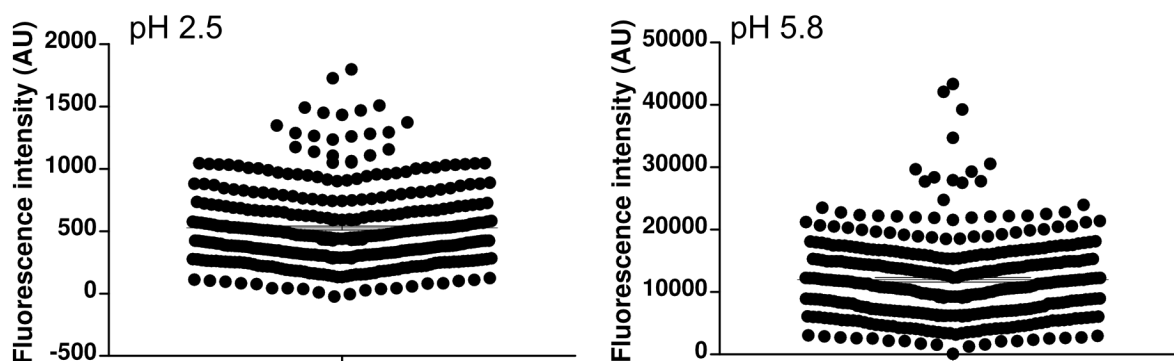


Figure 6.31 Column scatter plots for BabA binding profiles on CFG glycan array V3.1. The point distribution within each pH is quite different, indicating a lack of correlation between the two.

N	carbohydrate structure	pH 5.8		pH 2.5	
		FI	STD	FI	STD
337	GlcNAc $\alpha$ 1-4Galb1-4GlcNAcb1-3LeX1-3LeX-Sp0	43347	1235	1507	100
67	LeY-Sp8	42098	1574	431	134
216	NeuAca2-3LeAb1-3LeX Sp0	18554	3601	147	65
352	KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0	17714	2274	639	225
361	LeAb1-2Mana1-3[LeAb1-2Man a1-6]Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	17080	5542	306	78
57	<b>Fuca1-2Galb1-3GalNAca-Sp8</b>	16909	1927	205	110
116	LeA-Sp8	16807	1102	178	52
272	LeAb1-3LeAb-Sp0	14773	3261	133	37
78	GalNAca1-3LeY-Sp0	14698	10736	499	53
206	Neu5Aca2-3(6-O-Su)LeX-Sp8	14315	2506	1000	123
143	Galb1-4GlcNAcb1-3LeX1-3LeX-Sp0	14295	1317	174	78
326	NeuAca2-3LeAb1-3LeAb-Sp0	13991	2803	528	186
280	LeX1-3LeAb-Sp0	13561	3358	161	37
135	LeX1-4LeX-Sp0	13397	2407	316	51
66	LeY-Sp0	12218	2740	264	29
227	Neu5Aca2-3LeX1-3LeX1-3LeX-Sp0	11978	6629	154	30
64	LeY1-3LeX-Sp0	11892	3301	259	49
114	LeAb1-3Galb1-4GlcNAcb-Sp0	11794	2103	156	61
133	LeX-Sp0	11499	4575	376	37
117	LeAb-Sp8	11345	2062	371	117
71	<b>Fuca1-2Galb1-4GlcNAcb-Sp8</b>	10873	3241	464	115
113	LeAb1-3LeX-Sp0	10407	2513	686	326
56	LeB-Sp8	10069	3387	525	247
94	Gala1-3LeY-Sp0	10011	2287	591	143
33	[3OSO3]LeX-Sp8	9625	2462	594	147
229	Neu5Aca2-3LeX-Sp8	9613	1299	435	148
215	Neu5Aca2-3LeAb-Sp8	9518	2660	519	114
228	Neu5Aca2-3LeX-Sp0	8827	1485	507	159
231	Neu5Aca2-3LeX1-3Galb1-4GlcNAcb-Sp8	8636	3900	956	180
257	Neu5Gca2-3LeX-Sp0	8141	3228	1032	427
350	LeAb1-2Mana1-3[LeAb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp19	7041	4486	659	112

65	LeY1-3LeX1-3LeX-Sp0	6687	2305	453	107
232	Neu5Aca2-3Galb1-4GlcNAcb1-3LeX-Sp0	6098	1159	1372	242
134	LeX-Sp8	5488	1795	743	375
101	Gala1-3LeX-Sp8	4933	2925	478	204
368	Gala1-3LeX1-2Mana1-3[Gala1-3LeX1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp20	4749	956	558	486
358	LeY1-2Mana1-3[LeY1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp20	4349	1048	489	93
364	<b>Fuca1-2Galb1-3GlcNAcb1-3</b> [LeX1-6]Galb1-4Glc-Sp21	4115	2582	291	80
245	Neu5Aca2-6Galb1-4GlcNAcb1-3LeX1-3LeX-Sp0	3974	1459	689	195
136	LeX1-4LeX1-4LeX-Sp0	3187	1518	735	242
230	Neu5Aca2-3LeX1-3Galb-Sp8	2609	371	743	340

*Table 6.12: Binding of BabA 235 fragment to the Lewis antigen series present on the CFG glycan array 3.1 at pH 5.8 and 2.5. Oligosaccharides are ordered by decreasing binding intensity. Lewis X: Galb1-4(Fuca1-3)GlcNAcb. Lewis a: Galb1-3(Fuca1-4)GlcNAcb. Lewis Y: Fuca1-2Galb1-4(Fuca1-3)GlcNAcb. Lewis b: Fuca1-2Galb1-3(Fuca1-4)GlcNAcb. Bold: Lewis antigen fragments. FI: fluorescence intensity. STD: standard deviation of fluorescence intensity.*

### 6.2.6 Other Strategies for BabA235

Several other strategies were pursued for the expression, purification and characterization of BabA235. Even though no effort was spared to obtain BabA235 protein crystals, the concentration limitations at pH 5.8 and the aggregation proneness at pH 2.5 proved to be strong impediments. The strategies presented here were aimed to either produce pure, stable, homogeneous and soluble protein or to find alternative ways to structurally characterize BabA235. None of them yielded conclusive results, and they were set aside to proceed with the working system as described above. In this section, a small sample of the results of these other strategies will be presented.

#### 6.2.6.1 Alternative constructs to obtain soluble protein.

Several different constructs were tested for the BabA extracellular fragments in order to obtain soluble protein, as opposed to inclusion bodies. Table 4.10 shows all the different constructs produced for BabA.

Three different strategies were pursued to try and obtain soluble protein, namely:

- Cloning variants from other *H. pylori* strains (J99, G21 and P12), which had all been characterized as Lewis<sup>b</sup> binding. BabA is extremely polymorph, and therefore it is possible that different variants might present widely different physical properties (table 6.13).
- Targeting to the extracellular or periplasmic space through the use of the pET20 vector, which codes for an N-terminal *E. coli* protein export signal peptide within the multiple cloning site.
- Generating fusions with thioredoxin (Trx), maltose binding protein (MBP) and glutathione S-transferase (GST), which are all classical, highly soluble proteins which can be used both as affinity tags and as a way to increase protein solubility (table 6.14).

Obtained constructs were subjected to a much more exhaustive screening than that presented in table 5.5, section 5.2.1.1. *E. coli* strains BL21(DE3) Gold, Origami 2, Arctic Express, Shuffle, RIL, Rosetta and RosettaGami were tested at temperatures ranging from 12 °C to 37 °C, IPTG concentrations ranging from 0 to 1 mM and incubation periods from 4 to 72 hours (Data not shown). Under none of these conditions could soluble protein be obtained.

Strain	Gene ID	Fragments	length	Molecular Weight	pI	Identity with 26695
J99	890058 babA	210	21-231	22.5 kDa	7.9	87%
		235	21-256	25.2 kDa	8.8	
		255	21-276	27.2 kDa	8.3	
G27	6963184 babA	210	21-231	22.4 kDa	5.49	90%
		235	21-231	25.0 kDa	5.56	90%
		255	21-276	27.1 kDa	5.6	89%
P12	7010027 babA	210	21-231	22.4 kDa	4.78	90%
		235	21-231	25.1 kDa	5.02	90%
		255	21-276	27.2 kDa	5.07	89%

Table 6.13: BabA fragments from alternative strains. Molecular weights and isoelectric points were calculated with the ProtParam web tool[170], sequence identity was calculated using the blast2seq server[144]. It is important that pIs, molecular weights and sequence identities were calculated for the fragments without including affinity, tags, start methionines, etc.

Vector	feature
pET20b(+)	Periplasmic secretion signal PelB
pETM-21	Trx ORF
pETM-41	MBP ORF
pETM-30	GST ORF

Table 6.14 Alternative vectors used for optimizing expression. Vectors of the pETM series were kindly provided by the Protein Expression and Purification Core Facility, EMBL Heidelberg, Germany.

### 6.2.6.2 Alternative analytical methods tested on BabA 235

Alternative ways to characterize the BabA extracellular fragments, 235 in particular, were tested in cooperation with other research groups.

In order to characterize the different, pH dependent denaturation pathways, 1D  $^1\text{H}$ NMR spectra were to be measured by Dr. Xiulan Xie, Philipps Universität Marburg. Spectra could only be taken at pH 2.5, due to the concentration

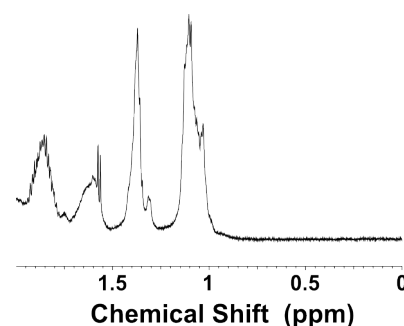


Figure 6.32 Methylene zone of the BabA235 NMR spectrum. Very low dispersion of peaks indicates a molten globule, or an unfolded state.



limitations at pH 6. The spectrum at pH 2.5 shows low dispersion within the methylene range, which is characteristic of molten globules[241] (Fig. 6.32). Thermal denaturation of BabA 235 at pH 2.5 could not be followed by this method, as the temperature range was not wide enough (the maximum temperature that the NMR device could achieve were 50°C) the concentration limitations made it impossible to proceed in the same way at pH 6, and therefore no comparison was possible.

An attempt to characterize the BabA conformational transition was performed using the hydrogen deuterium exchange mass spectrometry (HXMS) method[242][243].

BabA235 underwent an HXMS experiment at both pH 5.8 and pH 3 by Mrs. Katharina Krammer under supervision of Dr. Uwe Linne. Manual measuring procedures and climatization problems made the resulting data very difficult to evaluate, and the constants were not readily usable (data not shown).

### 6.2.7 Recombinant expression and purification of SabA extracellular fragments.

SabA is one of the other main adhesins in *H. pylori* pathogenesis, binding mainly sialyl-Lewis<sup>x</sup> molecules expressed by inflamed epithelial cells.

[92] As such, it was another very good candidate for characterization of pathogen-host adhesion mechanisms, and is also a very interesting target for drug-development .

Three different SabA gene fragments were isolated by PCR amplification from the genome of *H. pylori* 26695 (Table 6.15) and cloned over NcoI and XhoI restriction sites into the pET28a vector with a C-terminal 6-histidine-tag. No signal peptide was detected by the SignalP 3.0 server[244], as was expected due to the genetic regulation of SabA expression[245]. Finally, cloning success was validated by sequencing.

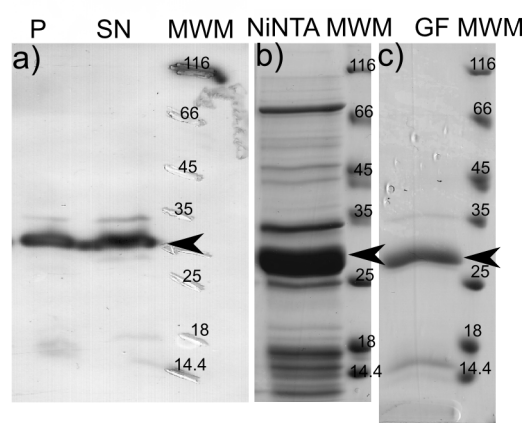


Figure 6.33: SabA260 production and purification. Arrows mark the SabA260 band. a) Western blot of a solubility test of a SabA expression in *E.coli* BL21(DE3) Gold. b) Sample purification. Left, pooled SabA260 containing fractions after NiNTA. Right, sample fraction after Gel-filtration.

Fragment	length	Molecular weight	pI
180	1-180	19.4 kDa	5.4
260	1-260	28.4 kDa	7.7
325	1-325	35.3 kDa	8.5

*Table 6.15 SabA fragments. Data was predicted using the ProtParam webtool, no tags were included for these calculations.*

After analytical expressions as described in table 5.5, Western blotting detected small but significant amounts of SabA260 in the supernatant (Fig. 6.33 a), while no expression could be detected for any of the other two fragments (data not shown). Purification by Ni-affinity columns and Superdex 200 size exclusion chromatography yielded relatively pure protein at a rate of ~0.1 mg per L of culture (Fig. 6.33 b).

Crystallization trials have been started without any positive results yet, and a functional characterization of the produced protein needs yet to be done.

## 7 Discussion

*C. glabrata* and *H. pylori* have received much attention in the field of infection biology, since they are considered to be to emergent, and widely spread pathogens[45][77]. The research performed on *C. glabrata* and *H. pylori* has been mainly based on *ex-vivo* clinical samples or whole cell studies. This approach has the advantage of setting the organism within its native ecological niche, and therefore allows for system-wide information to be obtained. On the other hand, it can't answer questions about the molecular nature of the interaction between the pathogen and its host, and is prone to produce ambiguous results, which are not easy to interpret. The *in vitro* biochemical, functional and structural studies performed in this work address precisely these weak points in current research, and complement it with strong molecular evidence.

Interestingly, the combination of *in vivo* and *in vitro* evidence highlights important similarities in the adhesion behavior of such distant organisms as *C. glabrata* and *H. pylori*. While the BabA and Epa adhesins seem to follow completely different paths to be exported and presented on the surface[63][106], they appear to show similarities in their binding specificities, targeting branches and cores of mucin O-glycans. The adhesion behavior of both organisms could be shown to vary rapidly between different colonization sites[90][27][26]. Compelling evidence could be presented for how such changes can occur on the molecular level. Finally, a flocculation-like behavior could be postulated for *H. pylori*, reminiscent of the functions of fungal adhesins[1].

In conclusion, the molecular evidence points towards a process of convergent evolution towards the mucin-mediated colonization of human epithelial tissues of two distantly related, but functionally similar, adhesion systems.

## 7.1 *H. pylori* adhesin BabA

*H. pylori* is one of the most prevalent pathogens in the human population[77], And is therefore considered one of the main targets for drug development. One of the differential systems that could be used for efficient therapy is the adhesin system, which has been described as crucial for infection and persistent colonization of the stomach. First and foremost in the adhesion system is the blood group antigen binding adhesin (BabA), which acts as primary receptor during colonization[4].

In this study, BabA could be produced for the first time in large quantities, and reliably refolded into functional, soluble protein. New biophysical and biochemical insights could be obtained, showing that BabA presents an extremely dynamic structure. Finally, a strong pH dependent activity could be studied via the Consortium for Functional Glycomics technologies, which could be set in a wider, more physiological context[90].

### 7.1.1 *The BabA extracellular domain can suffer profound pH-dependent conformational changes.*

Even though the Hp\_OMP domain (pfam accession number PF01856[246]) is unique to the *H. pylori* species[123], it has been shown that it has the characteristic C-terminal outer membrane insertion sequence[89], consisting of a C-terminal phenylalanine (Hop family) or tyrosine (Y-Hop sub-family) and an 11 amino-acid long C-terminal sequence composed mainly of aromatic residues. The outer membrane insertion sequence is a characteristic of the vast majority of OMPs[124], and therefore, it can be assumed that the Hp\_OMP domain will insert itself and show a similar  $\beta$ -barrel like structure. This claim was further supported by the fact that the C-terminal region of all Hp\_OMP domain containing proteins shows rows of amphipathic sequences, which are predicted to form anti-parallel  $\beta$ -sheets and can be modeled into  $\beta$ -barrels (figure 7.1). In the case of BabA, amino-acids 547 to 741 are predicted to contain a total of 5 to 7 long, amphipathic  $\beta$ -strands, which is within the norm presented in the pfam curated model[246]. All members of the Hop family show very conserved C- terminal and N-terminal regions. A highly variable central to N-terminal region is supposed to confer ligand specificity in an unknown manner[42].

All this information suggests that BabA will insert itself into the membrane in a similar way than the outer membrane proteins of other organisms, like OmpA, for example. If this would be the case, the C-terminal  $\beta$ -barrel like Hp\_OMP domain would first be built into the membrane, leaving the N-terminal region within the periplasmic space (figure 7.1). In an autotransporter like fashion, the

N-terminal region would then be “threaded” through the newly formed pore or passed through the membrane in a Bam system dependent manner and extruded into the extracellular space, where it would be folded[105][106][120].

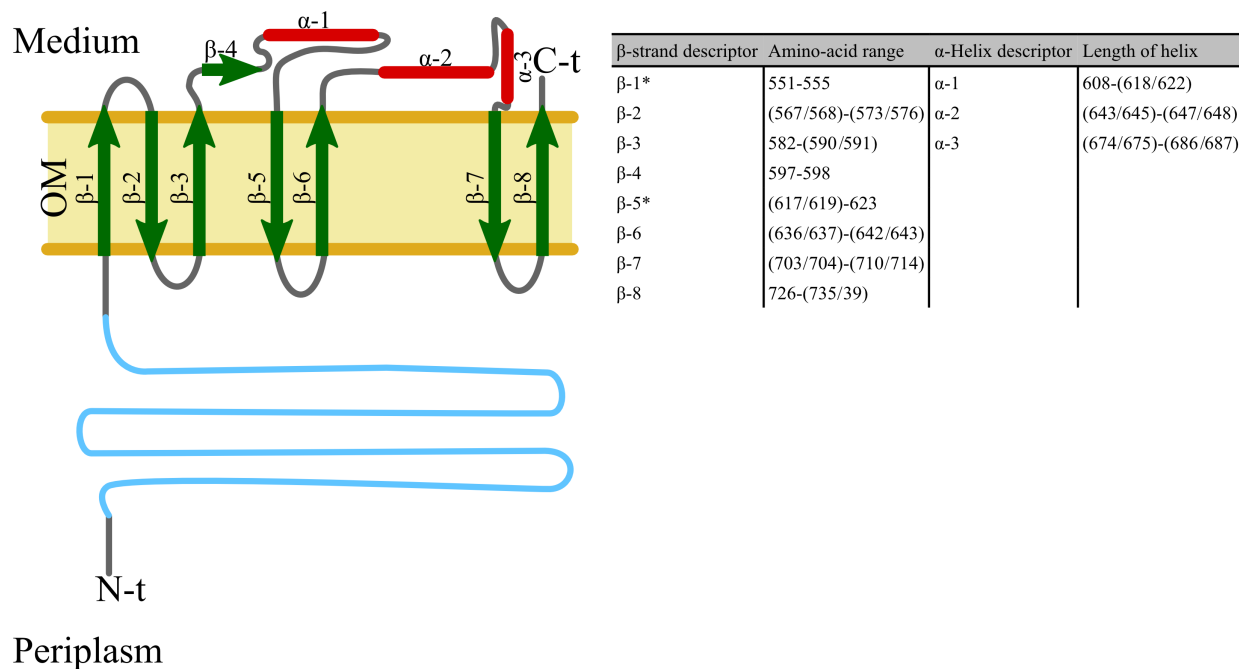


Figure 7.1: Transmembrane topology of the BabA HP\_OMP domain based on a JpredV3 prediction[232]. The putative transmembrane β-barrel is formed by 5 to 7 antiparallel, transmembrane β-strands (β-1-3 and 5 to 8), spanning through the outer membrane (OM). On the extracellular side, one short β-strand (β-4) and 3 α-helices can be found. The length of each secondary structure feature is given by the start and end amino-acids on the table on the right. Numbers in parentheses imply length discrepancies between the different evaluation methods offered by the Jpred server. β-strands marked with an asterisk (\*) were given low confidence by the Jpred algorithm. With or without the low probability strands, the N-terminal adhesion domain of BabA (blue) topologically ends in the periplasm, and needs to be exported into the medium across the membrane. This will most likely be accomplished in an autotransporter-like fashion[106].

The CD data in section 6.2.2 and the size exclusion chromatograms in section 6.2.3 support the idea that, once in the extracellular space, the N-terminal domain of BabA shows a high plasticity and flexibility. At very acidic or basic pHs, the BabA extracellular domain seems to aggregate even at relatively low concentrations and to be present in a relaxed, open conformation, as can be seen from the temperature dependent changes in dichroic behavior and the elution profile at pH 2.5. In contrast to the aggregation-prone behavior, solubility under these conditions is high. As soon as the protein is submitted to mildly acidic or neutral pHs, it forms a much better defined, compact structure, as observed on the elution profiles and CD melting curves at pH 6. Interestingly,

solubility is very low at mildly acidic pHs, even though the theoretical pI of the protein is 7.7.

The contrast between these two structural isoforms suggests that the BabA extracellular domain does not fold entirely at extreme pHs, but forms a molten-globule like structure[247]. The BabA molten globule structure is characterized by a strong tendency to aggregate maintaining high solubility. This suggests small sized aggregates[248] as corroborated by size exclusion chromatography, where most of the BabA aggregates at pH 2.5 elute within the column volume.

In contrast, at mildly acidic pHs, the conformation is much tighter, but presents a much higher tendency to precipitate at mid to high concentrations. The aggregates formed upon concentrating are very large, as they immediately precipitate and were visibly removed from the suspension upon sterile filtering, as evidenced by a decrease in turbidity and a tendency to clot the filter. As, of course, unfiltered samples would clot the equipment and seriously damage it, no attempt was made to characterize the unfiltered suspension over size exclusion chromatography.

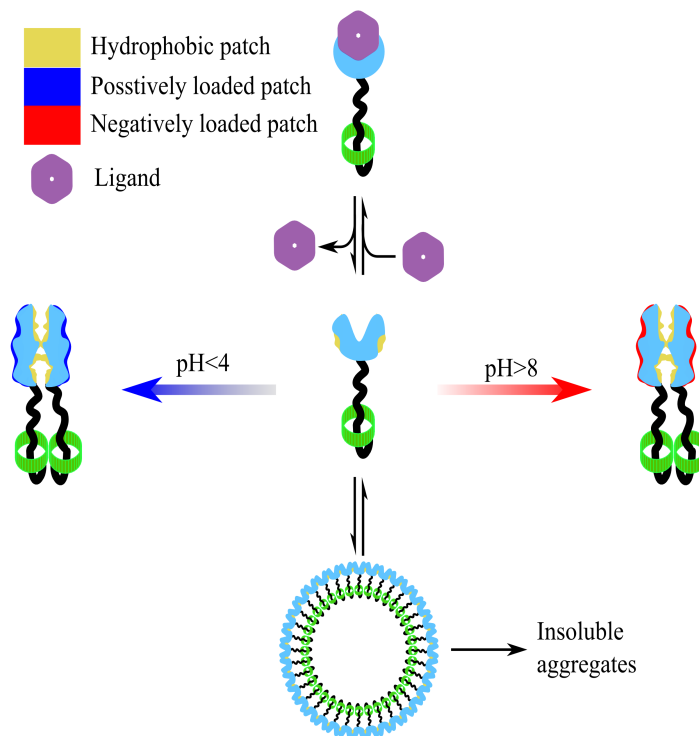
The compact state at mildly acidic pHs is the active state of the protein, probably stabilized by salt bridges, as is implied by the fact the relaxed conformation is obtained by providing an excess of either negative or positive charges to the protein (Figure 7.2). The instability of the active conformation might be explained through a still not completely folded state, which needs the ligand to perfectly fold, as was proposed before for some enzymes by Koshland[249]. This idea is further supported by the carbohydrate-binding profile of Fig. 6.28 b), which shows a moderate specificity, but high avidity for certain carbohydrates.

To summarize, BabA is present on the cell surface of *H. pylori* in a mixed state, ranging from completely molten globular at extreme pHs to partially folded at mild pHs. Upon binding to its ligand, BabA might undergo a third conformational change which yields the fully folded protein (Figure 7.2).

This model for BabA structure plasticity is in no way incompatible with previous studies, as the aggregation proneness of both the relaxed and compact conformations would be counteracted by the relative scarceness of BabA on the *H. pylori* surface, and the fact that it would only adopt its folded conformation upon entering the ecological niche where the ligand would be present (i.e., the deep mucous layer of the gastric antral region). It is also not incompatible with the fact that no Lewis<sup>b</sup> binding activity has been characterized for *H. pylori* strain 26695, as it could be shown that the specificity is shifted for the BabA presented in this work. Finally, it still holds true for any other *H. pylori* strain, because pH dependency in the binding and activity of *H. pylori* adhesion factors has

been studied and demonstrated for a wide variety of strains[90][250].

On the other hand, it very nicely complements BabA heterogeneity[122], polymorphisms[121] and *H. pylori* host-evasion mechanisms[251] by presenting an expression-independent and rapid BabA activity regulation mechanism. A pH-dependent conformational transition is not an uncommon mechanism to modulate activity, as has been shown for human growth hormone[243], or  $\gamma$ -conglutin[252]. It has also been recently pointed out as a feature of *H. pylori* binding to the gastric mucosa[87], specifically for the BabA-Muc5A interaction over Lewis<sup>b</sup> antigens.



*Figure 7.2: The BabA extracellular domain has four possible conformational states. At mild pHs, low concentrations and no ligand (center), the BabA extracellular domain is in a compact, but not fully folded state, which is aggregation prone. If the concentration is raised to unphysiological levels, BabA aggregates and, after the aggregates reach a critical size, they precipitate (center, low). If the ligand is supplied, the BabA extracellular domain fully folds around it (top). If the pH is either lowered or raised dramatically, salt-bridge partners stabilizing the tertiary structure of the compact conformation lose their charge, and therefore the domain unfolds, showing larger hydrophobic patches (left and right). The protein then tends to aggregate through the hydrophobic patches, but a stronger positive (at acidic pHs) or negative (at basic pHs) net charge prevents large aggregates from forming, and keeps small aggregates in solution.*

### 7.1.2 *H. pylori* interacts with the host in a pH dependent manner

Recent results have shown that *H. pylori* supports different binding modi for different anatomical parts of the host's gastrointestinal tract, and that the activation of the different binding modi is pH related[90]. The results in the present work (sections 6.2.2 to 6.2.4) highly support a pH dependent modulation of BabA binding activity (see 7.1.1.).

It is therefore possible to postulate a model of *H. pylori* mucosal binding in which BabA is only active on the cell surface at mildly acidic to neutral pHs, while it will remain inactive at strongly acidic pHs, not unlike the conditions to be found within the gastric lumen (Fig. 7.3). Under this model, the conformational flexibility of BabA contributes to the positioning and permanence of *H. pylori* in the antrum. As *H. pylori* would enter the gastric lumen, it would come in contact with highly glycosylated mucins in solution projected from the mucosal surface. With the gastric lumen at pHs well below four, BabA would not be active and no specific binding would occur under these conditions, only electrostatic interactions can take place[9][88]. The bacterium buries itself deeply into the mucin sheet above the epithelial cells in the lining of the stomach's antral region[128], the surrounding pH would start to raise, until it would surpass the threshold between three and four, at which point BabA would change its conformation and become active. The nearer *H. pylori* would get to the epithelial layer, the longer it would stay stationary. As peristaltic mucous clearance eliminates layers of mucin surrounding *H. pylori*, the pH would start becoming more acidic again, at which point BabA would become inactive, thus liberating the cell from the cleared mucin molecules and allowing it to return to the mucous layer. As such, BabA conformational transitions would form a rapid and short range complement to other conventional chemotactic mechanisms described for *H. pylori*, such as the CheA/Y system[83]. The mechanism of binding proposed in figure 7.3 would correlate quite well with other models, in which 4 different, tissue specific binding modi are proposed for *H. pylori*[90].



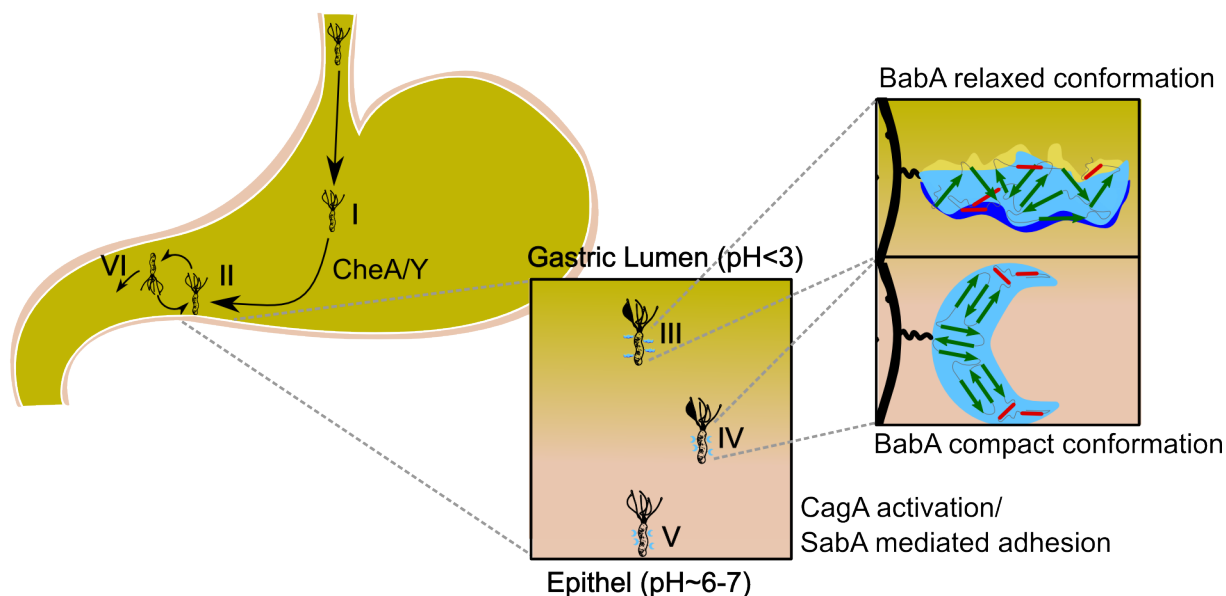


Figure 7.3: Model for BabA dependent gastric colonization by *H. pylori*. I: *H. pylori* enters the human stomach through the oral cavity, and is chemotactically led to the epithelium of the antrum[83]. II Once on the antral surface, *H. pylori* interacts electrostatically with the mucosal surface and (III) buries itself within it. BabA is in its relaxed conformation, and is not active under these conditions. IV: *H. pylori* enters the mucosal region in which the pH is above four. Part of the BabA population on the bacterial surface transitions to the packed conformation and becomes active. Around 90% of an *H. pylori* colony will be found at this position[88]. V: *H. pylori* has reached the epithelial cell surface. After colonization, CagA mediated infection is activated, and as a result, the stronger and more specific SabA binding is made possible[92]. Under these conditions, peristaltic mucus clearance can liberate some of the colony members into the gastric lumen (II) at which point, BabA would retransition to the relaxed conformation, thus becoming inactive. Again, chemotaxis can drive the first steps into the mucous layer or the bacterium might be washed from the stomach and into the duodenum (VI).

### 7.1.3 BabA specificity shift towards Lewis<sup>x</sup> and Lewis<sup>y</sup> blood group antigens.

*In vivo* analyses of *H. pylori* have repeatedly shown that BabA binds preferentially Lewis<sup>b</sup> blood group antigens, establishing a linear relation between pathogenicity and BabA expression[120]. This information has been recently challenged by more detailed research, which portrays no simple relationship between strong Lewis<sup>b</sup> binding and BabA expression[121].

No BabA N-terminal consensus sequence responsible for Lewis<sup>b</sup>-binding has been characterized yet[122], only a 6 amino acid sequence whose deletion fully disables, but does not modulate, BabA activity [253]. BabA shows a rate of variability within the extracellular region which correlates well with the highly polymorphous *H. pylori*[254]. Correspondingly, even though most bab<sup>-</sup> *H. pylori* strains do not bind Lewis<sup>b</sup> antigens, there are some that do, and within bab<sup>+</sup> strains, affinity for the oligosaccharide can be extremely variable[42][122][255].

In this context, the seemingly unexpected result that the refolded BabA235 gene product

preferentially binds Lewis<sup>X</sup> and Lewis<sup>Y</sup> instead of Lewis<sup>b</sup> (Fig. 6.29 and table 6.10) could be understood as a product of rapid evolution. The surface molecules of *H. pylori* are some of the most important targets in immune recognition of the pathogen, and are therefore subjected to a stronger evolutionary pressure than other, non-exposed molecules. Furthermore, the bab system is in constant fluctuation, as frequent recombination occurs between the *baba* and the *babb* gene loci[121], resulting in functional, yet varying gene products that are thought to keep the bacterium safe from immunological detection.

Still, in the results presented in section 6.2, a completely different specificity has evolved, and not a modulation of the specificity or a cessation of activity. The question arises, as to why a shift towards Lewis<sup>X</sup> and Lewis<sup>Y</sup> specificity for BabA might increase the survival potential of *H. pylori* within its human host. *H. pylori* produces Lewis<sup>X</sup> and Lewis<sup>Y</sup> antigens itself, and attaches them to its own lipopolysaccharide (LPS)[256][257]. LPS-bound Lewis<sup>X/Y</sup> recognition by epithelial cells is key in the binding of the gastric surface, even though the putative epithelial lectin has not been isolated[251], and can induce the production of anti-Lewis<sup>X</sup> and anti-Lewis<sup>Y</sup> autoantibodies[258].

Anti-Lewis<sup>X</sup> antibodies not only do not reduce *H. pylori* binding to epithelial surfaces, but make it better, and even increase both the epithelial damage and binding capacity in *baba*<sup>-</sup> mutants to levels that are superior to those of *baba*<sup>+</sup> strains[127]. The model proposed by Sheu *et. al.* was that *H. pylori* would form large complexes with Anti-Lewis<sup>X</sup> autoantibodies and epithelial Lewis<sup>X</sup> oligosaccharides present on tethered mucins, like muc5A, on the epithelial cell surface, *de facto* avoiding completely the use of a lectin and subverting antibody function (Figure 7.4 b)). The resulting bacterial clusters would be fixed on the epithelial surface, resulting in several advantages for the pathogen. The clusters would present a much higher avidity for the human cell surfaces than the single bacteria. The individuals within the cluster would be better protected from both the extreme conditions of the stomach and, more importantly, from the host's immune response.

If a *bab*<sup>-</sup> strain, such as 26695, would manage to shift BabA specificity towards Lewis<sup>X</sup> or Lewis<sup>Y</sup> instead of Lewis<sup>b</sup>, it would allow it to mimic Anti-Lewis<sup>X</sup> autoantibody behavior without having to rely on the host's specific immune response. The result of such a shift would be an *H. pylori* flocculating strain, which would not present a sparse Lewis<sup>b</sup> binding cells within the mucin matrix (Figure 7.4 a), but large, highly avid Lewis<sup>X</sup> binding cell aggregates or flocks (Figure 7.4 c). Incidentally, a very similar survival strategy has long been described for the non-pathogenic yeast *S. cerevisiae*. In the baker's yeast, aggregation or *flocculation*, is mediated by the Flo5 adhesin[1],

whose A domain bears remarkable similarity to Epa1A from *C. glabrata*.

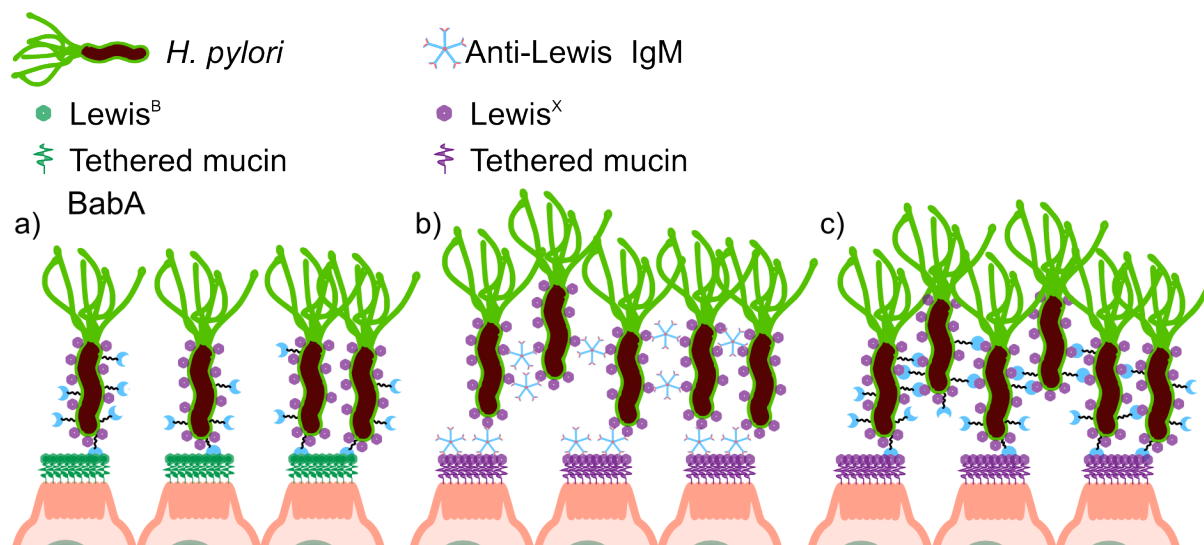


Figure 7.4: *Lewis<sup>x</sup> mediated, H. pylori-epithelium interactions.* a) Normal *H. pylori* cells colonize the antral surface by binding *Lewis<sup>B</sup>* antigens. b) In the model proposed by Sheu et. al. [127], anti-*Lewis<sup>x</sup>* IgM auto-antibodies act as bridges between the *Lewis<sup>x</sup>* molecules present on both the epithelial and bacterial cell-surfaces. b) In the model proposed here, specificity shifted BabA substitutes the IgMs to produce the same result, but conferring the additional advantage for the pathogen of not having to wait until the specific immune response is activated by the host.

#### 7.1.4 Outlook

The research presented in this work represents a first strong attempt to biochemically characterize BabA, one of *H. pylori*'s main adhesins. Even though strong arguments could be delivered by these results, many questions remain open.

It is imperative to test the same procedures presented here for the *H. pylori* 26695 BabA with the equivalent adhesin from a *Lewis<sup>b</sup>* binding strain. Results in this direction would shed light on the nature of the specificity shift of the 26695 BabA. The four-conformations model for the BabA extracellular domain predicts that protein stability should increase at mild pHs with a strong ligand. Stability tests through thermofluor assays might yield interesting results in this respect.

Obtaining soluble BabA through other methods than those attempted seems feasible, as for example the expression in *Pichia pastoris* has shown very good results in the expression of other recalcitrant, cysteine rich proteins (unpublished data).

## 7.2 Epithelial adhesins from *C. glabrata*

*C. glabrata* is an opportunistic pathogen which is responsible for the majority of iatrogenic fungal septicemia [44]. It can also be encountered in several, less severe diseases, like fungal vaginitis[45], infections of the oral cavity or the urinary tract[46]. A very important feature of *C. glabrata* is its extremely diverse and unique repertoire of adhesins, which allow it to adhere in several different ways to its human host, and control the initial stages in pathogenesis[56]. It has been proposed, but never confirmed, that the different epithelial adhesins (Epa) within the *C. glabrata* genome are responsible for the adhesion to different tissues[51][56].

The present work presents important discoveries for the understanding of the Epa adhesion system in the fields of biochemistry and structural biology. The structure of the A domain of Epa1 could be solved, yielding a profound and detailed knowledge of the carbohydrate binding pocket. As a result, the Epa1A binding pocket could be transformed through mutagenesis into that of its nearest relatives within the Epa family of proteins. These variant structures could be solved, and showed very enlightening differences within the Epa1A structure. The validity of the structural results was further confirmed through semi-quantitative, high-throughput binding assays in cooperation with the *Consortium for Functional Glycomics* (CFG), which was able to characterize the oligosaccharide specificity for Epa1A and variants. Epa1→6A binding profiles could be compared with binding profiles for Epa6 found in literature[41]. The comparison showed that the specificity had been successfully shifted in the Epa1→6A variant, increasing the reliability of the results for the other variants.

### 7.2.1 Comparison of the A domain from *Epa adhesins* with *S. cerevisiae Flo5 flocculin*

The Epa1A domain had been identified as a PA14 domain by bioinformatic analyses[64], but sequence identity was very low when compared to other, structurally characterized PA14 containing proteins, like the recently published Flo5A domain structure[1], or the name giving protective antigen from *B. anthracis*[141]. It was especially surprising, therefore, that the structure of Epa1A could be solved by molecular replacement using a highly simplified model based on Flo5A, as described in sections 5.5.1 and 6.1.1.

The overall structure of Epa1A is quite similar to that of Flo5A, with an RMSD value of 1.34 Å over 122 common C $\alpha$  atoms after aligning with the “*super*” command under pymol (Figure 7.5 a)). The entire  $\beta$ -sandwich is very well conserved, while the loops surrounding it are not, and vary strongly in length between the two proteins. Epa1A is lacking a Flo5-like subdomain from which several interactions with the  $\alpha$ -1,2-mannobiose arise in Flo5A. This typical Flo5A like sub-structure had been predicted to be missing from Epa1A, as can be seen in the alignment on Fig. 6.2.

Flo5A presents 4 disulfide bridges, two of which can also be found in Epa1A. These two cysteins are formed by cysteins 49 and 178, and 179 and 262, respectively, and bind the C- and N-terminal regions together (Fig. 7.5 b and c). The result is a closed structure, where the C- and N-termini are grouped on one of the extremes, where the putative link with the B-domain of the full length adhesin should be. The other two disulfide bridges of Flo5A are present within the subdomain, which is missing in Epa1A (blue region in Figure 7.5 a). Epa1A presents one disulfide bridge between C77 and C118 which is not found within Flo5A. This disulfide bridge could be of the utmost importance, as it covalently connects loop L1 and L2 (Figure 7.5 a), resulting in a firm outer shell for the binding pocket.

Loops L1, L2 and L3 are much more pronounced in Epa1A than in Flo5A and, although L2 is partially replaced by the subdomain, the Flo5A binding pocket is left much more exposed to solvent interaction (Figure 7.5 a). While both the CBL1 and the CBL2 loops are extremely well conserved in structure and position, there is no significant sequence homology in CBL2.

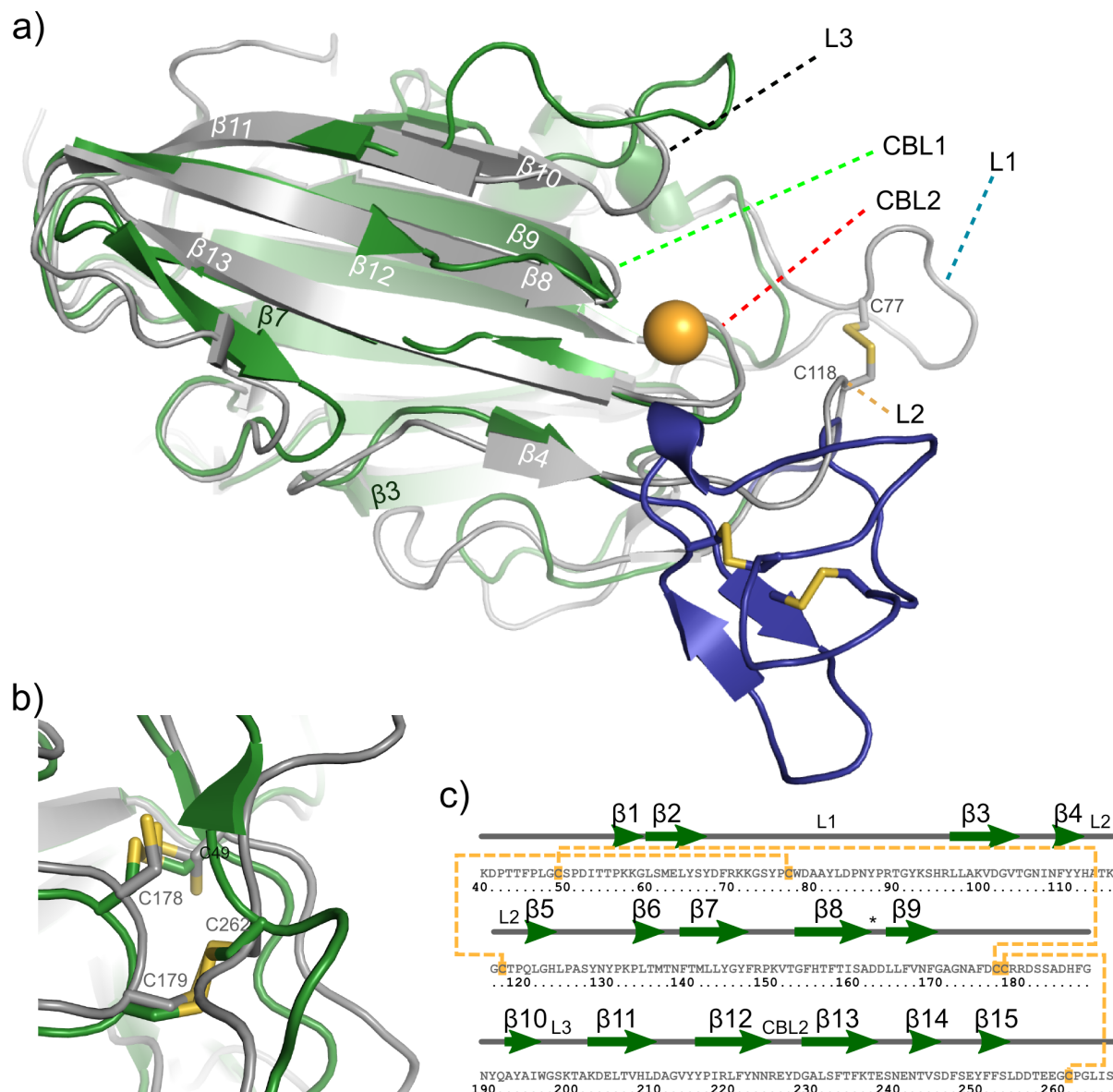


Figure 7.5 Comparison between *Epa1A* (grey) and *Flo5A* (green, subdomain in blue). a) Whole structure alignment performed with the super command in pymol (RMSD 1.34Å over 122 common Ca atoms). Numbering of the  $\beta$ -sheets is relative to *Epa1A* (Fig. 6.5). Loops L1 and L2 are covalently bound by a disulfide bridge which cannot be found in *Flo5A*. b) Detail of the other two, highly conserved, disulfide bridges in *Epa1A*. c) Schematic view of *Epa1A* secondary structure and disulfide connectivity.

When compared to the prototypical PA14 domain[64], the superposition performed with “super” is much worse than for *Flo5A*, with an RMSD of 2.642 Å over only 71 common Ca atoms (Fig. 7.6). The divergence is again mainly in the loop regions, with L1, L2 and L3 loops not having counterparts in the PA14 domain. The binding pocket is completely disrupted in the protective antigen, as no calcium ion is bound, and there is no *DcisD* motif on CBL1, with one of the

aspartates substituted for an asparagine and a trans-peptide bond. CBL2 does not align well, either, as it is prolonged over  $\beta 4$  before connecting to a shortened  $\beta$ -strand 13 equivalent (Fig 7.6).

The comparison with the protective antigen domain shows that, even with very few common sequential components, a  $\beta$ -sandwich motif can be built. The comparison with Flo5A is more interesting though, as it not only explains the surprising results that led to the solution of the structure by molecular replacement, but it also sheds light on the structural basis for the differences in specificity and affinity towards carbohydrates between the two proteins[1][155].

### 7.2.2 The binding pocket of *Epa1A*

Despite significant similarities in general fold, *Epa1A* appears to have quite a distinct carbohydrate binding mode from Flo5A. As could be shown previously[259], *Epa1A* binds calcium cations at trace concentrations, and does not need any calcium beyond the impurities present in the buffer to bind carbohydrates, whereas Flo5A needs calcium added to the buffer to be active[192]. On the other hand, *Epa1A* will not crystallize in the absence of a carbohydrate, while crystallization of Flo5A could be achieved without both glycan and calcium[1]. Finally, the binding of *Epa1A* to the T-antigen was characterized to be in the low micromolar range, while that of Flo5A to  $\alpha 1,2$  mannanose is within the millimolar range[1].

With the crystal structure of *Epa1A*, it was possible to define the binding pocket, and to explain the functional differences in terms of structural features. While the binding pocket of Flo5A is highly accessible to bulk solvent, the binding pocket of *Epa1A* is a less exposed (Fig. 7.7 a).

The binding of the calcium atom is very similar between the two proteins, with the *DcisD* motive on CBL1, the peptide backbone of CBL2 and an asparagine situated at the N-terminus of CBL2 as the

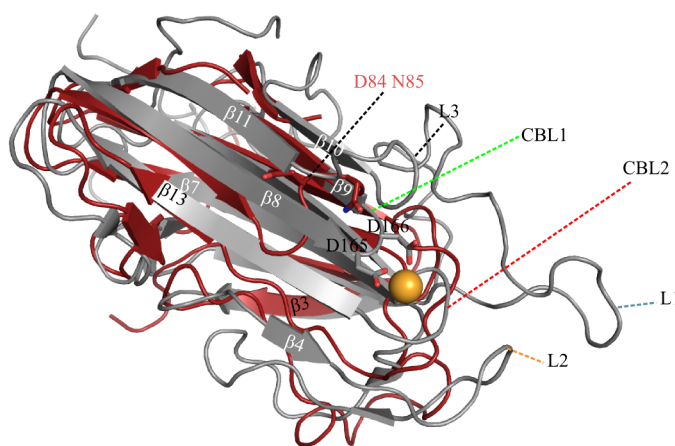


Figure 7.6: Structural alignment between *Epa1A* (grey) and the PA14 domain of the *B. anthracis* protective antigen. While the core  $\beta$ -sandwich structure is well conserved, the loops surrounding it are not. Of special interest is the fact that the entire binding pocket and associated loops (CBL1, CBL2, L1, L2 and L3) are missing or in a completely different conformation. CBL1 in particular does not present a *D-cis-D* motive anymore, with D84 and N85 in a *trans* conformation.



main coordination partners. All three features are present in both proteins, and that region is almost super-imposable between the two (Figure 7.7 a). Out of the 4 amino acids composing the CBL2 loop in Flo5A, one is polar (S227), one is aromatic but does not interact with the glycan (W228) and the other two are aliphatic (A225 and V226). In contrast, Epa1A presents three charged amino acids (R226, E227 and D229) and a tyrosine which can perform a  $\pi$ -C-H interaction [260], and hydrogen bonds (Figure 7.7 b and c). Furthermore, W198 from loop L3 can interact with the galactose moiety from the top of the binding pocket. The effects of the L3 interaction were further supported by calculating the interaction surface between Epa1A and the galactose with the PDBePISA server[261]. While the calculated surface for the Epa1A-galactose interaction was 160.1 Å<sup>2</sup>, the one between Flo5A and the corresponding mannose was only 135.8 Å<sup>2</sup>.

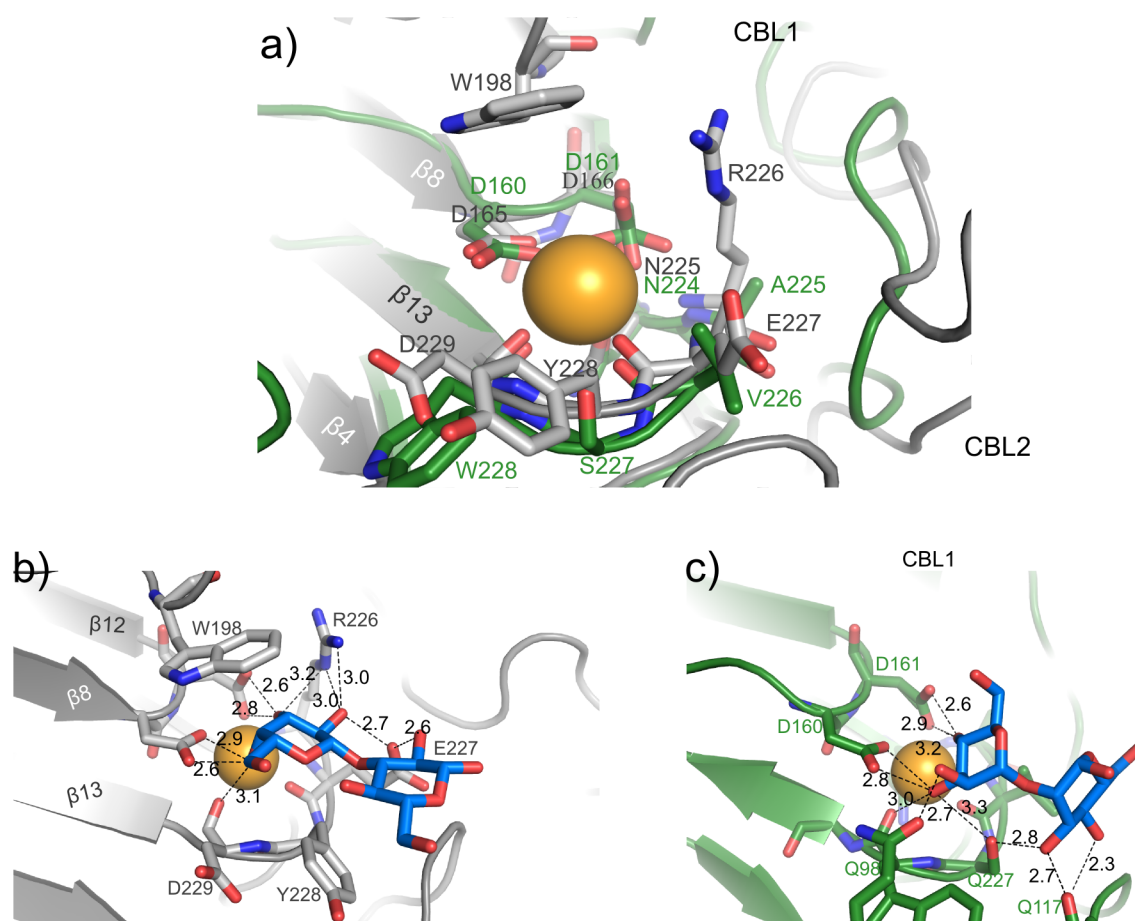


Figure 7.7: The binding pockets of Epa1A (grey) and Flo5A (green). a) Superposition of the binding pockets. Amino acid and  $\beta$ -strand positions in grey/white correspond to Epa1A, in green to Flo5A. b) Detail of the sugar binding for Epa1A with distances between putative hydrogen bond interaction partners. c) Detail on sugar binding for Flo5A with distances between putative hydrogen bond interaction partners.

The interaction of the respective glycans with the pocket is, in both cases, much tighter for the



non-reducing end than the reducing one (Fig. 7.7 a and b). In Flo5A's case interactions exist mainly between the *DcisD* motive, the calcium cation and the equatorial hydroxyl groups on C2 and C3 of the non-reducing mannose (Figure 7.7 c). On the other hand, the presence of W198 and R226 in Epa1A forces the galactose moiety to be in the opposite orientation, with the equatorial O2 interacting with R226 at the N-terminus of CBL2 and D166 on CBL1. It is with O3 and O4 that most of the interactions between CBL1 aspartates and the calcium cation (Figure 7.7 b)) take place.

In contrast to the tight interaction of the non-reducing moiety, the reducing moiety is bound in a more relaxed manner, which is quite similar between the two proteins. PDBePISA results show that both Flo5A and Epa1A have very similar interaction surfaces with the corresponding saccharides, with 113.2 Å<sup>2</sup> for the Flo5A-mannose interface and 109.3 Å<sup>2</sup> for the Epa1A-glucose interface. In this case, the hydrogen bond between E227 on CBL2 and the O2 of the glucose seems to be crucial in stabilizing the binding of the glucose moiety.

### 7.2.3 *A structural model for carbohydrate binding in Epa1A*

As described previously, loops L1, L2 and L3 act as a secondary compartment for the binding pocket, shielding it further from solvent. From the three, it is L3 which probably shows the biggest effect, as W198 interacts directly with the galactose ring moiety (Fig. 7.8). As a result, Epa1A's occluded pocket can stabilize the protein-oligosaccharide interaction much better by numerous vdW interactions than Flo5A, thus yielding higher affinities. Interestingly, the L3 tripeptide S<sub>199</sub>K<sub>200</sub>T<sub>201</sub> proved to be difficult to model during the structure refinement process, with an average B-factor for the region of 26.3 Å<sup>2</sup>, and the side chain of Lys200 could not be modeled at all (in comparison, Trp198 has an average B-factor of 13.2 Å<sup>2</sup>). It is conceivable that, in the absence of ligand, loop L3 will recede from its position directly on top of the galactose moiety, leaving a large gap through which the ligand might enter the binding pocket (Fig. 7.8). The entirety of L3 will then probably strongly shift back and forth, as was the case with the S<sub>199</sub>K<sub>200</sub>T<sub>201</sub> tripeptide in the complex structure, as the lack of interaction with the ligand will liberate the loop. As a result, the entire top region of the binding pocket will become much more flexible, hindering crystallization.

The shifting of L3 could act as a gate to the binding pocket, and explain both the stronger affinity of Epa1A, and its inability to crystallize in the absence of carbohydrates.

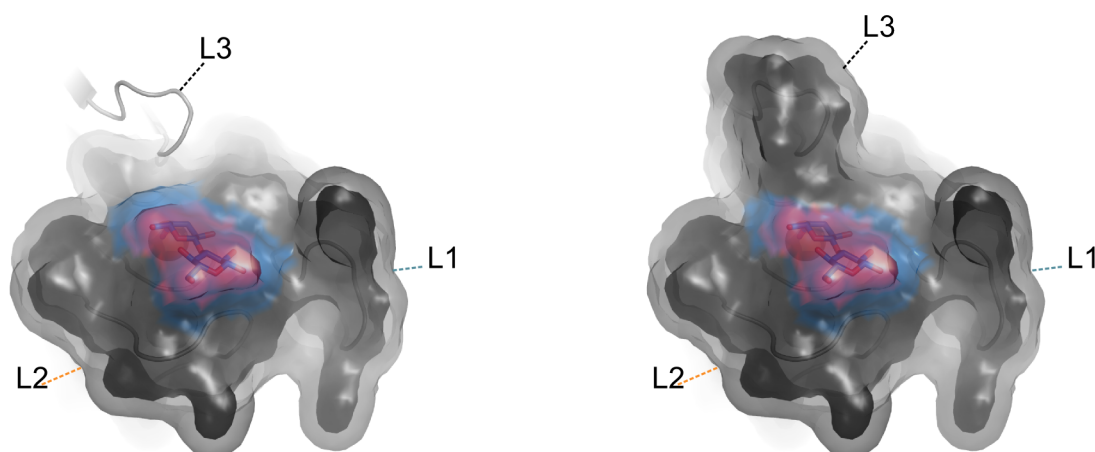


Figure 7.8: Surface view of the binding pocket, without (left) and with (right) the corresponding area of the L3 loop missing in Flo5A. Van der Waals surface of the protein is shown in black, while solvent accessible surface is shown in transparent grey. Correspondingly, van der Waals surface of the glycan is shown in red, solvent accessible surface in blue. While L1 and L2 will probably remain quite constant in their topology, mainly due to the disulfide bond between both (Fig. 6.5), L3 could show high flexibility when no ligand is bound. If L3 would flip upwards, removing the Trp-galactose interaction, the entirety of the galactose ring moiety would be exposed to the solvent (left), while with the interaction, almost the entire ring is covered by L3.

#### 7.2.4 The specificity of Epa1A and its variants.

Originally, epithelial adhesins from *C. glabrata* were characterized as binding lactosides[47], but more recent studies have challenged it[41]. The structural models of sections 6.1.1 and 6.1.6 present definite proof that members of the Epa1 subclass do not share a high affinity for lactose, but instead prefer  $\beta$ 1-3 disaccharides

From the binding results presented in section 6.1.7, it is clear that Epa1A binds almost exclusively galactose in the sub-pocket comprising the calcium, CBL1 and W198 (Fig. 6.18 a, b and c), while promiscuity within the second pocket is much higher (Fig. 6.19), accepting glucose, N-acetylglucosamine, galactose and N-acetyl-galactosamine. The preferred glycosidic bond between the two was  $\beta$ 1-3, and especially at limiting concentrations no  $\beta$ 1-4 substrates were bound.

When compared with the available literature[41], these data correlate well, as the measurements at a limiting concentration of  $1\mu\text{g/ml}$  (Fig. 6.15 a), highly increased the resolution of the assay, showing that the T-antigen (Galactose $\beta$ 1-3N-Acetyl-galactosamine) was the preferred binder of Epa1A. The -antigen is also called the core 1 mucin-type O-glycan[11] (Figure 2.6). This result is in itself quite remarkable, as it directly links *C. glabrata* epithelial adhesins to the cores of mucin-type O-glycans, which constitute the basis for the most common oligosaccharides present on mucins[40]. Therefore, the mere expression of Epa1p on the surface of *C. glabrata* should cause it to adhere to

mucosas and epithelia, explaining the observed clinical behavior[234].

In order to investigate Epa specificity and avidity further, the Epa family was divided into evolutionary subclasses (Fig. 6.7), from which homology models were derived (Fig. 6.8). The selected epithelial adhesins were chosen regarding their respective A domain sequence identity with Epa1A as a primary criterion. Sequence identity had to be at least 45% for an Epa to be accepted as a suitable candidate for the project, which significantly increased the reliability of the models. It was assumed, that any variation within the sequence corresponding to the  $\beta$ -sandwich structure would not be very significant and that differences in the specificity and avidity would be mainly centered around CBL1 and CBL2.. The reason for this assumption was that the structural comparison of Epa1A and Flo5A (Fig. 7.5), presented a structurally highly conserved  $\beta$ -sandwich, but variable loops.

While CBL1 was very strongly conserved within all tested Epa family members, CBL2 showed great variation, and three subclasses could be defined (Fig. 6.8). Variants Epa1 $\rightarrow$ 2A, Epa1 $\rightarrow$ 3A and Epa1 $\rightarrow$ 6A could be then generated by mutating the REYD CBL2 tetra-peptide of Epa1A to the corresponding tetra-peptide present in each of the subclass members. This tetra-peptide was part of the specificity determining penta-peptide presented by Zupancic *et. al.* in 2008[41].

The Epa1 $\rightarrow$ 6A variant results could be compared with literature as Epa6p had already been functionally characterized before, along with Epa1p and Epa7p[41]. A correlation analysis with the results presented by Zupancic *et. al.* In 2008 [41] yielded a Pearson R<sup>2</sup> value of 54.6% for all common glycans with a confidence interval of 95%, showing that the Epa1 $\rightarrow$ 6A conversion was successful. The CFG array binding profiles are very similar between Epa1 $\rightarrow$ 6A and Epa6p (Fig. 7.8). Both of them share a highly increased promiscuity in their binding, not showing the marked preference for galactose $\beta$ 1- joined disaccharides that Epa1A presents, but indiscriminately binding galactose $\beta$ 1-3,  $\beta$ 1-4 and  $\alpha$  joined glycans (Fig. 6.15 b). Interestingly, the strong preference for galactose as the first glycan is maintained, as the primary binding pocket, formed by CBL1, the calcium cation and W198 from L3 are predicted to remain unchanged.

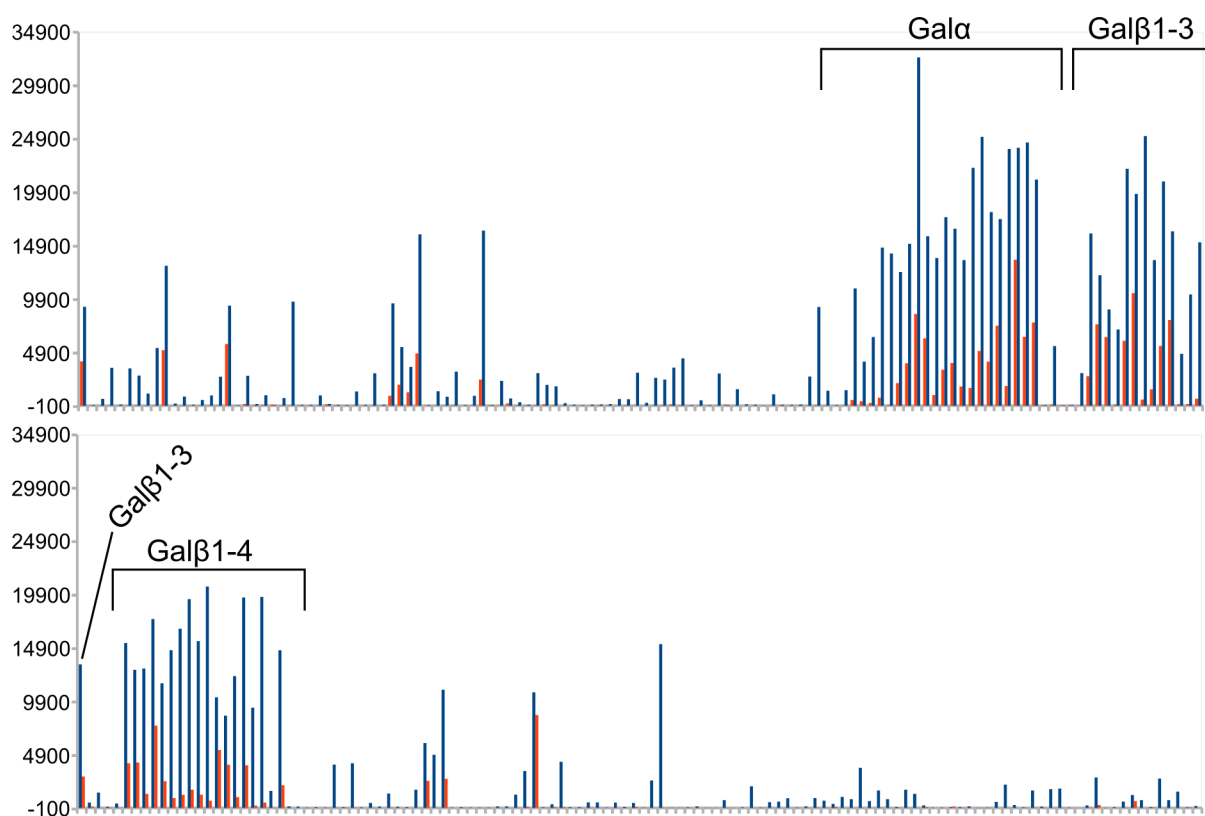
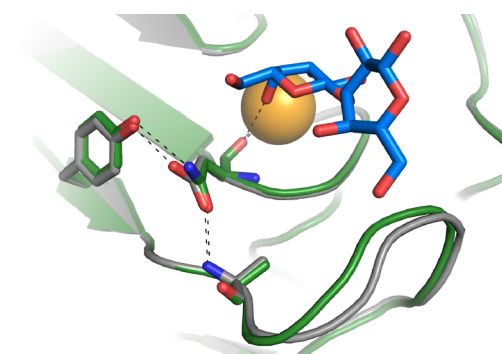


Figure 7.9: Glycan array profiles for *S. cerevisiae* presenting Epa6 from Zupancic *et. al.* 2008[41] (blue) and Epa1→6A (orange). Maximum binding is clustered around the same types of glycans, namely galactose linked, and galβ1-4 or galβ1-3 joined oligosaccharides, without much differences in preference. In order to superimposed both profiles, only common oligosaccharides were presented. Correlation analysis for common glycans yielded a Pearson coefficient of 0.74 and an  $R^2$  Pearson of 54% for a confidence interval of 95% and a two tailed probability of less than 0.01%

The reasons for such remarkable behavior could be easily explained after the variant had been structurally characterized in complex with a galactose-β1-3glucose disaccharide. The suspicions that the galactose was solely bound by the primary binding pocket could be corroborated, and reasons for the increased promiscuity of the second binding pocket, where CBL2 is the main actor, became immediately evident. The Epa1A tetra-peptide REYD had been shifted to RDND (Fig. 6.8 d)). The resulting pocket was much wider and more flexible (Fig. 6.13 a)), and the reducing sugar of the disaccharide could rotate in it. The carbohydrate appeared in two major conformations, one similar to the one seen in Epa1A (Fig. 7.7 b)), where both pyranose moieties are on the same plane, and another in which the glucose had rotated around 90° along the β1-3 axis. It is to be expected that in such a pocket more sterically challenging substrates could be accommodated, for example with larger substituents or even α-bound disaccharides

After the validity of the method could be addressed over the Epa1→6A variant, the other variants were structurally and functionally characterized. The Epa1→2A variant showed a marked preference for sulfated and  $\alpha$  joined glycans, albeit with a much lower general affinity for any glycan than either Epa1A or Epa1→6A. Sulfated carbohydrates are common on a wide variety of mucins[7], and have been pointed at as major targets in the invasion of the oral cavity by different microorganisms[29].

The only change within CBL2 between the Epa2 subclass and that of Epa6 is the substitution of aspartate 229 for asparagine. The resulting reduction of negative charges within the binding pocket could explain the relative preference for sulfated sugars. This difference in behavior is still puzzling, though, since asparagine 229 does not interact directly with the carbohydrate over the side chain. There are also no major differences in the hydrogen bonding network and conformation between it and its equivalent aspartate in either Epa1A or Epa1→6A (Fig. 7.10). Finally, although this single amino-acid change appears to promote the perpendicular conformation of



*Figure 7.10: Hydrogen bond network of D/N229. D229 from Epa1A (grey) shows almost the same behavior in orientation and hydrogen bonding as N229 from Epa1→2A (green) which makes it very difficult to understand how functional differences can arise by such a mutation.*

the glycan (Fig. 6.3 b), this might be misleading, as the resolution for the Epa1→2A dataset was around 0.8 Å worse than that of Epa1→6A. It seems more feasible to think that the binding will be similar to that of Epa1→6A, with a relatively free, rotating glucose which will have both the planar and the perpendicular conformations as preferred states. Epa1→2A was also by far the most aggregation prone of all variants, and also the most difficult to crystallize. How the D229N substitution might have caused this is again challenging to explain, and might need further experiments, especially of a co-crystal structure with a sulfate glycan, to be fully comprehended.

The changes performed on CBL2 to shift Epa1A to the Epa3 subclass and generate the Epa1→3A variant were the most profound, as it involved changing the tetra-peptide from REYD to IGKD. The change of R226 to isoleucine essentially impeded the only interaction between CBL2 and galactose, possibly destabilizing even the primary binding pocket. Also, while the rest of the variants have a CBL2 loop going from positive to negative in an N- to C-terminal fashion, the Epa1→3A CBL2 loop behaves completely differently. A hydrophobic N-terminal portion is

followed by a positive C-terminal region in the Epa3 sub-class. Accordingly, no strong affinity can be found for any carbohydrate, but Epa1→3A still appears to follow similar preference patterns as Epa1→2A (Fig. 6.18), albeit with an even lower affinity. The crystal structure of Epa1→3A presents an even less defined disaccharide as in either of the other two variants. Here the galactose is even less defined, which is probably the result of the loss of the R226-galactose interaction. For Epa1→3A it is only possible to determine that galactose is within the primary binding pocket, and that there is a second monomer within the secondary binding pocket. It is very difficult to assess the nature of the second carbohydrate, and the type of bonding.

All three variants were very different in their specificities to the original Epa1A, and those specificities could be partially characterized. The CBL2-dependent specificity of the different Epa subclasses ties in well with Epa differential expression, pointing at a specialization within Epa family. Specialized Epa subclasses that are activated according to the situation have been proposed as a very important colonization mechanism[56].

Expression of Epa6p has been shown to be activated in the presence of topic antimycotics used in vulvovaginitis treatment[27] or during urethral infection as a response to limiting nicotinic acid concentrations[26]. Epa6 and 7 have also been linked to biofilm formation on hydrophobic surfaces, while Epa1 to Epa5 did not contribute to it[66]. High substrate promiscuity could explain why Epa6 is immediately expressed in any of these stressful situations, as the adhesin might act as an emergency system, increasing adhesion to the substrate and protection through biofilm formation. Targeting Epa6 mediated emergency adhesion might be an interesting alternative or complementary therapy against candidiasis. A structure based explanation for Epa6 promiscuity was provided here, perhaps allowing the development of structure based anti-adhesive, antimycotic drugs against this protein.

The Epa2 subclass needs to be further studied, but it could be significant in the colonization of the oral or intestinal cavity, where sulfo-mucins are specially abundant, or it could be related to host evasion, as sulfo-mucins have been linked to lymphocytes and the immune system[7]. Epa3 has no reported function in literature, and from the characteristics presented in this work, it is difficult to predict a functionality for it.

### 7.2.5 *Epa A domains do not bind lactose*

As previously mentioned, Epa1 was originally thought to bind lactose[47], but recent studies have shown otherwise[41]. The CFG binding profiles show clearly that lactose is by no means the preferred binder of either Epa1A or its variants (Fig 7.11), but that it is bound ~13 worse than lactose for an Epa1A concentration of 200  $\mu\text{g/ml}$ .

The dissociation constants ( $K_d$ ) presented in this work clearly show that Epa1A binds the T-antigen at least ~20 better than lactose. These values are in accordance with the CFG measurements for 200  $\mu\text{g/ml}$ , but not for the measurements performed at more limiting protein concentrations, where affinities are up to ~50 times better for the T-antigen than for lactose (Fig. 7.11). This discrepancy could be explained with the solution of the crystal structures of Epa1A and Epa1 $\rightarrow$ 6A.

The crystals employed to obtain the datasets for these structures had been grown in the presence of 50 mM lactose, so it was surprising to find no lactose (galactose- $\beta$ 1-4glucose) but a galactose- $\beta$ 1-3 disaccharide within the binding pocket, which was modeled as galactose- $\beta$ 1-3glucose. The

explanation for such an occurrence could be the fact that the lactose employed had not been chemically synthesized, but isolated from milk, and had a 98% purity. It has been shown that bovine mucl is present within fat glomerules in milk[262], and that human milk contains several carbohydrates which are normally

targeted by microbial lectins[2]. The function of such carbohydrates has been proposed to be the protection of the intestinal microbiota. It seems very much possible that the contaminating mucin-type O- or N- glycans present in milk, and therefore present as impurities in the lactose isolate, would actually displace lactose from the binding pocket and bind instead.

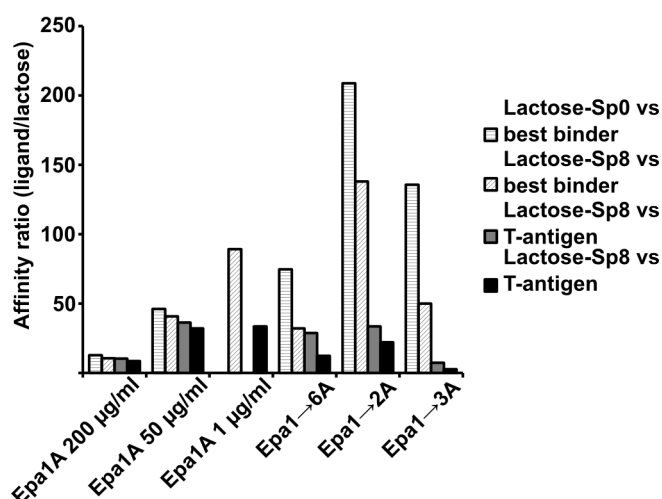


Figure 7.11: Relative decrease in affinity between best binder and the T-antigen versus lactose on the CFG V4.1 glycan array. Values were calculated by dividing either the best binder or the T-antigen for each concentration and or variant by the corresponding intensity of the lactose. While affinities The best binders for Epa1A and Epa1 $\rightarrow$ 6A are comparable to the T-antigen, Epa1 $\rightarrow$ 2A and Epa1 $\rightarrow$ 3A show a clearly different preference.

The implication of this theory, of course, is that the dissociation constants ( $K_d$ ) presented by Kalugin [259], Eulenburg [155] and the present work for lactose would be wrong, as they were measured against a mixture of low affinity lactose and high affinity contaminant. The resulting  $K_d$  would therefore be an “average”  $K_d$  of the two. It can be expected that the  $K_d$  for Epa1A and the contaminant will not be in the micromolar range, but in a range at least 13 times lower, i.e. the nanomolar range. These assumptions are further supported by fluorescence titration results which present the T-antigen as being a ~20 times better binder than the contaminated lactose.

### 7.2.6 Outlook

The structural and functional characterization of Epa1A and variants yielded a profound knowledge on the binding mechanics of the *C. glabrata* epithelial adhesins with unprecedented detail. Results corroborate that, at least Epa1, has a high affinity for  $\beta$ 1-3 sugars, while Epa6 is much more promiscuous[41]. It could also be shown that there is a definite and different specificity for each of the Epa family members, which coupled with their differential expression seems to point to a specialization of epithelial adhesins for different mucins or mucosas[56].

Even with such exceptional results, much remains unclear in the field, and the present work sets the basis to proceed in a structurally relevant way. In order to fully characterize the binding pocket, it will be necessary to obtain affinity constants for a wider variety of ligands. Chemically synthesized, pure lactose will be absolutely necessary to clarify the affinity towards this disaccharide. The T-antigen showed very promising results, but several other mucin-type O-glycan cores are also good binding candidates, and therefore quantitative titrations will be performed against them. Type 1 and type 2 N-lactosamine branching disaccharides characteristic of the branching of O- and N- glycans[40] are very similar to the disaccharide found in the structures, and will be of primary interest. Also, obtaining the corresponding complex structures would result in an even more detailed knowledge of the pocket.

On first sight, it appears that the contaminated lactose might explain the non-linear Scatchard plots which resulted from some of the fluorescence titrations (section 6.1.8). It has been conclusively demonstrated, though, that ligand-heterogeneity never mimics negative cooperativity or multiple classes of binding sites[263]. A second, low affinity binding site, as with Flo5A[1], would nicely explain the Scatchard plots, but no such secondary site could be found within the crystal structures. In order to test this hypothesis, T-antigen co-crystals will be obtained, as they will show whether, at high concentrations of a high affinity ligand, some other pocket is to be occupied. Protein instability



at low concentrations might be a different explanation, as aggregates would still bind the ligand, albeit at a much lower affinity than the free protein, resulting in two different binding site classes. Stability problems at low concentrations could be solved by performing the fluorescence titrations with a mixture of Epa1A and BSA, which would raise the protein concentration without interacting with the titration itself.

Finally, even with the same CBL2, Epa1p and Epa7p have been shown to have slightly different specificities[41], which points at long range effects, as has previously been observed for Flo5[1]. Loops L1 and L2 might be involved in such long range effects. It would be most interesting to characterize these more subtle differences either through the production of Epa7A or by shifting the L1 and L2 loops of Epa1A towards that of Epa7p. Also, the Epa adhesin palette from *C. glabrata* is much wider than the ones presented here. The other members of the Epa family were not taken into account during this study because there were larger differences with Epa1A, and therefore it was deemed implausible that simple mutagenesis of Epa1A would yield reliable information about these other proteins. It would still be central to the understanding of *C. glabrata's* binding and host-evasion abilities to study these other Epa members. For such an undertaking, it would be necessary to clone and produce the different gene products directly, an effort which exceeded the scope of the present work.

## 8 Acknowledgements

First and foremost, I would like to thank Professor Lars-Oliver Essen, who offered me the fantastic opportunity of working in structural biology through the study of microbial adhesins. His insight into the intricacies of biochemistry, and more specifically, structural biology left me in stunned realization in more than one occasion. I am certain that I am now not only a much better scientist, but a much more reliable and dependable person through his direct influence.

I would also like to thank Professor Hans-Ulrich Mösch for kindly taking up the secondary supervision of this work. Many thanks for the discussions and collaboration, along with Ms. Rike Diderrich, on the Epa subject.

Mr Maik Veelders' priceless services must be mentioned at this point. Without him, his competence, and his mastery of matters he claims not to have a clue about, most of this work would not have been possible. As a crystallographer, a scientist, a corrector and a friend he has been absolutely priceless.

Drs Georgios Psakis and Louisa Tsougaraki, his lovely wife, spent long hours correcting my convoluted English, supporting and discussing critical issues within and without the scientific spheres.

Mr Wolfgang Große merits a brief but dashing appearance in these acknowledgements as the person who managed to get me out of my hole at the back of the lab during one of the inevitable downs of research. Without his constant and unyielding optimism, I might still be a doctor, but a much less happy one.

Mr Vitali Kalugin and Mr Hannes Lübden worked for different projects under my tutelage. I would like to thank them for the great results they produced, and for their respect and the great atmosphere in the lab during their stay. And for stories about nurses in the cinema.

The acknowledgements wouldn't be complete without Mrs Petra Gnau, who has always been down to earth, responsible and great fun!

To all the people in the lab; the lab, diploma and Ph.D. students, go my most heartfelt thanks for having provided for such a great time.

To friends, old and new, in Spain, Germany, Japan, Taiwan and everywhere else they might be; they should feel acknowledged at this moment, wherever they are. A special mention I would like to

## Acknowledgements

make for Juan Prats, who will not be able to read this and should have been.

Yu-Chen Hung, my already (and ever) lovely wife, also needs a mention here, as I would face terrible punishment and doom if not. Thank you very much for all these years, for your companionship and love. I don't know what would become of me without you.

Bacchus and Tethys, a beacon of platonic love, should also be thanked. Especially Bacchus, for such wonderful corrections in my Ph.D. thesis as gaaaaaaaaaaaaaaaaaagggggggggggllllllllllhhhhh.

I would like to thank my parents last, their unconditional support, love and intelligence have always been an inspiration to me, even if I do not show it all the time. Hopefully, the foundation for the promotion and protection of the Manuel Maestre Reyna will start bearing fruits soon. Ok, and the little idiot can feel himself acknowledged, as well.

*Īa Shub-Niggurath!*

## Bibliography

1. Veelders, M., Brückner, S., Ott, D., Unverzagt, C., Mösch, H. & Essen, L. (2010). *Structural basis of flocculin-mediated social behavior in yeast*, Proc Natl Acad Sci U S A. **107**, 52, 22511-22516
2. Sharon, N. & Ofek, I. (2000). *Safe as mother's milk: carbohydrates as future anti-adhesion drugs for bacterial diseases*, Glycoconj J. **17**, 7-9, 659-664
3. Hatakeyama, M. & Brzozowski, T. (2006). *Pathogenesis of Helicobacter pylori infection*, Helicobacter. **11**, Suppl 1, 14-20
4. Atherton, J.C. (2006). *The pathogenesis of Helicobacter pylori-induced gastro-duodenal diseases*, Annu Rev Pathol. **1**, 1, 63-96
5. Henderson, I.R., Navarro-Garcia, F., Desvaux, M., Fernandez, R.C. & Ala'Aldeen, D. (2004). *Type V protein secretion pathway: the autotransporter story*, Microbiol Mol Biol Rev. **68**, 4, 692-744
6. Hattrup, C.L. & Gendler, S.J. (2008). *Structure and function of the cell surface (tethered) mucins*, Annu Rev Physiol. **70**, 431-457
7. Nieuw Amerongen, A.V., Bolscher, J.G., Bloemena, E. & Veerman, E.C. (1998). *Sulfomucins in the human body*, Biol Chem. **379**, 1, 1-18
8. Filler, S.G. (2006). *Candida-host cell receptor-ligand interactions*, Curr Opin Microbiol. **9**, 4, 333-339
9. Nordman, H., Borén, T., Davies, J.R., Engstrand, L. & Carlstedt, I. (1999). *pH-dependent binding of Helicobacter pylori to pig gastric mucins*, FEMS Immunol Med Microbiol. **24**, 2, 175-181
10. Sheehan, J.K., Kesimer, M. & Pickles, R. (2006). *Innate immunity and mucus structure and function*, Novartis Found Symp. **279**, 155-66; discussion 167-9, 216-9
11. McEntyre, J., Kornfeld, S., Varki, A., Sharon, N., Bertozzi, C., Rabuka, D., Esko, J., Colley, K., Freeze, H., Elbein, A. et al. (2009) Essentials of Glycobiology, 2nd edition
12. Knowles, M.R. & Boucher, R.C. (2002). *Mucus clearance as a primary innate defense mechanism for mammalian airways*, J Clin Invest. **109**, 5, 571-577
13. Gendler, S.J. (2001). *MUC1, the renaissance molecule*, J Mammary Gland Biol Neoplasia. **6**, 3, 339-353
14. Schroeder, J.A., Thompson, M.C., Gardner, M.M. & Gendler, S.J. (2001). *Transgenic MUC1 interacts with epidermal growth factor receptor and correlates with mitogen-activated protein kinase activation in the mouse mammary gland*, J Biol Chem. **276**, 16, 13057-13064
15. Lagow, E., DeSouza, M.M. & Carson, D.D. (1999). *Mammalian reproductive tract mucins*, Hum Reprod Update. **5**, 4, 280-292
16. Bäckström, M., Link, T., Olson, F.J., Karlsson, H., Graham, R., Picco, G., Burchell, J., Taylor-Papadimitriou, J., Noll, T. & Hansson, G.C. (2003). *Recombinant MUC1 mucin with a breast cancer-like O-glycosylation produced in large amounts in Chinese-hamster ovary cells*, Biochem J. **376**, Pt 3, 677-686
17. Ordoñez, C.L., Khashayar, R., Wong, H.H., Ferrando, R., Wu, R., Hyde, D.M., Hotchkiss, J.A., Zhang, Y., Novikov, A., Dolganov, G. et al. (2001). *Mild and moderate asthma is associated with airway goblet cell hyperplasia and abnormalities in mucin gene expression*, Am J Respir Crit Care Med. **163**, 2, 517-523
18. Gronowitz, E., Pitkänen, S., Kjellmer, I., Heikinheimo, M. & Strandvik, B. (2003). *Association between serum oncofetal antigens CA 19-9 and CA 125 and clinical status in patients with cystic fibrosis*, Acta Paediatr. **92**, 11, 1267-1271
19. Lillehoj, E.P., Kim, B.T. & Kim, K.C. (2002). *Identification of Pseudomonas aeruginosa*

- flagellin as an adhesin for Muc1 mucin*, Am J Physiol Lung Cell Mol Physiol. **282**, 4, L751-6
20. Karlsson, K.A. (2000). *The human gastric colonizer Helicobacter pylori: a challenge for host-parasite glycobiology*, Glycobiology. **10**, 8, 761-771
  21. Venegas, M.F., Navas, E.L., Gaffney, R.A., Duncan, J.L., Anderson, B.E. & Schaeffer, A.J. (1995). *Binding of type 1-piliated Escherichia coli to vaginal mucus*, Infect Immun. **63**, 2, 416-422
  22. Russo, C.L., Spurr-Michaud, S., Tisdale, A., Pudney, J., Anderson, D. & Gipson, I.K. (2006). *Mucin gene expression in human male urogenital tract epithelia*, Hum Reprod. **21**, 11, 2783-2793
  23. Bartman, A.E., Buisine, M.P., Aubert, J.P., Niehans, G.A., Toribara, N.W., Kim, Y.S., Kelly, E.J., Crabtree, J.E. & Ho, S.B. (1998). *The MUC6 secretory mucin gene is expressed in a wide variety of epithelial tissues*, J Pathol. **186**, 4, 398-405
  24. Gipson, I.K., Ho, S.B., Spurr-Michaud, S.J., Tisdale, A.S., Zhan, Q., Torlakovic, E., Pudney, J., Anderson, D.J., Toribara, N.W. & Hill, J.A.3. (1997). *Mucin genes expressed by human female reproductive tract epithelia*, Biol Reprod. **56**, 4, 999-1011
  25. Fukuda, M. (2002). *Roles of mucin-type O-glycans in cell adhesion*, Biochim Biophys Acta. **1573**, 3, 394-405
  26. Domergue, R., Castaño, I., De Las Peñas, A., Zupancic, M., Lockett, V., Hebel, J.R., Johnson, D. & Cormack, B.P. (2005). *Nicotinic acid limitation regulates silencing of Candida adhesins during UTI*, Science. **308**, 5723, 866-870
  27. Mundy, R.D. & Cormack, B. (2009). *Expression of Candida glabrata adhesins after exposure to chemical preservatives*, J Infect Dis. **199**, 12, 1891-1898
  28. Dong, Q., Wang, Q., Xin, Y., Li, N. & Xuan, S. (2009). *Comparative genomics of Helicobacter pylori*, World J Gastroenterol. **15**, 32, 3984-3991
  29. Veerman, E.C., Bank, C.M., Namavar, F., Appelmelk, B.J., Bolscher, J.G. & Nieuw Amerongen, A.V. (1997). *Sulfated glycans on oral mucin as receptors for Helicobacter pylori*, Glycobiology. **7**, 6, 737-743
  30. Kouznetsova, I., Gerlach, K.L., Zahl, C. & Hoffmann, W. (2010). *Expression Analysis of Human Salivary Glands by Laser Microdissection: Differences Between Submandibular and Labial Glands*, Cell Physiol Biochem. **26**, 3, 375-382
  31. Kerschner, J.E. (2007). *Mucin gene expression in human middle ear epithelium*, Laryngoscope. **117**, 9, 1666-1676
  32. Bu, X., Li, N., Tian, X., Li, L., Wang, J., Yu, X. & Huang, P. (2010). *Altered expression of MUC2 and MUC5AC in progression of colorectal carcinoma*, World J Gastroenterol. **16**, 32, 4089-4094
  33. DeSouza, M.M., Surveyor, G.A., Price, R.E., Julian, J., Kardon, R., Zhou, X., Gendler, S., Hilkens, J. & Carson, D.D. (1999). *MUC1/episialin: a critical barrier in the female reproductive tract*, J Reprod Immunol. **45**, 2, 127-158
  34. Carson, D.D., DeSouza, M.M., Kardon, R., Zhou, X., Lagow, E. & Julian, J. (1998). *Mucin expression and function in the female reproductive tract*, Hum Reprod Update. **4**, 5, 459-464
  35. Lindén, S.K., Florin, T.H.J. & McGuckin, M.A. (2008). *Mucin dynamics in intestinal bacterial infection*, PLoS One. **3**, 12, e3952
  36. Schachter, H. & Brockhausen, I. (1992) Glycoconjugates. Composition, Structure and Function. The biosynthesis of serine(threonine)-N-Acetylgalactosamine-linked carbohydrate moieties
  37. Lo-Guidice, J.M., Herz, H., Lamblin, G., Plancke, Y., Roussel, P. & Lhermitte, M. (1997). *Structures of sulfated oligosaccharides isolated from the respiratory mucins of a non-secretor (O, Le(a + b -)) patient suffering from chronic bronchitis*, Glycoconj J. **14**, 1, 113-125
  38. Lo-Guidice, J.M., Merten, M.D., Lamblin, G., Porchet, N., Houvenaghel, M.C., Figarella, C., Roussel, P. & Perini, J.M. (1997). *Mucins secreted by a transformed cell line derived from human tracheal gland cells*, Biochem J. **326 ( Pt 2)**, 431-437

39. Prakobphol, A., Leffler, H. & Fisher, S.J. (1993). *The high-molecular-weight human mucin is the primary salivary carrier of ABH, Le(a), and Le(b) blood group antigens*, Crit Rev Oral Biol Med. **4**, 3-4, 325-333
40. Linden, S.K., Sutton, P., Karlsson, N.G., Korolik, V. & McGuckin, M.A. (2008). *Mucins in the mucosal barrier to infection*, Mucosal Immunol. **1**, 3, 183-197
41. Zupancic, M.L., Frieman, M., Smith, D., Alvarez, R.A., Cummings, R.D. & Cormack, B.P. (2008). *Glycan microarray analysis of Candida glabrata adhesin ligand specificity*, Mol Microbiol. **68**, 3, 547-559
42. Aspholm-Hurtig, M., Dailide, G., Lahmann, M., Kalia, A., Ilver, D., Roche, N., Vikström, S., Sjöström, R., Lindén, S., Bäckström, A. et al. (2004). *Functional adaptation of BabA, the H. pylori ABO blood group antigen binding adhesin*, Science. **305**, 5683, 519-522
43. Kaur, R., Domergue, R., Zupancic, M.L. & Cormack, B.P. (2005). *A yeast by any other name: Candida glabrata and its interaction with the host*, Curr Opin Microbiol. **8**, 4, 378-384
44. Gulia, J., Aryal, S., Saadlla, H. & Shorr, A.F. (2010). *Healthcare-associated candidemia--a distinct entity?*, J Hosp Med. **5**, 5, 298-301
45. Buitrón García-Figueroa, R., Araiza-Santibáñez, J., Basurto-Kuba, E. & Bonifaz-Trujillo, A. (2009). *Candida glabrata: an emergent opportunist in vulvovaginitis*, Cir Cir. **77**, 6, 423-427
46. Pignato, S., Salvo, S., Coniglio, M.A., Marranzano, M., Faro, G. & Giammanco, G. (2009). *Persistent oral and urinary Candida spp. carriage in Italian HIV-seropositive asymptomatic subjects*, J Prev Med Hyg. **50**, 4, 232-235
47. Cormack, B.P., Ghori, N. & Falkow, S. (1999). *An adhesin of the yeast pathogen Candida glabrata mediating adherence to human epithelial cells*, Science. **285**, 5427, 578-582
48. Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E. et al. (2004). *Genome evolution in yeasts*, Nature. **430**, 6995, 35-44
49. Wong, S., Fares, M.A., Zimmermann, W., Butler, G. & Wolfe, K.H. (2003). *Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata*, Genome Biol. **4**, 2, R10
50. Dodgson, A.R., Pujol, C., Pfaller, M.A., Denning, D.W. & Soll, D.R. (2005). *Evidence for recombination in Candida glabrata*, Fungal Genet Biol. **42**, 3, 233-243
51. Castaño, I., Pan, S., Zupancic, M., Hennequin, C., Dujon, B. & Cormack, B.P. (2005). *Telomere length control and transcriptional regulation of subtelomeric adhesins in Candida glabrata*, Mol Microbiol. **55**, 4, 1246-1258
52. Rusche, L.N., Kirchmaier, A.L. & Rine, J. (2003). *The establishment, inheritance, and function of silenced chromatin in Saccharomyces cerevisiae*, Annu Rev Biochem. **72**, , 481-516
53. Tanny, J.C. & Moazed, D. (2001). *Coupling of histone deacetylation to NAD breakdown by the yeast silencing protein Sir2: Evidence for acetyl transfer from substrate to an NAD breakdown product*, Proc Natl Acad Sci U S A. **98**, 2, 415-420
54. Sedighi, M. & Sengupta, A.M. (2007). *Epigenetic chromatin silencing: bistability and front propagation*, Phys Biol. **4**, 4, 246-255
55. Suka, N., Luo, K. & Grunstein, M. (2002). *Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin*, Nat Genet. **32**, 3, 378-383
56. de Groot, P.W.J., Kraneveld, E.A., Yin, Q.Y., Dekker, H.L., Gross, U., Crielaard, W., de Koster, C.G., Bader, O., Klis, F.M. & Weig, M. (2008). *The cell wall of the human pathogen Candida glabrata: differential incorporation of novel adhesin-like wall proteins*, Eukaryot Cell. **7**, 11, 1951-1964
57. Klis, F., Ram, A. & De Groot, P. (2007) A molecular and genomic view of the fungal cell wall
58. Osumi, M. (1998). *The ultrastructure of yeast: cell wall structure and formation*, Micron. **29**, 2-3, 207-233
59. Zlotnik, H., Fernandez, M.P., Bowers, B. & Cabib, E. (1984). *Saccharomyces cerevisiae*

- mannoproteins form an external cell wall layer that determines wall porosity*, J Bacteriol. **159**, 3, 1018-1026
60. Wheeler, R.T. & Fink, G.R. (2006). *A drug-sensitive genetic network masks fungi from the immune system*, PLoS Pathog. **2**, 4, e35
  61. Kaur, R., Ma, B. & Cormack, B.P. (2007). *A family of glycosylphosphatidylinositol-linked aspartyl proteases is required for virulence of Candida glabrata*, Proc Natl Acad Sci U S A. **104**, 18, 7628-7633
  62. Verstrepen, K.J., Reynolds, T.B. & Fink, G.R. (2004). *Origins of variation in the fungal cell surface*, Nat Rev Microbiol. **2**, 7, 533-540
  63. Frieman, M.B., McCaffery, J.M. & Cormack, B.P. (2002). *Modular domain structure in the Candida glabrata adhesin Epa1p, a beta1,6 glucan-cross-linked cell wall protein*, Mol Microbiol. **46**, 2, 479-492
  64. Rigden, D.J., Mello, L.V. & Galperin, M.Y. (2004). *The PA14 domain, a conserved all-beta domain in bacterial toxins, enzymes, adhesins and signaling molecules*, Trends Biochem Sci. **29**, 7, 335-339
  65. Linder, T. & Gustafsson, C.M. (2008). *Molecular phylogenetics of ascomycotal adhesins--a novel family of putative cell-surface adhesive proteins in fission yeasts*, Fungal Genet Biol. **45**, 4, 485-497
  66. Iraqui, I., Garcia-Sanchez, S., Aubert, S., Dromer, F., Ghigo, J., d'Enfert, C. & Janbon, G. (2005). *The Yak1p kinase controls expression of adhesins and biofilm formation in Candida glabrata in a Sir4p-dependent pathway*, Mol Microbiol. **55**, 4, 1259-1271
  67. Fidel, P.L.J., Vazquez, J.A. & Sobel, J.D. (1999). *Candida glabrata: review of epidemiology, pathogenesis, and clinical disease with comparison to C. albicans*, Clin Microbiol Rev. **12**, 1, 80-96
  68. Cassone, A., De Bernardis, F., Mondello, F., Ceddia, T. & Agatensi, L. (1987). *Evidence for a correlation between proteinase secretion and vulvovaginal candidosis*, J Infect Dis. **156**, 5, 777-783
  69. Soll, D.R. (1988). *High-frequency switching in Candida albicans and its relations to vaginal candidiasis*, Am J Obstet Gynecol. **158**, 4, 997-1001
  70. Lachke, S.A., Joly, S., Daniels, K. & Soll, D.R. (2002). *Phenotypic switching and filamentation in Candida glabrata*, Microbiology. **148**, Pt 9, 2661-2674
  71. Csank, C. & Haynes, K. (2000). *Candida glabrata displays pseudohyphal growth*, FEMS Microbiol Lett. **189**, 1, 115-120
  72. Whiteway, M. & Oberholzer, U. (2004). *Candida morphogenesis and host-pathogen interactions*, Curr Opin Microbiol. **7**, 4, 350-357
  73. Sundstrom, P. (1999). *Adhesins in Candida albicans*, Curr Opin Microbiol. **2**, 4, 353-357
  74. Kantarcioglu, A.S. & Yücel, A. (2002). *Phospholipase and protease activities in clinical Candida isolates with reference to the sources of strains*, Mycoses. **45**, 5-6, 160-165
  75. Miyazaki, H., Miyazaki, Y., Geber, A., Parkinson, T., Hitchcock, C., Falconer, D.J., Ward, D.J., Marsden, K. & Bennett, J.E. (1998). *Fluconazole resistance associated with drug efflux and increased transcription of a drug transporter gene, PDH1, in Candida glabrata*, Antimicrob Agents Chemother. **42**, 7, 1695-1701
  76. Marshall, B.J. & Warren, J.R. (1984). *Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration*, Lancet. **1**, 8390, 1311-1315
  77. Blaser, M.J. & Atherton, J.C. (2004). *Helicobacter pylori persistence: biology and disease*, J Clin Invest. **113**, 3, 321-333
  78. Press Release: The 2005 Nobel Prize in Physiology or Medicine (2010), [http://nobelprize.org/nobel\\_prizes/medicine/laureates/2005/press.html](http://nobelprize.org/nobel_prizes/medicine/laureates/2005/press.html)
  79. Clyne, M., Dolan, B. & Reeves, E.P. (2007). *Bacterial factors that mediate colonization of the stomach and virulence of Helicobacter pylori*, FEMS Microbiol Lett. **268**, 2, 135-143

80. Hansson, L.E., Nyrén, O., Hsing, A.W., Bergström, R., Josefsson, S., Chow, W.H., Fraumeni, J.F.J. & Adami, H.O. (1996). *The risk of stomach cancer in patients with gastric or duodenal ulcer disease*, N Engl J Med. **335**, 4, 242-249
81. Morris, A. & Nicholson, G. (1987). *Ingestion of Campylobacter pyloridis causes gastritis and raised fasting gastric pH*, Am J Gastroenterol. **82**, 3, 192-199
82. Graham, D.Y., Alpert, L.C., Smith, J.L. & Yoshimura, H.H. (1988). *Iatrogenic Campylobacter pylori infection is a cause of epidemic achlorhydria*, Am J Gastroenterol. **83**, 9, 974-980
83. Foynes, S., Dorrell, N., Ward, S.J., Stabler, R.A., McColm, A.A., Rycroft, A.N. & Wren, B.W. (2000). *Helicobacter pylori possesses two CheY response regulators and a histidine kinase sensor; CheA, which are essential for chemotaxis and colonization of the gastric mucosa*, Infect Immun. **68**, 4, 2016-2023
84. Wen, Y., Marcus, E.A., Matrubutham, U., Gleeson, M.A., Scott, D.R. & Sachs, G. (2003). *Acid-adaptive genes of Helicobacter pylori*, Infect Immun. **71**, 10, 5921-5939
85. Weeks, D.L., Eskandari, S., Scott, D.R. & Sachs, G. (2000). *A H<sup>+</sup>-gated urea channel: the link between Helicobacter pylori urease and gastric colonization*, Science. **287**, 5452, 482-485
86. Rektorschek, M., Buhmann, A., Weeks, D., Schwan, D., Bensch, K.W., Eskandari, S., Scott, D., Sachs, G. & Melchers, K. (2000). *Acid resistance of Helicobacter pylori depends on the UreI membrane protein and an inner membrane proton barrier*, Mol Microbiol. **36**, 1, 141-152
87. Lindén, S., Mahdavi, J., Hedenbro, J., Borén, T. & Carlstedt, I. (2004). *Effects of pH on Helicobacter pylori binding to human gastric mucins: identification of binding to non-MUC5AC mucins*, Biochem J. **384**, Pt 2, 263-270
88. Testerman, T., McGee, D. & Mobley, H. (2001) *Helicobacter pylori*
89. Yamaoka, Y. & Alm, R. (2008) *Helicobacter pylori: molecular genetics and cellular biology*
90. Lindén, S.K., Wickström, C., Lindell, G., Gilshenan, K. & Carlstedt, I. (2008). *Four modes of adhesion are used during Helicobacter pylori binding to human mucins in the oral and gastric niches*, Helicobacter. **13**, 2, 81-93
91. Odenbreit, S. (2005). *Adherence properties of Helicobacter pylori: impact on pathogenesis and adaptation to the host*, Int J Med Microbiol. **295**, 5, 317-324
92. Torres, J. & Backert, S. (2008). *Pathogenesis of Helicobacter pylori infection*, Helicobacter. **13**, Suppl 1, 13-17
93. McGuckin, M.A., Every, A.L., Skene, C.D., Linden, S.K., Chionh, Y.T., Swierczak, A., McAuley, J., Harbour, S., Kaparakis, M., Ferrero, R. et al. (2007). *Muc1 mucin limits both Helicobacter pylori colonization of the murine gastric mucosa and associated gastritis*, Gastroenterology. **133**, 4, 1210-1218
94. Lindén, S., Mahdavi, J., Semino-Mora, C., Olsen, C., Carlstedt, I., Borén, T. & Dubois, A. (2008). *Role of ABO secretor status in mucosal innate immunity and H. pylori infection*, PLoS Pathog. **4**, 1, e2
95. Censini, S., Lange, C., Xiang, Z., Crabtree, J.E., Ghiara, P., Borodovsky, M., Rappuoli, R. & Covacci, A. (1996). *cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors*, Proc Natl Acad Sci U S A. **93**, 25, 14648-14653
96. Nilsson, C., Sillén, A., Eriksson, L., Strand, M., Enroth, H., Normark, S., Falk, P. & Engstrand, L. (2003). *Correlation between cag pathogenicity island composition and Helicobacter pylori-associated gastroduodenal disease*, Infect Immun. **71**, 11, 6573-6581
97. Odenbreit, S., Püls, J., Sedlmaier, B., Gerland, E., Fischer, W. & Haas, R. (2000). *Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion*, Science. **287**, 5457, 1497-1500
98. Stein, M., Bagnoli, F., Halenbeck, R., Rappuoli, R., Fantl, W.J. & Covacci, A. (2002). *c-Src/Lyn kinases activate Helicobacter pylori CagA through tyrosine phosphorylation of the EPIYA motifs*, Mol Microbiol. **43**, 4, 971-980



99. Meyer-ter-Vehn, T., Covacci, A., Kist, M. & Pahl, H.L. (2000). *Helicobacter pylori* activates mitogen-activated protein kinase cascades and induces expression of the proto-oncogenes *c-fos* and *c-jun*, *J Biol Chem.* **275**, 21, 16064-16072
100. Segal, E.D., Cha, J., Lo, J., Falkow, S. & Tompkins, L.S. (1999). *Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by Helicobacter pylori*, *Proc Natl Acad Sci U S A.* **96**, 25, 14559-14564
101. Churin, Y., Al-Ghoul, L., Kepp, O., Meyer, T.F., Birchmeier, W. & Naumann, M. (2003). *Helicobacter pylori CagA protein targets the c-Met receptor and enhances the motogenic response*, *J Cell Biol.* **161**, 2, 249-255
102. Amieva, M.R., Vogelmann, R., Covacci, A., Tompkins, L.S., Nelson, W.J. & Falkow, S. (2003). *Disruption of the epithelial apical-junctional complex by Helicobacter pylori CagA*, *Science.* **300**, 5624, 1430-1434
103. Higashi, H., Tsutsumi, R., Fujita, A., Yamazaki, S., Asaka, M., Azuma, T. & Hatakeyama, M. (2002). *Biological activity of the Helicobacter pylori virulence factor CagA is determined by variation in the tyrosine phosphorylation sites*, *Proc Natl Acad Sci U S A.* **99**, 22, 14428-14433
104. Ito, Y., Azuma, T., Ito, S., Suto, H., Miyaji, H., Yamazaki, Y., Kohli, Y. & Kuriyama, M. (1998). *Full-length sequence analysis of the vacA gene from cytotoxic and noncytotoxic Helicobacter pylori*, *J Infect Dis.* **178**, 5, 1391-1398
105. Dautin, N. & Bernstein, H.D. (2007). *Protein secretion in gram-negative bacteria via the autotransporter pathway*, *Annu Rev Microbiol.* **61**, 89-112
106. Henderson, I.R. & Nataro, J.P. (2001). *Virulence functions of autotransporter proteins*, *Infect Immun.* **69**, 3, 1231-1243
107. Telford, J.L., Ghiara, P., Dell'Orco, M., Comanducci, M., Burroni, D., Bugnoli, M., Tecce, M.F., Censini, S., Covacci, A., Xiang, Z. et al. (1994). *Gene structure of the Helicobacter pylori cytotoxin and evidence of its key role in gastric disease*, *J Exp Med.* **179**, 5, 1653-1658
108. Yahiro, K., Wada, A., Nakayama, M., Kimura, T., Ogushi, K., Niidome, T., Aoyagi, H., Yoshino, K., Yonezawa, K., Moss, J. et al. (2003). *Protein-tyrosine phosphatase alpha, RPTP alpha, is a Helicobacter pylori VacA receptor*, *J Biol Chem.* **278**, 21, 19183-19189
109. Seto, K., Hayashi-Kuwabara, Y., Yoneta, T., Suda, H. & Tamaki, H. (1998). *Vacuolation induced by cytotoxin from Helicobacter pylori is mediated by the EGF receptor in HeLa cells*, *FEBS Lett.* **431**, 3, 347-350
110. Nakayama, M., Hisatsune, J., Yamasaki, E., Nishi, Y., Wada, A., Kurazono, H., Sap, J., Yahiro, K., Moss, J. & Hirayama, T. (2006). *Clustering of Helicobacter pylori VacA in lipid rafts, mediated by its receptor; receptor-like protein tyrosine phosphatase beta, is required for intoxication in AZ-521 Cells*, *Infect Immun.* **74**, 12, 6571-6580
111. Kuo, C. & Wang, W. (2003). *Binding and internalization of Helicobacter pylori VacA via cellular lipid rafts in epithelial cells*, *Biochem Biophys Res Commun.* **303**, 2, 640-644
112. McClain, M.S., Schraw, W., Ricci, V., Boquet, P. & Cover, T.L. (2000). *Acid activation of Helicobacter pylori vacuolating cytotoxin (VacA) results in toxin internalization by eukaryotic cells*, *Mol Microbiol.* **37**, 2, 433-442
113. Iwamoto, H., Czajkowsky, D.M., Cover, T.L., Szabo, G. & Shao, Z. (1999). *VacA from Helicobacter pylori: a hexameric chloride channel*, *FEBS Lett.* **450**, 1-2, 101-104
114. Tombola, F., Oregna, F., Brutsche, S., Szabò, I., Del Giudice, G., Rappuoli, R., Montecucco, C., Papini, E. & Zoratti, M. (1999). *Inhibition of the vacuolating and anion channel activities of the VacA toxin of Helicobacter pylori*, *FEBS Lett.* **460**, 2, 221-225
115. Papini, E., de Bernard, M., Milia, E., Bugnoli, M., Zerial, M., Rappuoli, R. & Montecucco, C. (1994). *Cellular vacuoles induced by Helicobacter pylori originate from late endosomal compartments*, *Proc Natl Acad Sci U S A.* **91**, 21, 9720-9724
116. Atherton, J.C., Cao, P., Peek, R.M.J., Tummuru, M.K., Blaser, M.J. & Cover, T.L. (1995). *Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori. Association of specific vacA*

- types with cytotoxin production and peptic ulceration*, J Biol Chem. **270**, 30, 17771-17777
117. Solnick, J.V., Hansen, L.M., Salama, N.R., Boonjakuakul, J.K. & Syvanen, M. (2004). *Modification of Helicobacter pylori outer membrane protein expression during experimental infection of rhesus macaques*, Proc Natl Acad Sci U S A. **101**, 7, 2106-2111
118. Borén, T., Falk, P., Roth, K.A., Larson, G. & Normark, S. (1993). *Attachment of Helicobacter pylori to human gastric epithelium mediated by blood group antigens*, Science. **262**, 5141, 1892-1895
119. Yu, J., Leung, W.K., Go, M.Y.Y., Chan, M.C.W., To, K.F., Ng, E.K.W., Chan, F.K.L., Ling, T.K.W., Chung, S.C.S. & Sung, J.J.Y. (2002). *Relationship between Helicobacter pylori babA2 status with gastric epithelial cell turnover and premalignant gastric lesions*, Gut. **51**, 4, 480-484
120. Ilver, D., Arnqvist, A., Ogren, J., Frick, I.M., Kersulyte, D., Incecik, E.T., Berg, D.E., Covacci, A., Engstrand, L. & Borén, T. (1998). *Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging*, Science. **279**, 5349, 373-377
121. Bäckström, A., Lundberg, C., Kersulyte, D., Berg, D.E., Borén, T. & Arnqvist, A. (2004). *Metastability of Helicobacter pylori bab adhesin genes and dynamics in Lewis b antigen binding*, Proc Natl Acad Sci U S A. **101**, 48, 16923-16928
122. Hennig, E.E., Mernaugh, R., Edl, J., Cao, P. & Cover, T.L. (2004). *Heterogeneity among Helicobacter pylori strains in expression of the outer membrane protein BabA*, Infect Immun. **72**, 6, 3429-3435
123. Tomb, J.F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A. et al. (1997). *The complete genome sequence of the gastric pathogen Helicobacter pylori*, Nature. **388**, 6642, 539-547
124. Struyvé, M., Moons, M. & Tommassen, J. (1991). *Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein*, J Mol Biol. **218**, 1, 141-148
125. Jong, W.S.P., Sauri, A. & Luirink, J. (2010). *Extracellular production of recombinant proteins using bacterial autotransporters*, Curr Opin Biotechnol. **21**, 5, 646-652
126. Mahdavi, J., Sondén, B., Hurtig, M., Olfat, F.O., Forsberg, L., Roche, N., Angstrom, J., Larsson, T., Teneberg, S., Karlsson, K. et al. (2002). *Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation*, Science. **297**, 5581, 573-578
127. Sheu, S., Sheu, B., Yang, H., Lei, H. & Wu, J. (2007). *Anti-Lewis X antibody promotes Helicobacter pylori adhesion to gastric epithelial cells*, Infect Immun. **75**, 6, 2661-2667
128. Mobley, H., Mendz, G., Hazell, S., Hzel, M., Andersen, L., Wadstöm, T., Solnick, J., Vandamme, P., O'Rourke, J., Bode, G. et al. (2001) *Helicobacter pylori: Physiology and Genetics*
129. Yamaoka, Y. (2008). *Roles of Helicobacter pylori BabA in gastroduodenal pathogenesis*, World J Gastroenterol. **14**, 27, 4265-4272
130. Walz, A., Odenbreit, S., Stühler, K., Wattenberg, A., Meyer, H.E., Mahdavi, J., Borén, T. & Ruhl, S. (2009). *Identification of glycoprotein receptors within the human salivary proteome for the lectin-like BabA and SabA adhesins of Helicobacter pylori by fluorescence-based 2-D bacterial overlay*, Proteomics. **9**, 6, 1582-1592
131. PEMBERTON, R. (1994). *AGGLUTININS (LECTINS) FROM SOME BRITISH HIGHER FUNGI*, MYCOLOGICAL RESEARCH. **98**, Part 3, 277-290
132. Karlsson, K.A. (2001). *Pathogen-host protein-carbohydrate interactions as the basis of important infections*, Adv Exp Med Biol. **491**, , 431-443
133. Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S.J. & Knight, S.D. (1999). *X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic Escherichia coli*, Science. **285**, 5430, 1061-1066
134. Hamburger, Z.A., Brown, M.S., Isberg, R.R. & Bjorkman, P.J. (1999). *Crystal structure of invasins: a bacterial integrin-binding protein*, Science. **286**, 5438, 291-295

135. Wimmerova, M., Mitchell, E., Sanchez, J., Gautier, C. & Imberty, A. (2003). *Crystal structure of fungal lectin: six-bladed beta-propeller fold and novel fucose recognition mode for Aleuria aurantia lectin*, J Biol Chem. **278**, 29, 27059-27067
136. Imberty, A. & Varrot, A. (2008). *Microbial recognition of human cell surface glycoconjugates*, Curr Opin Struct Biol. **18**, 5, 567-576
137. Mancheño, J.M., Tateno, H., Goldstein, I.J., Martínez-Ripoll, M. & Hermoso, J.A. (2005). *Structural analysis of the Laetiporus sulphureus hemolytic pore-forming lectin in complex with sugars*, J Biol Chem. **280**, 17, 17251-17259
138. Imberty, A., Mitchell, E.P. & Wimmerová, M. (2005). *Structural basis of high-affinity glycan recognition by bacterial and fungal lectins*, Curr Opin Struct Biol. **15**, 5, 525-534
139. Remaut, H. & Waksman, G. (2004). *Structural biology of bacterial pathogenesis*, Curr Opin Struct Biol. **14**, 2, 161-170
140. Perret, S., Sabin, C., Dumon, C., Pokorná, M., Gautier, C., Galanina, O., Ilia, S., Bovin, N., Nicaise, M., Desmadril, M. et al. (2005). *Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-III of Pseudomonas aeruginosa*, Biochem J. **389**, Pt 2, 325-332
141. Petosa, C., Collier, R.J., Klimpel, K.R., Leppla, S.H. & Liddington, R.C. (1997). *Crystal structure of the anthrax toxin protective antigen*, Nature. **385**, 6619, 833-838
142. Milne, J.C. & Collier, R.J. (1993). *pH-dependent permeabilization of the plasma membrane of mammalian cells by anthrax protective antigen*, Mol Microbiol. **10**, 3, 647-653
143. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). *Basic local alignment search tool*, J Mol Biol. **215**, 3, 403-410
144. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T.L. (2008). *NCBI BLAST: a better web interface*, Nucleic Acids Res. **36**, Web Server issue, W5-9
145. Yoshida, E., Hidaka, M., Fushinobu, S., Koyanagi, T., Minami, H., Tamaki, H., Kitaoka, M., Katayama, T. & Kumagai, H. (2010). *Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 beta-glucosidase from Kluyveromyces marxianus*, Biochem J. , ,
146. Pohlner, J., Halter, R., Beyreuther, K. & Meyer, T.F. (1987). *Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA protease*, Nature. **325**, 6103, 458-462
147. Jose, J., Jähnig, F. & Meyer, T.F. (1995). *Common structural features of IgA1 protease-like outer membrane protein autotransporters*, Mol Microbiol. **18**, 2, 378-380
148. Brandon, L.D., Goehring, N., Janakiraman, A., Yan, A.W., Wu, T., Beckwith, J. & Goldberg, M.B. (2003). *IcsA, a polarly localized autotransporter with an atypical signal peptide, uses the Sec apparatus for secretion, although the Sec apparatus is circumferentially distributed*, Mol Microbiol. **50**, 1, 45-60
149. Veiga, E., de Lorenzo, V. & Fernández, L.A. (1999). *Probing secretion and translocation of a beta-autotransporter using a reporter single-chain Fv as a cognate passenger domain*, Mol Microbiol. **33**, 6, 1232-1243
150. Oliver, D.C., Huang, G., Nodel, E., Pleasance, S. & Fernandez, R.C. (2003). *A conserved region within the Bordetella pertussis autotransporter BrkA is necessary for folding of its passenger domain*, Mol Microbiol. **47**, 5, 1367-1383
151. de Cock, H., Brandenburg, K., Wiese, A., Holst, O. & Seydel, U. (1999). *Non-lamellar structure and negative charges of lipopolysaccharides required for efficient folding of outer membrane protein PhoE of Escherichia coli*, J Biol Chem. **274**, 8, 5114-5119
152. Voulhoux, R., Bos, M.P., Geurtsen, J., Mols, M. & Tommassen, J. (2003). *Role of a highly conserved bacterial protein in outer membrane protein assembly*, Science. **299**, 5604, 262-265
153. Sauri, A., Soprova, Z., Wickström, D., de Gier, J., Van der Schors, R.C., Smit, A.B., Jong, W.S.P. & Luirink, J. (2009). *The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease*, Microbiology. **155**, Pt 12, 3982-3991

154. Bos, M.P., Robert, V. & Tommassen, J. (2007). *Biogenesis of the gram-negative bacterial outer membrane*, *Annu Rev Microbiol.* **61**, 191-214
155. Graf zu Eulenburg, G.A.B. (2009). *Strukturelle und funktionelle Charakterisierung von Adhäsinen aus humanpathogenen Candida-Spezies*, Diploma thesis.
156. Dubendorff, J.W. & Studier, F.W. (1991). *Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor*, *J Mol Biol.* **219**, 1, 45-59
157. Bessette, P.H., Aslund, F., Beckwith, J. & Georgiou, G. (1999). *Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm*, *Proc Natl Acad Sci U S A.* **96**, 24, 13703-13708
158. de Marco, A. (2009). *Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli*, *Microb Cell Fact.* **8**, 26
159. Mullis, K.B. & Faloona, F.A. (1987). *Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction*, *Methods Enzymol.* **155**, 335-350
160. Integrated DNA Technologies ( ) , <http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/>
161. Braman, J., Papworth, C. & Greener, A. (1996). *Site-directed mutagenesis using double-stranded plasmid DNA templates*, *Methods Mol Biol.* **57**, 31-44
162. Birnboim, H.C. & Doly, J. (1979). *A rapid alkaline extraction procedure for screening recombinant plasmid DNA*, *Nucleic Acids Res.* **7**, 6, 1513-1523
163. Dower, W.J., Miller, J.F. & Ragsdale, C.W. (1988). *High efficiency transformation of E. coli by high voltage electroporation*, *Nucleic Acids Res.* **16**, 13, 6127-6145
164. Kaiser, C. (2001) Untersuchung am Adhäsin BabA aus dem Humanpathogen Helicobacter pylori
165. Porath, J., Carlsson, J., Olsson, I. & Belfrage, G. (1975). *Metal chelate affinity chromatography, a new approach to protein fractionation*, *Nature.* **258**, 5536, 598-599
166. Stone, M.C. & Carta, G. (2007). *Protein adsorption and transport in agarose and dextran-grafted agarose media for ion exchange chromatography*, *J Chromatogr A.* **1146**, 2, 202-215
167. GE Healthcare (2010), PD-10 Desalting columns handbook
168. GE Healthcare (2010), illustra MicroSpin G50 Columns
169. Novagen/ Merck (2010), iFOLD Protein Refolding System 2 user protocol
170. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. & Bairoch, A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*, *The Proteomics Protocols Handbook.* 571-607
171. Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J. & Klenk, D.C. (1985). *Measurement of protein using bicinchoninic acid*, *Anal Biochem.* **150**, 1, 76-85
172. Laemmli, U.K. (1970). *Cleavage of structural proteins during the assembly of the head of bacteriophage T4*, *Nature.* **227**, 5259, 680-685
173. Bennett, J. & Scott, K.J. (1971). *Quantitative staining of fraction I protein in polyacrylamide gels using Coomassie brilliant blue*, *Anal Biochem.* **43**, 1, 173-182
174. Burnette, W.N. (1981). *"Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A*, *Anal Biochem.* **112**, 2, 195-203
175. Greenfield, N.J. (1996). *Methods to estimate the conformation of proteins and polypeptides from circular dichroism data*, *Anal Biochem.* **235**, 1, 1-10
176. Kelly, S.M. & Price, N.C. (2000). *The use of circular dichroism in the investigation of protein structure and function*, *Curr Protein Pept Sci.* **1**, 4, 349-384
177. Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M.E., Alvarez, R., Bryan, M.C., Fazio, F., Calarese, D., Stevens, J. et al. (2004). *Printed covalent glycan array for ligand profiling of diverse glycan binding proteins*, *Proc Natl Acad Sci U S A.* **101**, 49, 17033-17038

178. invitrogen, Technical Tip: Grater Control and More Reaction Time with SDP Ester, 2007
179. Lakowicz, J.R. (1995). *Fluorescence spectroscopy of biomolecules*, Encyclopedia of Molecular Biology and Molecular Medicine. 294 - 306
180. Rosenthal, H.E. (1967). *A graphic method for the determination and presentation of binding parameters in a complex system*, Anal Biochem. **20**, 3, 525-532
181. Zierler, K. (1989). *Misuse of nonlinear Scatchard plots*, Trends Biochem Sci. **14**, 8, 314-317
182. Jancarik, J., Scott, W.G., Milligan, D.L., Koshland, D.E.J. & Kim, S.H. (1991). *Crystallization and preliminary X-ray diffraction study of the ligand-binding domain of the bacterial chemotaxis-mediating aspartate receptor of Salmonella typhimurium*, J Mol Biol. **221**, 1, 31-34
183. Qiagen (2010) ,  
<http://www.qiagen.com/products/protein/crystallization/screeningprotein/crystallizationconditions/introduction.aspx#scre>
184. Leonard, G., McCarthy, J., Nurizzo, D. & Thibault, X. (2007). *Technical Report: Automatic Experiments at the European Synchrotron Radiation Facility MX Beam-lines* , Synchrotron Radiation News. **20**, 18 - 24
185. Leslie, A. (1992). *Recent changes to the MOSFLM package for processing film and image plate data*, Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography. **26**
186. Stein, N. (2008). *CHAINSAW: a program for mutating pdb files used as templates in molecular replacement.*, J. Appl. Cryst.. **41**, 641 - 643
187. Sali, A. & Blundell, T.L. (1993). *Comparative protein modelling by satisfaction of spatial restraints*, J Mol Biol. **234**, 3, 779-815
188. Kuntal, B.K., Aparoy, P. & Reddanna, P. (2010). *EasyModeller: A graphical interface to MODELLER*, BMC Res Notes. **3**, 226
189. Dunbrack, R.L.J. (2006). *Sequence comparison and protein structure prediction*, Curr Opin Struct Biol. **16**, 3, 374-384
190. Tramontano, A. ( 2006) Protein Structure Prediction
191. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995). *SCOP: a structural classification of proteins database for the investigation of sequences and structures*, J Mol Biol. **247**, 4, 536-540
192. Veelders, M. (2008). *Flokkuline aus Saccharomyces cerevisiae: funktionelle und strukturelle Untersuchungen*, Diploma thesis.
193. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*, Nucleic Acids Res. **25**, 17, 3389-3402
194. Gotoh, O. & Tagashira, Y. (1986). *Sequence search on a supercomputer*, Nucleic Acids Res. **14**, 1, 57-64
195. Dayhoff, M., Schwarz, R. & Orcutt, B. ( 1978) Atlas of Protein Sequence and Structure
196. Henikoff, S. & Henikoff, J.G. (1992). *Amino acid substitution matrices from protein blocks*, Proc Natl Acad Sci U S A. **89**, 22, 10915-10919
197. Zhang, Y. (2009). *Protein structure prediction: when is it useful?*, Curr Opin Struct Biol. **19**, 2, 145-155
198. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. et al. (2007). *Clustal W and Clustal X version 2.0*, Bioinformatics. **23**, 21, 2947-2948
199. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*, Nucleic Acids Res. **22**, 22, 4673-4680
200. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997). *The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by*

- quality analysis tools*, Nucleic Acids Res. **25**, 24, 4876-4882
201. Thompson, J.D. (1995). *Introducing variable gap penalties to sequence alignment in linear space*, Comput Appl Biosci. **11**, 2, 181-186
202. Saitou, N. & Nei, M. (1987). *The neighbor-joining method: a new method for reconstructing phylogenetic trees*, Mol Biol Evol. **4**, 4, 406-425
203. Gascuel, O. & Steel, M. (2006). *Neighbor-joining revealed*, Mol Biol Evol. **23**, 11, 1997-2000
204. Desper, R. & Gascuel, O. (2004). *Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting*, Mol Biol Evol. **21**, 3, 587-598
205. Felsenstein, J. (2004) *Inferring phylogenies*
206. Schwarzenbacher, R., Godzik, A., Grzechnik, S.K. & Jaroszewski, L. (2004). *The importance of alignment accuracy for molecular replacement*, Acta Crystallogr D Biol Crystallogr. **60**, Pt 7, 1229-1236
207. Eswar, N., Marti-Renom, M., Webb, M., Madhusudhan, M., Eramian, D., Shen, M., Pieper, U. & Sali, A. (2006). *Comparative Protein Structure Modeling With MODELLER.*, Current Protocols in Bioinformatics. **Supplement 15**, 5.6.1-5.6.30
208. Bushan, K. (2009), <http://sites.google.com/site/bioinformatikz/>
209. Shen, M. & Sali, A. (2006). *Statistical potential for assessment and prediction of protein structures*, Protein Sci. **15**, 11, 2507-2524
210. The PyMOL Molecular Graphics System, Version 1.2r1, Schrödinger, LLC.
211. Kabsch, W. (2010). *XDS*, Acta Crystallogr D Biol Crystallogr. **66**, Pt 2, 125-132
212. The Computational Collaborative Project Number 4, (1994). *The CCP4 suite: programs for protein crystallography*, Acta Crystallogr D Biol Crystallogr. **50**, Pt 5, 760-763
213. Diederichs, K. & Karplus, P.A. (1997). *Improved R-factors for diffraction data analysis in macromolecular crystallography*, Nat Struct Biol. **4**, 4, 269-275
214. Drenth, J. (1999). *Principles of Protein X-Ray Crystallography*, Springer.
215. Rossmann, M. & Blow, D. (1962). *The Detection of Sub-Units Within the Crystallographic Asymmetric Unit*, Acta Crystallographica. **15**, 24-31
216. McCoy, A., Grosse-Kunstleve, R., Adams, P., Winn, M., Storoni, L. & Read, R. (2007). *Phaser Crystallographic Software*, Journal of Applied Crystallography. **40**, 658-674
217. Fisher, R. (1922). *Philos. Trans. R. Soc. A.* **222**, 309-368
218. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010). *Features and development of Coot*, Acta Crystallogr D Biol Crystallogr. **66**, Pt 4, 486-501
219. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). *Refinement of macromolecular structures by the maximum-likelihood method*, Acta Crystallogr D Biol Crystallogr. **53**, Pt 3, 240-255
220. Langer, G., Cohen, S.X., Lamzin, V.S. & Perrakis, A. (2008). *Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7*, Nat Protoc. **3**, 7, 1171-1179
221. Winn, M.D., Murshudov, G.N. & Papiz, M.Z. (2003). *Macromolecular TLS refinement in REFMAC at moderate resolutions*, Methods Enzymol. **374**, 300-321
222. Kostrewa, D. (1997). *Bulk Solvent Correction: Practical Application and Effects in Reciprocal and Real Space*, CCP4 newsletter. **34**, 9-22
223. Painter, J. & Merritt, E.A. (2006). *Optimal description of a protein structure in terms of multiple groups undergoing TLS motion*, Acta Crystallogr D Biol Crystallogr. **62**, Pt 4, 439-450
224. Painter, J. & Merritt, E. (2006). *TLSMD web server for the generation of multi-group TLS models*, Journal of Applied Crystallography. **39**, 109-111
225. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M., Appel, R. & Bairoch, A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*, The Proteomics Protocols Handbook. **52**, 571-607

226. Qtiplot (2010) , <http://www.qtiplot.ro>
227. Greenfield, N.J. (2006). *Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions*, Nat Protoc. **1**, 6, 2527-2535
228. Andrade, M.A., Chacón, P., Merelo, J.J. & Morán, F. (1993). *Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network*, Protein Eng. **6**, 4, 383-390
229. Kohonen, T., Mäkisara, K. & Saramäki, T. (1984). *Phonotopic maps - Insightful representation of phonological features for speech recognition*, Proceedings of the international conference on pattern recognition. , , 182-185
230. Campbell, R. () , <http://www.ks.uiuc.edu/Research/vmd/>
231. Shindyalov, I.N. & Bourne, P.E. (1998). *Protein structure alignment by incremental combinatorial extension (CE) of the optimal path*, Protein Eng. **11**, 9, 739-747
232. Cole, C., Barber, J.D. & Barton, G.J. (2008). *The Jpred 3 secondary structure prediction server*, Nucleic Acids Res. **36**, Web Server issue, W197-201
233. Cuff, J.A. & Barton, G.J. (2000). *Application of multiple sequence alignment profiles to improve protein secondary structure prediction*, Proteins. **40**, 3, 502-511
234. Arendrup, M.C. (2010). *Epidemiology of invasive candidiasis*, Curr Opin Crit Care. **16**, 5, 445-452
235. Prescott, L., Harley, J. & Klein, D. (1996) Microbiology
236. Papadopoulos, J.S. & Agarwala, R. (2007). *COBALT: constraint-based alignment tool for multiple protein sequences*, Bioinformatics. **23**, 9, 1073-1079
237. Avogadro (2010), [http://avogadro.openmolecules.net/wiki/Main\\_Page](http://avogadro.openmolecules.net/wiki/Main_Page)
238. Scatchard, G. (1949). *The attractions of proteins for small molecules and ions*, annals of the new yor academy of sciences. **51**, 660-672
239. Bai, Y., Zhang, Y., Chen, Y., Jin, J., Zhang, Z. & Zhou, D. (2004). *Cloning and expression and immunogenicity of Helicobacter pylori BabA2 gene*, World J Gastroenterol. **10**, 17, 2560-2562
240. Yang, J.T., Wu, C.S. & Martinez, H.M. (1986). *Calculation of protein conformation from circular dichroism*, Methods Enzymol. **130**, 208-269
241. Redfield, C.H.R.I.S.T.I.N.A. (2004). *NMR studies of partially folded molten-globule states*, Methods Mol Biol. **278**
242. Fernandez, E. & Tobler, S. (2006) Misbehaving Proteins, Protein (Mis)Folding, Aggregation, and Stability
243. Hamuro, Y., Coales, S.J., Southern, M.R., Nemeth-Cawley, J.F., Stranz, D.D. & Griffin, P.R. (2003). *Rapid analysis of protein structure and dynamics by hydrogen/deuterium exchange mass spectrometry*, J Biomol Tech. **14**, 3, 171-182
244. Bendtsen, J.D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). *Improved prediction of signal peptides: SignalP 3.0*, J Mol Biol. **340**, 4, 783-795
245. Goodwin, A.C., Weinberger, D.M., Ford, C.B., Nelson, J.C., Snider, J.D., Hall, J.D., Paules, C.I., Peek, R.M.J. & Forsyth, M.H. (2008). *Expression of the Helicobacter pylori adhesin SabA is controlled via phase variation and the ArsRS signal transduction system*, Microbiology. **154**, Pt 8, 2231-2240
246. Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K. et al. (2010). *The Pfam protein families database*, Nucleic Acids Res. **38**, Database issue, D211-22
247. Ohgushi, M. & Wada, A. (1983). *'Molten-globule state': a compact form of globular proteins with mobile side-chains*, FEBS Lett. **164**, 1, 21-24
248. Glatz, C. (1992) Modeling of aggregation-precipitation phenomena
249. Koshland, D.E. (1958). *Application of a Theory of Enzyme Specificity to Protein Synthesis*, Proc Natl Acad Sci U S A. **44**, 2, 98-104

250. Papadogiannakis, N., Willén, R., Carlén, B., Sjöstedt, S., Wadström, T. & Gad, A. (2000). *Modes of adherence of Helicobacter pylori to gastric surface epithelium in gastroduodenal disease: a possible sequence of events leading to internalisation*, APMIS. **108**, 6, 439-447
251. Edwards, N.J., Monteiro, M.A., Faller, G., Walsh, E.J., Moran, A.P., Roberts, I.S. & High, N.J. (2000). *Lewis X structures in the O antigen side-chain promote adhesion of Helicobacter pylori to the gastric epithelium*, Mol Microbiol. **35**, 6, 1530-1539
252. Capraro, J., Spotti, P., Magni, C., Scarafoni, A. & Duranti, M. (2010). *Spectroscopic studies on the pH-dependent structural dynamics of gamma-conglutin, the blood glucose-lowering protein of lupin seeds*, Int J Biol Macromol.
253. Styer, C.M., Hansen, L.M., Cooke, C.L., Gundersen, A.M., Choi, S.S., Berg, D.E., Benghezal, M., Marshall, B.J., Peek, R.M.J., Borén, T. et al. (2010). *Expression of the BabA adhesin during experimental infection with Helicobacter pylori*, Infect Immun. **78**, 4, 1593-1600
254. Pride, D.T., Meinersmann, R.J. & Blaser, M.J. (2001). *Allelic Variation within Helicobacter pylori babA and babB*, Infect Immun. **69**, 2, 1160-1171
255. Lindén, S., Nordman, H., Hedenbro, J., Hurtig, M., Borén, T. & Carlstedt, I. (2002). *Strain- and blood group-dependent binding of Helicobacter pylori to human gastric MUC5AC glycoforms*, Gastroenterology. **123**, 6, 1923-1930
256. Appelmelk, B.J., Monteiro, M.A., Martin, S.L., Moran, A.P. & Vandenbroucke-Grauls, C.M. (2000). *Why Helicobacter pylori has Lewis antigens*, Trends Microbiol. **8**, 12, 565-570
257. Martin, S.L., Edbrooke, M.R., Hodgman, T.C., van den Eijnden, D.H. & Bird, M.I. (1997). *Lewis X biosynthesis in Helicobacter pylori. Molecular cloning of an alpha(1,3)-fucosyltransferase gene*, J Biol Chem. **272**, 34, 21349-21356
258. Moran, A.P., Prendergast, M.M. & Appelmelk, B.J. (1996). *Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease*, FEMS Immunol Med Microbiol. **16**, 2, 105-115
259. Kalugin, V., Herstellung und Charakterisierung von Epa-Mutanten, 2009. Bachelorarbeit
260. Chávez, M.I., Andreu, C., Vidal, P., Aboitiz, N., Freire, F., Groves, P., Asensio, J.L., Asensio, G., Muraki, M., Cañada, F.J. et al. (2005). *On the importance of carbohydrate-aromatic interactions for the molecular recognition of oligosaccharides by proteins: NMR studies of the structure and binding affinity of AcAMP2-like peptides with non-natural naphthyl and fluoroaromatic residues*, Chemistry. **11**, 23, 7060-7074
261. Krissinel, E. & Henrick, K. (2007). *Inference of macromolecular assemblies from crystalline state*, J Mol Biol. **372**, 3, 774-797
262. Sando, L., Pearson, R., Gray, C., Parker, P., Hawken, R., Thomson, P.C., Meadows, J.R.S., Kongsuwan, K., Smith, S. & Tellam, R.L. (2009). *Bovine Muc1 is a highly polymorphic gene encoding an extensively glycosylated mucin that binds bacteria*, J Dairy Sci. **92**, 10, 5276-5291
263. Mendel, C.M., Licko, V. & Kane, J.P. (1985). *The effect of ligand heterogeneity on the Scatchard plot. Particular relevance to lipoprotein binding analysis*, J Biol Chem. **260**, 6, 3451-3455





## 9 Appendices

### 9.1 Appendix I: Glycan Arrays

All data obtained directly from the Consortium for Functional Glycomics data bank (<http://www.functionalglycomics.org/static/consortium/resources/resourcecoreh.shtml>)

#### 9.1.1 Array V3.1

Number	Glycan
1	Neu5Aca2-8Neu5Acb-Sp17
2	Neu5Aca2-8Neu5Aca2-8Neu5Acb-Sp8
3	Neu5Gcb2-6Galb1-4GlcNAc-Sp8
4	Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
5	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
6	a-D-Gal-Sp8
7	a-D-Glc-Sp8
8	a-D-Man-Sp8
9	a-GalNAc-Sp8
10	a-L-Fuc-Sp8
11	a-L-Fuc-Sp9
12	a-L-Rha-Sp8
13	a-Neu5Ac-Sp8
14	a-Neu5Ac-Sp11
15	b-Neu5Ac-Sp8
16	b-D-Gal-Sp8
17	b-D-Glc-Sp8
18	b-D-Man-Sp8
19	b-GalNAc-Sp8

Number	Glycan
20	b-GlcNAc–Sp0
21	b-GlcNAc–Sp8
22	b-GlcN(Gc)-Sp8
23	(Galb1-4GlcNAcb)2-3,6-GalNAca-Sp8
24	GlcNAcb1-3(GlcNAcb1-4)(GlcNAcb1-6)GlcNAc-Sp8
25	[3OSO3][6OSO3]Galb1-4[6OSO3]GlcNAcb-Sp0
26	[3OSO3][6OSO3]Galb1-4GlcNAcb-Sp0
27	[3OSO3]Galb1-4Glc-Sp8
28	[3OSO3]Galb1-4(6OSO3)Glc–Sp0
29	[3OSO3]Galb1-4(6OSO3)Glc–Sp8
30	[3OSO3]Galb1-3(Fuca1-4)GlcNAcb–Sp8
31	[3OSO3]Galb1-3GalNAca-Sp8
32	[3OSO3]Galb1-3GlcNAcb–Sp8
33	[3OSO3]Galb1-4(Fuca1-3)GlcNAcb–Sp8
34	[3OSO3]Galb1-4[6OSO3]GlcNAcb-Sp8
35	[3OSO3]Galb1-4GlcNAcb–Sp0
36	[3OSO3]Galb1-4GlcNAcb-Sp8
37	[3OSO3]Galb–Sp8
38	[4OSO3][6OSO3]Galb1-4GlcNAcb-Sp0
39	[4OSO3]Galb1-4GlcNAcb-Sp8
40	6-H2PO3Mana-Sp8
41	[6OSO3]Galb1-4Glc–Sp0
42	[6OSO3]Galb1-4Glc–Sp8
43	[6OSO3]Galb1-4GlcNAcb–Sp8
44	[6OSO3]Galb1-4[6OSO3]Glc-Sp8
45	NeuAca2-3[6OSO3]Galb1-4GlcNAcb–Sp8
46	[6OSO3]GlcNAcb–Sp8
47	9NAcNeu5Aca-Sp8
48	9NAcNeu5Aca2-6Galb1-4GlcNAcb-Sp8
49	Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp13
50	GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp13
51	Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp13
52	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-

Number	Glycan
	6)Manb1-4GlcNAcb1-4GlcNAcb-Sp13
53	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp8
54	Fuca1-2Galb1-3GalNAcb1-3Gala-Sp9
55	Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp9
56	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb-Sp8
57	Fuca1-2Galb1-3GalNAca-Sp8
58	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp0
59	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp9
60	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glc-Sp10
61	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glc-Sp8
62	Fuca1-2Galb1-3GlcNAcb-Sp0
63	Fuca1-2Galb1-3GlcNAcb-Sp8
64	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
65	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
66	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp0
67	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp8
68	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAc-Sp0
69	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
70	Fuca1-2Galb1-4GlcNAcb-Sp0
71	Fuca1-2Galb1-4GlcNAcb-Sp8
72	Fuca1-2Galb1-4Glc-Sp0
73	Fuca1-2Galb-Sp8
74	Fuca1-3GlcNAcb-Sp8
75	Fuca1-4GlcNAcb-Sp8
76	Fucb1-3GlcNAcb-Sp8
77	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0
78	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0
79	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0
80	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8
81	GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0
82	GalNAca1-3(Fuca1-2)Galb-Sp8
83	GalNAca1-3GalNAcb-Sp8

Number	Glycan
84	GalNAca1-3Galb-Sp8
85	GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8
86	GalNAcb1-3GalNAca-Sp8
87	GalNAcb1-3(Fuca1-2)Galb-Sp8
88	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0
89	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0
90	GalNAcb1-4GlcNAcb-Sp0
91	GalNAcb1-4GlcNAcb-Sp8
92	Gala1-2Galb-Sp8
93	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0
94	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0
95	Gala1-3(Fuca1-2)Galb1-4GlcNAc-Sp0
96	Gala1-3(Fuca1-2)Galb1-4Glc-Sp0
97	Gala1-3(Fuca1-2)Galb-Sp8
98	Gala1-3(Gala1-4)Galb1-4GlcNAcb-Sp8
99	Gala1-3GalNAca-Sp8
100	Gala1-3GalNAcb-Sp8
101	Gala1-3Galb1-4(Fuca1-3)GlcNAcb-Sp8
102	Gala1-3Galb1-3GlcNAcb-Sp0
103	Gala1-3Galb1-4GlcNAcb-Sp8
104	Gala1-3Galb1-4Glc-Sp0
105	Gala1-3Galb-Sp8
106	Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8
107	Gala1-4Galb1-4GlcNAcb-Sp0
108	Gala1-4Galb1-4GlcNAcb-Sp8
109	Gala1-4Galb1-4Glc-Sp0
110	Gala1-4GlcNAcb-Sp8
111	Gala1-6Glc-Sp8
112	Galb1-2Galb-Sp8
113	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
114	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4GlcNAcb-Sp0
115	Galb1-3(Fuca1-4)GlcNAc-Sp0
116	Galb1-3(Fuca1-4)GlcNAc-Sp8

Number	Glycan
117	Galb1-3(Fuca1-4)GlcNAcb-Sp8
118	Galb1-3(Galb1-4GlcNAcb1-6)GalNAca-Sp8
119	Galb1-3(GlcNAcb1-6)GalNAca-Sp8
120	Galb1-3(Neu5Aca2-6)GalNAca-Sp8
121	Galb1-3(Neu5Acb2-6)GalNAca-Sp8
122	Galb1-3(Neu5Aca2-6)GlcNAcb1-4Galb1-4Glc-Sp10
123	Galb1-3GalNAca-Sp8
124	Galb1-3GalNAcb-Sp8
125	Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0
126	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp0
127	Galb1-3GalNAcb1-4Galb1-4Glc-Sp8
128	Galb1-3Galb-Sp8
129	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0
130	Galb1-3GlcNAcb1-3Galb1-4Glc-Sp10
131	Galb1-3GlcNAcb-Sp0
132	Galb1-3GlcNAcb-Sp8
133	Galb1-4(Fuca1-3)GlcNAcb-Sp0
134	Galb1-4(Fuca1-3)GlcNAcb-Sp8
135	Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb-Sp0
136	Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb-Sp0
137	Galb1-4[6OSO3]Glc-Sp0
138	Galb1-4[6OSO3]Glc-Sp8
139	Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8
140	Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8
141	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
142	Galb1-4GlcNAcb1-3GalNAca-Sp8
143	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
144	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
145	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
146	Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0
147	Galb1-4GlcNAcb1-3Galb1-4Glc-Sp8
148	Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp8

Number	Glycan
149	Galb1-4GlcNAcb1-6GalNAca-Sp8
150	Galb1-4GlcNAcb-Sp0
151	Galb1-4GlcNAcb-Sp8
152	Galb1-4Glc-Sp0
153	Galb1-4Glc-Sp8
154	GlcNAca1-3Galb1-4GlcNAcb-Sp8
155	GlcNAca1-6Galb1-4GlcNAcb-Sp8
156	GlcNAcb1-2Galb1-3GalNAca-Sp8
157	GlcNAcb1-3(GlcNAcb1-6)GalNAca-Sp8
158	GlcNAcb1-3(GlcNAcb1-6)Galb1-4GlcNAcb-Sp8
159	GlcNAcb1-3GalNAca-Sp8
160	GlcNAcb1-3Galb-Sp8
161	GlcNAcb1-3Galb1-3GalNAca-Sp8
162	GlcNAcb1-3Galb1-4GlcNAcb-Sp0
163	GlcNAcb1-3Galb1-4GlcNAcb-Sp8
164	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
165	GlcNAcb1-3Galb1-4Glc-Sp0
166	GlcNAcb1-4MDPLys
167	GlcNAcb1-4(GlcNAcb1-6)GalNAca-Sp8
168	GlcNAcb1-4Galb1-4GlcNAcb-Sp8
169	(GlcNAcb1-4)6b-Sp8
170	(GlcNAcb1-4)5b-Sp8
171	GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8
172	GlcNAcb1-6(Galb1-3)GalNAca-Sp8
173	GlcNAcb1-6GalNAca-Sp8
174	GlcNAcb1-6Galb1-4GlcNAcb-Sp8
175	Glca1-4Glc-Sp8
176	Glca1-4Glca-Sp8
177	Glca1-6Glca1-6Glc-Sp8
178	Glc1-4Glc-Sp8
179	Glc1-6Glc-Sp8
180	G-ol-Sp8
181	GlcAa-Sp8

Number	Glycan
182	GlcAb-Sp8
183	GlcAb1-3Galb-Sp8
184	GlcAb1-6Galb-Sp8
185	KDNa2-3Galb1-3GlcNAcb-Sp0
186	KDNa2-3Galb1-4GlcNAcb-Sp0
187	Mana1-2Mana1-2Mana1-3Mana-Sp9
188	Mana1-2Mana1-3(Mana1-2Mana1-6)Mana-Sp9
189	Mana1-2Mana1-3Mana-Sp9
190	Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
191	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana2Mana2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
192	Mana1-2Mana1-2Mana1-3(Mana1-2Mana1-3(Mana1-2Mana1-6)Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
193	Mana1-3(Mana1-6)Mana-Sp9
194	Mana1-3(Mana1-2Mana1-2Mana1-6)Mana-Sp9
195	Mana1-6(Mana1-3)Mana1-6(Mana2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
196	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
197	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
198	Manb1-4GlcNAcb-Sp0
199	Fuca1-3(Galb1-4)GlcNAcb1-2Mana1-3(Fuca1-3(Galb1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
200	Neu5Aca2-3Galb1-3GalNAca-Sp8
201	NeuAca2-8NeuAca2-8NeuAca2-8NeuAca2-3(GalNAcb1-4)Galb1-4Glc-Sp0
202	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4Glc-Sp0
203	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4Glc-Sp0
204	Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4Glc-Sp0
205	Neu5Aca2-8Neu5Aca2-8Neu5Aca-Sp8
206	Neu5Aca2-3(6-O-Su)Galb1-4(Fuca1-3)GlcNAcb-Sp8
207	Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb-Sp0
208	Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb-Sp8
209	Neu5Aca2-3(GalNAcb1-4)Galb1-4Glc-Sp0
210	NeuAca2-3(NeuAca2-3Galb1-3GalNAcb1-4)Galb1-4Glc-Sp0
211	Neu5Aca2-3(Neu5Aca2-6)GalNAca-Sp8



Number	Glycan
212	Neu5Aca2-3GalNAca-Sp8
213	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0
214	Neu5Aca2-3Galb1-3(6OSO3)GlcNAc-Sp8
215	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp8
216	NeuAca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb Sp0
217	Neu5Aca2-3Galb1-3(Neu5Aca2-3Galb1-4)GlcNAcb-Sp8
218	Neu5Aca2-3Galb1-3[6OSO3]GalNAca-Sp8
219	Neu5Aca2-3Galb1-3(Neu5Aca2-6)GalNAca-Sp8
220	Neu5Aca2-3Galb-Sp8
221	NeuAca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0
222	NeuAca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0
223	Neu5Aca2-3Galb1-3GlcNAcb-Sp0
224	Neu5Aca2-3Galb1-3GlcNAcb-Sp8
225	Neu5Aca2-3Galb1-4[6OSO3]GlcNAcb-Sp8
226	Neu5Aca2-3Galb1-4(Fuca1-3)(6OSO3)GlcNAcb-Sp8
227	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
228	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
229	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp8
230	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8
231	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8
232	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAc-Sp0
233	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
234	Neu5Aca2-3Galb1-4GlcNAcb-Sp0
235	Neu5Aca2-3Galb1-4GlcNAcb-Sp8
236	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
237	Neu5Aca2-3Galb1-4Glc-Sp0
238	Neu5Aca2-3Galb1-4Glc-Sp8
239	Galb1-4GlcNAcb1-2Mana1-3(Fuca1-3(Galb1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
240	Neu5Aca2-6GalNAca-Sp8
241	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0
242	Neu5Aca2-6Galb1-4[6OSO3]GlcNAcb-Sp8
243	Neu5Aca2-6Galb1-4GlcNAcb-Sp0

Number	Glycan
244	Neu5Aca2-6Galb1-4GlcNAcb-Sp8
245	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
246	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
247	Neu5Aca2-6Galb1-4Glc-Sp0
248	Neu5Aca2-6Galb1-4Glc-Sp8
249	Neu5Aca2-6Galb-Sp8
250	Neu5Aca2-8Neu5Aca-Sp8
251	Neu5Aca2-8Neu5Aca2-3Galb1-4Glc-Sp0
252	Neu5Acb2-6GalNAca-Sp8
253	Neu5Acb2-6Galb1-4GlcNAcb-Sp8
254	Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
255	Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
256	Neu5Gca2-3Galb1-3GlcNAcb-Sp0
257	Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
258	Neu5Gca2-3Galb1-4GlcNAcb-Sp0
259	Neu5Gca2-3Galb1-4Glc-Sp0
260	Neu5Gca2-6GalNAca-Sp0
261	Neu5Gca2-6Galb1-4GlcNAcb-Sp0
262	Neu5Gca-Sp8
263	[3OSO3]Galb1-4(Fuca1-3)(6OSO3)Glc-Sp0
264	[3OSO3]Galb1-4(Fuca1-3)Glc-Sp0
265	[3OSO3]Galb1-4[Fuca1-3][6OSO3]GlcNAc-Sp8
266	[3OSO3]Galb1-4[Fuca1-3]GlcNAc-Sp0
267	Fuca1-2[6OSO3]Galb1-4GlcNAc-Sp0
268	Fuca1-2Galb1-4[6OSO3]GlcNAc-Sp8
269	Fuca1-2[6OSO3]Galb1-4[6OSO3]Glc-Sp0
270	Fuca1-2-(6OSO3)-Galb1-4Glc-Sp0
271	Fuca1-2-Galb1-4[6OSO3]Glc-Sp0
272	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
273	Galb1-3-(Galb1-4GlcNAcb1-6)GalNAc-Sp14
274	Galb1-3(GlcNAcb1-6)GalNAc-Sp14
275	Galb1-3-(Neu5Aa2-3Galb1-4GlcNAcb1-6)GalNAc-Sp14

Number	Glycan
276	Galb1-3GalNAc-Sp14
277	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0
278	Galb1-4[Fuca1-3][6OSO3]GlcNAc-Sp0
279	Galb1-4[Fuca1-3][6OSO3]Glc-Sp0
280	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
281	Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0
282	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0
283	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0
284	[3OSO3]Galb1-4[6OSO3]GlcNAcb-Sp0
285	[3OSO3][4OSO3]Galb1-4GlcNAcb-SpSp0
286	[6OSO3]Galb1-4[6OSO3]GlcNAcb-Sp0
287	6-H2PO3Glc-Sp10
288	Gala1-3(Fuca1-2)Galb-Sp18
289	Gala1-3GalNAca-Sp16
290	Galb1-3GalNAca-Sp16
291	Galb1-3(Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6)GalNAc-Sp14
292	Galb1-3Galb1-4GlcNAcb-Sp8
293	Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
294	Galb1-4GlcNAcb1-3(Galb1-4GlcNAcb1-6)Galb1-4GlcNAc-Sp0
295	Galb1-4GlcNAcb1-3(GlcNAcb1-6)Galb1-4GlcNAc-Sp0
296	Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0
297	Galb1-4GlcNAcb1-6Galb1-4GlcNAcb-Sp0
298	GalNAca-Sp15
299	GalNAca1-3(Fuca1-2)Galb-Sp18
300	GalNAcb1-3Galb-Sp8
301	GlcAb1-3GlcNAcb-Sp8
302	GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
303	GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
304	GlcNAcb1-3Man-Sp10
305	GlcNAcb1-4GlcNAcb-Sp10
306	GlcNAcb1-4GlcNAcb-Sp12
307	HOOC(CH <sub>3</sub> )CH-3-O-GlcNAcb1-4GlcNAcb-Sp10

Number	Glycan
308	Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
309	Mana1-6Manb-Sp10
310	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb-Sp10
311	Mana1-2Mana1-2Mana1-3(Mana1-2Mana1-6(Mana1-3)Mana1-6)Mana-Sp9
312	Mana1-2Mana1-2Mana1-3(Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6)Mana-Sp9
313	Neu5Aca2-3Galb1-3(Neu5Aca2-3Galb1-4GlcNAcb1-6)GalNAc-Sp14
314	Neu5Aca2-3Galb1-3(Neu5Aca2-6)GalNAc-Sp14
315	Neu5Aca2-3Galb1-3GalNAc-Sp14
316	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
317	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
318	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
319	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-N(LT)AVL
320	Fuca1-2Galb1-3GalNAca-Sp14
321	Galb1-3(Neu5Aca2-6)GalNAca-Sp14
322	Galb1-4GlcNAcb1-3GalNAc-Sp14
323	NeuAc(9Ac)a2-3Galb1-4GlcNAcb-Sp0
324	NeuAc(9Ac)a2-3Galb1-3GlcNAcb-Sp0
325	NeuAca2-6Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0
326	NeuAca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
327	NeuAca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
328	Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0
329	GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0
330	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
331	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
332	(Neu5Aca2-3-Galb1-3)(((Neu5Aca2-3-Galb1-4(Fuca1-3))GlcNAcb1-6)GalNAc-Sp14
333	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
334	GlcNAca1-4Galb1-4GlcNAcb-Sp0
335	GlcNAca1-4Galb1-3GlcNAcb-Sp0
336	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0
337	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0

Number	Glycan
338	GlcNAc1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
339	GlcNAc1-4Galb1-3GalNAc-Sp14
340	Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12
341	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12
342	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12
343	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12
344	Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12
345	Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12
346	Galb1-4GlcNAcb1-2Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
347	GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
348	Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
349	Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
350	Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3[Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp19
351	(6SO3)GlcNAcb1-3Gal b1-4GlcNAc-b-Sp0
352	KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0
353	KDNa2-6Galb1-4GlcNAc-Sp0
354	KDNa2-3Galb1-4Glc-Sp0
355	KDNa2-3Galb1-3GalNAc-Sp14
356	Fuca1-2Galb1-3GlcNAcb1-2Mana1-3(Fuca1-2Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
357	Fuca1-2Galb1-4GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
358	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3[Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp20
359	Gala1-3Galb1-4GlcNAcb1-2Mana1-3(Gala1-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
360	Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
361	Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3[Galb1-3(Fuca1-4)GlcNAcb1-2Man a1-6]Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
362	Neu5Aca2-6GlcNAcb1-4GlcNAc-Sp21

Number	Glycan
363	Neu5Aca2-6GlcNAcb1-4GlcNAcb1-4GlcNAc-Sp21
364	Fuca1-2Galb1-3GlcNAcb1-3[Galb1-4(Fuca1-3)GlcNAcb1-6]Galb1-4Glc-Sp21
365	Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3[Galb1-4GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAc-Sp21
366	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3[GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp20
367	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3[Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp20
368	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3[Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp20
369	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3[GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp20
370	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3[Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp20
371	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3[Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp19
372	NeuAca2-3Galb1-4GlcNAcb1-3GalNAc-Sp14
373	NeuAca2-6Galb1-4GlcNAcb1-3GalNAc-Sp14
374	NeuAca2-3[Fuca1-3]Galb1-4GlcNAcb1-3GalNAc-Sp14
375	GalNAcb1-4GlcNAcb1-2Mana1-6(GalNAcb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12
376	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp14
377	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAc-Sp14

Table 9.1: Glycan Array V3.1

## 9.1.2 Glycan array V4.1

Number	Glycan
1	Gala-Sp8
2	Glca-Sp8
3	Mana-Sp8
4	GalNAca-Sp8
5	GalNAca-Sp15
6	Fuca-Sp8
7	Fuca-Sp9
8	Rhaa-Sp8
9	Neu5Aca-Sp8
10	Neu5Aca-Sp11
11	Neu5Acb-Sp8
12	Galb-Sp8
13	Glcb-Sp8
14	Manb-Sp8
15	GalNAcb-Sp8
16	GlcNAcb-Sp0
17	GlcNAcb-Sp8
18	GlcN(Gc)b-Sp8
19	Galb1-4GlcNAcb1-3(Galb1-4GlcNAcb1-6)GalNAca-Sp8
20	GlcNAcb1-3(GlcNAcb1-4)(GlcNAcb1-6)GlcNAc-Sp8
21	[3OSO3][6OSO3]Galb1-4[6OSO3]GlcNAcb-Sp0
22	[3OSO3][6OSO3]Galb1-4GlcNAcb-Sp0
23	[3OSO3]Galb1-4(Fuca1-3)[6OSO3]Glc-Sp0
24	[3OSO3]Galb1-4Glcb-Sp8
25	[3OSO3]Galb1-4[6OSO3]Glcb-Sp0
26	[3OSO3]Galb1-4[6OSO3]Glcb-Sp8
27	[3OSO3]Galb1-3(Fuca1-4)GlcNAcb-Sp8

Number	Glycan
28	[3OSO3]Galb1-3GalNAca-Sp8
29	[3OSO3]Galb1-3GlcNAcb-Sp0
30	[3OSO3]Galb1-3GlcNAcb-Sp8
31	[3OSO3]Galb1-4(Fuca1-3)GlcNAc-Sp0
32	[3OSO3]Galb1-4(Fuca1-3)GlcNAcb-Sp8
33	[3OSO3]Galb1-4[6OSO3]GlcNAcb-Sp0
34	[3OSO3]Galb1-4[6OSO3]GlcNAcb-Sp8
35	[3OSO3]Galb1-4GlcNAcb-Sp0
36	[3OSO3]Galb1-4GlcNAcb-Sp8
37	[3OSO3]Galb-Sp8
38	[4OSO3][6OSO3]Galb1-4GlcNAcb-Sp0
39	[4OSO3]Galb1-4GlcNAcb-Sp8
40	6-H2PO3Mana-Sp8
41	[6OSO3]Galb1-4Glc-Sp0
42	[6OSO3]Galb1-4Glc-Sp8
43	[6OSO3]Galb1-4GlcNAcb-Sp8
44	[6OSO3]Galb1-4[6OSO3]Glc-Sp8
45	Neu5Aca2-3[6OSO3]Galb1-4GlcNAcb-Sp8
46	[6OSO3]GlcNAcb-Sp8
47	[9NAc]Neu5Aca-Sp8
48	[9NAc]Neu5Aca2-6Galb1-4GlcNAcb-Sp8
49	Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
50	Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp13
51	GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
52	GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp13
53	Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
54	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-N(LT)AVL
55	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp8
56	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
57	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp13



Number	Glycan
58	Fuca1-2Galb1-3GalNAcb1-3Gala-Sp9
59	Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp9
60	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb-Sp8
61	Fuca1-2Galb1-3GalNAca-Sp8
62	Fuca1-2Galb1-3GalNAca-Sp14
63	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp0
64	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp9
65	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glc-Sp8
66	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glc-Sp10
67	Fuca1-2Galb1-3GlcNAcb-Sp0
68	Fuca1-2Galb1-3GlcNAcb-Sp8
69	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
70	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
71	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp0
72	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp8
73	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
74	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
75	Fuca1-2Galb1-4GlcNAcb-Sp0
76	Fuca1-2Galb1-4GlcNAcb-Sp8
77	Fuca1-2Galb1-4Glc-Sp0
78	Fuca1-2Galb-Sp8
79	Fuca1-3GlcNAcb-Sp8
80	Fuca1-4GlcNAcb-Sp8
81	Fucb1-3GlcNAcb-Sp8
82	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0
83	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0
84	[3OSO3]Galb1-4(Fuca1-3)Glc-Sp0
85	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0
86	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8
87	GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0
88	GlcNAcb1-3Galb1-3GalNAca-Sp8
89	GalNAca1-3(Fuca1-2)Galb-Sp8
90	GalNAca1-3(Fuca1-2)Galb-Sp18

Number	Glycan
91	GalNAca1-3GalNAcb-Sp8
92	GalNAca1-3Galb-Sp8
93	GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8
94	GalNAcb1-3GalNAca-Sp8
95	GalNAcb1-3(Fuca1-2)Galb-Sp8
96	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0
97	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0
98	GalNAcb1-4GlcNAcb-Sp0
99	GalNAcb1-4GlcNAcb-Sp8
100	Gala1-2Galb-Sp8
101	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0
102	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp8
103	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0
104	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp8
105	Gala1-3(Fuca1-2)Galb1-4GlcNAc-Sp0
106	Gala1-3(Fuca1-2)Galb1-4Glc-Sp0
107	Gala1-3(Fuca1-2)Galb-Sp8
108	Gala1-3(Fuca1-2)Galb-Sp18
109	Gala1-3(Gala1-4)Galb1-4GlcNAcb-Sp8
110	Gala1-3GalNAca-Sp8
111	Gala1-3GalNAca-Sp16
112	Gala1-3GalNAcb-Sp8
113	Gala1-3Galb1-4(Fuca1-3)GlcNAcb-Sp8
114	Gala1-3Galb1-3GlcNAcb-Sp0
115	Gala1-3Galb1-4GlcNAcb-Sp8
116	Gala1-3Galb1-4Glc-Sp0
117	Gala1-3Galb-Sp8
118	Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8
119	Gala1-4Galb1-4GlcNAcb-Sp0
120	Gala1-4Galb1-4GlcNAcb-Sp8
121	Gala1-4Galb1-4Glc-Sp0
122	Gala1-4GlcNAcb-Sp8
123	Gala1-6Glc-Sp8

Number	Glycan
124	Galb1-2Galb-Sp8
125	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
126	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4GlcNAcb-Sp0
127	Galb1-3(Fuca1-4)GlcNAcb-Sp0
128	Galb1-3(Fuca1-4)GlcNAcb-Sp8
129	Galb1-3(Fuca1-4)GlcNAc-Sp8
130	Galb1-4GlcNAcb1-6GalNAca-Sp8
131	Galb1-3(GlcNAcb1-6)GalNAca-Sp8
132	Galb1-3(GlcNAcb1-6)GalNAc-Sp14
133	Galb1-3(Neu5Aca2-6)GalNAca-Sp8
134	Galb1-3(Neu5Aca2-6)GalNAca-Sp14
135	Galb1-3(Neu5Acb2-6)GalNAca-Sp8
136	Galb1-3(Neu5Aca2-6)GlcNAcb1-4Galb1-4Glc-Sp10
137	Galb1-3GalNAca-Sp8
138	Galb1-3GalNAca-Sp14
139	Galb1-3GalNAca-Sp16
140	Galb1-3GalNAcb-Sp8
141	Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0
142	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp0
143	Galb1-3GalNAcb1-4Galb1-4Glc-Sp8
144	Galb1-3Galb-Sp8
145	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0
146	Galb1-3GlcNAcb1-3Galb1-4Glc-Sp10
147	Galb1-3GlcNAcb-Sp0
148	Galb1-3GlcNAcb-Sp8
149	Galb1-4(Fuca1-3)GlcNAcb-Sp0
150	Galb1-4(Fuca1-3)GlcNAcb-Sp8
151	Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb-Sp0
152	Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb-Sp0
153	Galb1-4[6OSO3]Glc-Sp0
154	Galb1-4[6OSO3]Glc-Sp8
155	Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8
156	Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8

Number	Glycan
157	Galb1-4GlcNAcb1-3GalNAca-Sp8
158	Galb1-4GlcNAcb1-3GalNAca-Sp14
159	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
160	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
161	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
162	Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0
163	Galb1-4GlcNAcb1-3Galb1-4Glc-Sp8
164	Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp8
165	Galb1-3(Galb1-4GlcNAcb1-6)GalNAca-Sp8
166	Galb1-3(Galb1-4GlcNAcb1-6)GalNAc-Sp14
167	Galb1-4GlcNAcb-Sp0
168	Galb1-4GlcNAcb-Sp8
169	Galb1-4Glc-Sp0
170	Galb1-4Glc-Sp8
171	GlcNAca1-3Galb1-4GlcNAcb-Sp8
172	GlcNAca1-6Galb1-4GlcNAcb-Sp8
173	GlcNAcb1-2Galb1-3GalNAca-Sp8
174	GlcNAcb1-3(GlcNAcb1-6)GalNAca-Sp8
175	GlcNAcb1-3(GlcNAcb1-6)Galb1-4GlcNAcb-Sp8
176	GlcNAcb1-3GalNAca-Sp8
177	GlcNAcb1-3GalNAca-Sp14
178	GlcNAcb1-3Galb-Sp8
179	GlcNAcb1-3Galb1-4GlcNAcb-Sp0
180	GlcNAcb1-3Galb1-4GlcNAcb-Sp8
181	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
182	GlcNAcb1-3Galb1-4Glc-Sp0
183	GlcNAcb1-4-MDPLys
184	GlcNAcb1-4(GlcNAcb1-6)GalNAca-Sp8
185	GlcNAcb1-4Galb1-4GlcNAcb-Sp8
186	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4b-Sp8
187	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8
188	GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8
189	GlcNAcb1-6(Galb1-3)GalNAca-Sp8

Number	Glycan
190	GlcNAc1-6GalNAc1-Sp8
191	GlcNAc1-6GalNAc1-Sp14
192	GlcNAc1-6Gal1-4GlcNAc1-Sp8
193	Glc1-4Glc1-Sp8
194	Glc1-4Glc1-Sp8
195	Glc1-6Glc1-6Glc1-Sp8
196	Glc1-4Glc1-Sp8
197	Glc1-6Glc1-Sp8
198	G-ol-Sp8
199	GlcAa-Sp8
200	GlcAb-Sp8
201	GlcAb1-3Gal1-Sp8
202	GlcAb1-6Gal1-Sp8
203	KDNa2-3Gal1-3GlcNAc1-Sp0
204	KDNa2-3Gal1-4GlcNAc1-Sp0
205	Mana1-2Mana1-2Mana1-3Mana1-Sp9
206	Mana1-2Mana1-3(Mana1-2Mana1-6)Mana1-Sp9
207	Mana1-2Mana1-3Mana1-Sp9
208	Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAc1-4GlcNAc1-Sp12
209	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAc1-4GlcNAc1-Sp12
210	Mana1-2Mana1-2Mana1-3(Mana1-2Mana1-3(Mana1-2Mana1-6)Mana1-6)Manb1-4GlcNAc1-4GlcNAc1-Sp12
211	Mana1-3(Mana1-6)Mana1-Sp9
212	Mana1-3(Mana1-2Mana1-2Mana1-6)Mana1-Sp9
213	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAc1-4GlcNAc1-Sp12
214	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAc1-4GlcNAc1-Sp12
215	Manb1-4GlcNAc1-Sp0
216	Neu5Aca2-3Gal1-4GlcNAc1-3Gal1-4(Fuca1-3)GlcNAc1-Sp0
217	[3OSO3]Gal1-4(Fuca1-3)[6OSO3]GlcNAc1-Sp8
218	Fuca1-2[6OSO3]Gal1-4GlcNAc1-Sp0
219	Fuca1-2Gal1-4[6OSO3]GlcNAc1-Sp8
220	Fuca1-2[6OSO3]Gal1-4[6OSO3]GlcNAc1-Sp0

Number	Glycan
221	Neu5Aca2-3Galb1-3GalNAca-Sp8
222	Neu5Aca2-3Galb1-3GalNAca-Sp14
223	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcB-Sp0
224	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcB-Sp0
225	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4GlcB-Sp0
226	Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcB-Sp0
227	Neu5Aca2-8Neu5Aca2-8Neu5Aca-Sp8
228	Neu5Aca2-3(6-O-Su)Galb1-4(Fuca1-3)GlcNAcb-Sp8
229	Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb-Sp0
230	Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb-Sp8
231	Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcB-Sp0
232	Neu5Aca2-3(Neu5Aca2-3Galb1-3GalNAcb1-4)Galb1-4GlcB-Sp0
233	Neu5Aca2-3(Neu5Aca2-6)GalNAca-Sp8
234	Neu5Aca2-3GalNAca-Sp8
235	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0
236	Neu5Aca2-3Galb1-3[6OSO3]GlcNAc-Sp8
237	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp8
238	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
239	Neu5Aca2-3Galb1-3(Neu5Aca2-3Galb1-4)GlcNAcb-Sp8
240	Neu5Aca2-3Galb1-3[6OSO3]GalNAca-Sp8
241	Neu5Aca2-3Galb1-3(Neu5Aca2-6)GalNAca-Sp8
242	Neu5Aca2-3Galb1-3(Neu5Aca2-6)GalNAca-Sp14
243	Neu5Aca2-3Galb-Sp8
244	Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4GlcB-Sp0
245	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0
246	Fuca1-2[6OSO3]Galb1-4Glc-Sp0
247	Neu5Aca2-3Galb1-3GlcNAcb-Sp0
248	Neu5Aca2-3Galb1-3GlcNAcb-Sp8
249	Neu5Aca2-3Galb1-4[6OSO3]GlcNAcb-Sp8
250	Neu5Aca2-3Galb1-4(Fuca1-3)[6OSO3]GlcNAcb-Sp8
251	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
252	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
253	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp8

Number	Glycan
254	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8
255	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8
256	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
257	Neu5Aca2-3Galb1-4GlcNAcb-Sp0
258	Neu5Aca2-3Galb1-4GlcNAcb-Sp8
259	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
260	Fuca1-2Galb1-4[6OSO3]Glc-Sp0
261	Neu5Aca2-3Galb1-4Glc-Sp0
262	Neu5Aca2-3Galb1-4Glc-Sp8
263	Neu5Aca2-6GalNAca-Sp8
264	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0
265	Neu5Aca2-6Galb1-4[6OSO3]GlcNAcb-Sp8
266	Neu5Aca2-6Galb1-4GlcNAcb-Sp0
267	Neu5Aca2-6Galb1-4GlcNAcb-Sp8
268	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
269	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
270	Neu5Aca2-6Galb1-4Glc-Sp0
271	Neu5Aca2-6Galb1-4Glc-Sp8
272	Neu5Aca2-6Galb-Sp8
273	Neu5Aca2-8Neu5Aca-Sp8
274	Neu5Aca2-8Neu5Aca2-3Galb1-4Glc-Sp0
275	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
276	Neu5Acb2-6GalNAca-Sp8
277	Neu5Acb2-6Galb1-4GlcNAcb-Sp8
278	Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
279	Neu5Gca2-3Galb1-3GlcNAcb-Sp0
280	Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
281	Neu5Gca2-3Galb1-4GlcNAcb-Sp0
282	Neu5Gca2-3Galb1-4Glc-Sp0
283	Neu5Gca2-6GalNAca-Sp0
284	Neu5Gca2-6Galb1-4GlcNAcb-Sp0
285	Neu5Gca-Sp8
286	Galb1-3(Neu5Aca2-3Galb1-4GlcNAcb1-6)GalNAca-Sp14

Number	Glycan
287	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0
288	Galb1-4(Fuca1-3)[6OSO3]GlcNAc-Sp0
289	Galb1-4(Fuca1-3)[6OSO3]Glc-Sp0
290	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
291	Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0
292	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0
293	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0
294	[3OSO3][4OSO3]Galb1-4GlcNAcb-Sp0
295	[6OSO3]Galb1-4[6OSO3]GlcNAcb-Sp0
296	6-H2PO3Glc-Sp10
297	Galb1-3(Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6)GalNAca-Sp14
298	Galb1-3Galb1-4GlcNAcb-Sp8
299	Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
300	Galb1-4GlcNAcb1-3(Galb1-4GlcNAcb1-6)Galb1-4GlcNAc-Sp0
301	Galb1-4GlcNAcb1-3(GlcNAcb1-6)Galb1-4GlcNAc-Sp0
302	Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0
303	Galb1-4GlcNAcb1-6Galb1-4GlcNAcb-Sp0
304	GalNAcb1-3Galb-Sp8
305	GlcAb1-3GlcNAcb-Sp8
306	GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
307	GlcNAcb1-3Man-Sp10
308	GlcNAcb1-4GlcNAcb-Sp10
309	GlcNAcb1-4GlcNAcb-Sp12
310	HOOC(CH3)CH-3-O-GlcNAcb1-4GlcNAcb-Sp10
311	Mana1-6Manb-Sp10
312	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb-Sp10
313	Mana1-2Mana1-2Mana1-3(Mana1-2Mana1-6(Mana1-3)Mana1-6)Mana-Sp9
314	Mana1-2Mana1-2Mana1-3(Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6)Mana-Sp9
315	Neu5Aca2-3Galb1-3(Neu5Aca2-3Galb1-4GlcNAcb1-6)GalNAca-Sp14
316	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
317	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-



Number	Glycan
	4GlcNAcb1-4GlcNAcb-Sp12
318	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
319	Neu5Aca2-8Neu5Acb-Sp17
320	Neu5Aca2-8Neu5Aca2-8Neu5Acb-Sp8
321	Neu5Gcb2-6Galb1-4GlcNAc-Sp8
322	Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
323	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
324	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
325	Fuca1-3(Galb1-4)GlcNAcb1-2Mana1-3(Fuca1-3(Galb1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
326	Neu5Ac(9Ac)a2-3Galb1-4GlcNAcb-Sp0
327	Neu5Ac(9Ac)a2-3Galb1-3GlcNAcb-Sp0
328	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0
329	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0
330	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
331	Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0
332	GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0
333	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
334	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
335	Neu5Aca2-3-Galb1-3(Galb1-4(Fuca1-3)GlcNAcb1-6)GalNAc-Sp14
336	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
337	GlcNAca1-4Galb1-4GlcNAcb-Sp0
338	GlcNAca1-4Galb1-3GlcNAcb-Sp0
339	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0
340	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0
341	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0
342	GlcNAca1-4Galb1-3GalNAc-Sp14
343	Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12
344	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12

Number	Glycan
345	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12
346	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12
347	Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12
348	Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12
349	Galb1-4GlcNAcb1-2Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
350	GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
351	Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
352	Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
353	Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
354	[6OSO3]GlcNAcb1-3Galb1-4GlcNAc-b-Sp0
355	KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0
356	KDNa2-6Galb1-4GlcNAc-Sp0
357	KDNa2-3Galb1-4Glc-Sp0
358	KDNa2-3Galb1-3GalNAca-Sp14
359	Fuca1-2Galb1-3GlcNAcb1-2Mana1-3(Fuca1-2Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
360	Fuca1-2Galb1-4GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
361	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
362	Gala1-3Galb1-4GlcNAcb1-2Mana1-3(Gala1-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
363	Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
364	Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
365	Neu5Aca2-6GlcNAcb1-4GlcNAc-Sp21
366	Neu5Aca2-6GlcNAcb1-4GlcNAcb1-4GlcNAc-Sp21
367	Fuca1-2Galb1-3GlcNAcb1-3(Galb1-4(Fuca1-3)GlcNAcb1-6)Galb1-4Glc-Sp21
368	Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
369	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20

Number	Glycan
370	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
371	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
372	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
373	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
374	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
375	Neu5Aca2-3Galb1-4GlcNAcb1-3GalNAc-Sp14
376	Neu5Aca2-6Galb1-4GlcNAcb1-3GalNAc-Sp14
377	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14
378	(GalNAcb1-4GlcNAcb1-2Mana1-6)GalNAcb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12
379	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0
380	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAc-Sp0
381	Galb1-3GlcNAcb1-3(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6)Galb1-4Glc-Sp0
382	Galb1-3GlcNAcb1-3(Galb1-4(Fuca1-3)GlcNAcb1-6)Galb1-4Glc-Sp21
383	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-3(Galb1-4GlcNAcb1-6)Galb1-4Glc-Sp21
384	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-3(Galb1-4(Fuca1-3)GlcNAcb1-6)Galb1-4Glc-Sp21
385	Galb1-3GlcNAcb1-3(Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6)Galb1-4Glc-Sp21
386	Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3(Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-6)Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp21
387	GlcNAcb1-2(GlcNAcb1-4)Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
388	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0
389	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0
390	Galb1-3GlcNAcb1-3GalNAca-Sp14
391	Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb1-3GalNAca-Sp14
392	GalNAca1-3(Fuca1-2)Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0
393	Gala1-3Galb1-3GlcNAcb1-2Mana1-3(Gala1-3Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp19
394	Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp19

Number	Glycan
395	Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12
396	GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12
397	Neu5Aca2-3Galb1-3GlcNAcb1-3GalNAca-Sp14
398	Fuca1-2Galb1-4GlcNAcb1-3GalNAca-Sp14
399	Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14
400	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0
401	Gala1-4Galb1-3GlcNAcb1-2Mana1-3(Gala1-4Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
402	Gala1-4Galb1-4GlcNAcb1-2Mana1-3(Gala1-4Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-LVaNKT
403	Gala1-3Galb1-4GlcNAcb1-3GalNAca-Sp14
404	Galb1-3GlcNAcb1-6Galb1-4GlcNAcb-Sp0
405	Galb1-3GlcNAca1-6Galb1-4GlcNAcb-Sp0
406	GalNAcb1-3Gala1-6Galb1-4Glc-Sp8
407	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp14
408	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)Glc-Sp21
409	Neu5Aca2-6Galb1-3GlcNAcb1-3(Galb1-4GlcNAcb1-6)Galb1-4Glc-Sp21
410	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glc-Sp0
411	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glc-Sp0
412	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14
413	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14
414	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0
415	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14
416	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14
417	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14
418	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
419	Fuca1-2Galb1-4GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
420	GlcNAcb1-2Mana1-3(GlcNAcb1-2(GlcNAcb1-6)Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
421	Fuca1-2Galb1-3GlcNAcb1-3GalNAc-Sp14
422	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14

Number	Glycan
423	GalNAc1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14
424	Gala1-3Galb1-3GlcNAcb1-3GalNAc-Sp14
425	Fuca1-2Galb1-3GlcNAcb1-2Mana1-3(Fuca1-2Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
426	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
427	Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2(Galb1-3GlcNAcb1-6)Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
428	Fuca1-2Galb1-3GlcNAcb1-3(Galb1-4GlcNAcb1-6)Galb1-4Glc-Sp21
429	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3GlcNAcb1-6)Galb1-4Glc-Sp21
430	GlcNAcb1-2Mana1-3(GlcNAcb1-4)(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
431	GlcNAcb1-4(GlcNAcb1-2)Mana1-3(GlcNAcb1-4)(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
432	GlcNAcb1-2Mana1-3(GlcNAcb1-4)(GlcNAcb1-6(GlcNAcb1-2)Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
433	GlcNAcb1-4(GlcNAcb1-2)Mana1-3(GlcNAcb1-4)(GlcNAcb1-6(GlcNAcb1-2)Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
434	Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-4)(Galb1-4GlcNAcb1-2)Manb1-4GlcNAcb1-4GlcNAc-Sp21
435	Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
436	Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-4)(Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
437	Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3(GlcNAcb1-4)(Galb1-4GlcNAcb1-6(Galb1-4-GlcNAcb1-2)Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
438	Gala1-3Galb1-4Glc-Sp10
439	Galb1-4Galb-Sp10
440	Galb1-6Galb-Sp10
441	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb-Sp8
442	GalNAcb1-6GalNAcb-Sp8
443	[6OSO3]Galb1-3GlcNAcb-Sp0
444	[6OSO3]Galb1-3[6OSO3]GlcNAc-Sp0
445	Fuca1-2Galb1-4GlcNAcb1-2(Fuca1-2Galb1-4GlcNAcb1-4)Mana1-3(Fuca1-2Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
446	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-4)Mana1-3(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-

Number	Glycan
	Sp12
447	Galb1-4GlcNAcb1-3(Galb1-4GlcNAcb1-6)GalNAc-Sp14
448	Galb1-4GlcNAcb1-6GalNAc-Sp14
449	Galb1-4(Fuca1-3)GlcNAcb1-6GalNAc-Sp14
450	Galb1-4GlcNAcb1-2Mana-Sp0
451	Fuca1-2Galb1-4GlcNAcb1-3(Fuca1-2Galb1-4GlcNAcb1-6)GalNAc-Sp14
452	Gala1-3Fuca1-2Galb1-4GlcNAcb1-3(Gala1-3Fuca1-2Galb1-4GlcNAcb1-6)GalNAc-Sp14
453	GalNAca1-3Fuca1-2Galb1-4GlcNAcb1-3(GalNAca1-3Fuca1-2Galb1-4GlcNAcb1-6)GalNAc-Sp14
454	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glc-b-sp0
455	GalNAcb1-4Galb1-4Glc-b-sp0
456	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
457	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
458	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb-4Glc-Sp21
459	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
460	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
461	Galb1-4GlcNAcb-(OCH <sub>2</sub> CH <sub>2</sub> ) <sub>6</sub> NH <sub>2</sub>
462	Gala1-3(Fuca1-2)Galb1-3GalNAca-Sp8
463	Gala1-3(Fuca1-2)Galb1-3GalNAcb-Sp8
464	Glea1-6Glea1-6Glea1-6Glc-b-Sp10
465	Glea1-4Glea1-4Glea1-4Glc-b-Sp10

Table 9.2 Glycan Array V4.1

## 9.2 Appendix II: Protein sequences

The sequences presented here are the ones obtained from sequencing of the generated constructs, and not from database searches.

### 9.2.1 Sequences of *Epa1A* and variants

#### 9.2.1.1 Protein sequence from *Epa1A*

```

      10      20      30      40      50      60
MGSSHHHHHH SSGLVPRGSH MTSSNDISLA SKDPTTFPLG CSPDITTPKK GLSMELYSYD

      70      80      90     100     110     120
FRKKGSYPCW DAAYLDPNYP RTGYKSHRLL AKVDGVTGNI NFYYHATKGC TPQLGHLPAS

     130     140     150     160     170     180
YNYPKPLTMT NFTMLLYGYF RPKVTGFHTF TISADDLLFV NFGAGNAFDC CRRDSSADHF

     190     200     210     220     230     240
GNYQAYAIWG SKTAKDELTV HLDAGVYYPI RLFYNNREYD GALSFTFKTE SNENTVSDFS

     250     260
EYFFSLDDTE EGCPGLISYD SS

```

#### 9.2.1.2 Protein sequence from *Epa1*→*2A*

```

      10      20      30      40      50      60
MGSSHHHHHH SSGLVPRGSH MTSSNDISLA SKDPTTFPLG CSPDITTPKK GLSMELYSYD

      70      80      90     100     110     120
FRKKGSYPCW DAAYLDPNYP RTGYKSHRLL AKVDGVTGNI NFYYHATKGC TPQLGHLPAS

     130     140     150     160     170     180
YNYPKPLTMT NFTMLLYGYF RPKVTGFHTF TISADDLLFV NFGAGNAFDC CRRDSSADHF

     190     200     210     220     230     240
GNYQAYAIWG SKTAKDELTV HLDAGVYYPI RLFYNNRDNN GALSFTFKTE SNENTVSDFS

     250     260
EYFFSLDDTE EGCPGLISYD SS

```

**9.2.1.3 Protein sequence from *Epa1*→3A**

10            20            30            40            50            60  
 MGSSHHHHHHH S SGLVPRGSH MTSSNDISLA SKDPTTFPLG CSPDITTPKK GLSMELYSYD  
  
70            80            90            100            110            120  
 FRKKGSYPCW DAAYLDPNYP RTGYKSHRLL AKVDGVTGNI NFYYHATKGC TPQLGHLPAS  
  
130            140            150            160            170            180  
 YNYPKPLTMT NFTMLLYGYF RPKVTGFHTF TISADDLLFV NFGAGNAFDC CRRDSSADHF  
  
190            200            210            220            230            240  
 GNYQAYAIWG SKTAKDELTV HLDAGVYYPI RLFYNNIGKD GALSFTFKTE SNENTVSDFS  
  
250            260  
 EYFFSLDDTE EGCPGLISYD SS

**9.2.1.4 Protein sequence from *Epa1*→6A**

10            20            30            40            50            60  
 MGSSHHHHHHH S SGLVPRGSH MTSSNDISLA SKDPTTFPLG CSPDITTPKK GLSMELYSYD  
  
70            80            90            100            110            120  
 FRKKGSYPCW DAAYLDPNYP RTGYKSHRLL AKVDGVTGNI NFYYHATKGC TPQLGHLPAS  
  
130            140            150            160            170            180  
 YNYPKPLTMT NFTMLLYGYF RPKVTGFHTF TISADDLLFV NFGAGNAFDC CRRDSSADHF  
  
190            200            210            220            230            240  
 GNYQAYAIWG SKTAKDELTV HLDAGVYYPI RLFYNNRDND GALSFTFKTE SNENTVSDFS  
  
250            260  
 EYFFSLDDTE EGCPGLISYD SS



## 9.2.2 Sequences of BabA extracellular fragments from *H. pylori* 26695

### 9.2.2.1 BabA 235

10                    20                    30                    40                    50                    60  
 MEDDGFYTSV GYQIGEEAAQM VTNTKGIQDL SDNYENLSKL LTRYSTLNTL IKLSADPSAI  
  
70                    80                    90                    100                    110                    120  
 NAARENLGAS AKNLIGDTKN SPAYQAVLLA INAAVGFWNV LGYATQCGGN ANGQESTSST  
  
130                    140                    150                    160                    170                    180  
 TIFNNEPGYR STSITCSLNR YKPGYYGPMS IENFKKLNEA YQILQTALNK GLPALKENNG  
  
190                    200                    210                    220                    230                    240  
 TVSVTYTYTC SGEGNDNCSK KATGVSDQNG GTKTKTQTID GKTVTTTTISS KVVDSGLLPR  
  
 GSHHHHHHH

### 9.2.2.2 BabA 210

10                    20                    30                    40                    50                    60  
 MEDDGFYTSV GYQIGEEAAQM VTNTKGIQDL SDNYENLSKL LTRYSTLNTL IKLSADPSAI  
  
70                    80                    90                    100                    110                    120  
 NAARENLGAS AKNLIGDTKN SPAYQAVLLA INAAVGFWNV LGYATQCGGN ANGQESTSST  
  
130                    140                    150                    160                    170                    180  
 TIFNNEPGYR STSITCSLNR YKPGYYGPMS IENFKKLNEA YQILQTALNK GLPALKENNG  
  
190                    200                    210                    220  
 TVSVTYTYTC SGEGNDNCSK KATGVSDQNG GLLPRGSHHH HHHH

### 9.2.2.3 BabA 144

10                    20                    30                    40                    50                    60  
 MEDDGFYTSV GYQIGEEAAQM VTNTKGIQDL SDNYENLSKL LTRYSTLNTL IKLSADPSAI  
  
70                    80                    90                    100                    110                    120  
 NAARENLGAS AKNLIGDTKN SPAYQAVLLA INAAVGFWNV LGYATQCGGN ANGQESTSST  
  
130                    140                    150  
 TIFNNEPGYR STSITCSLNR YKPGGLLPRG SHHHHHHH

### 9.2.3 Sequences of BabA extracellular fragments from *H. pylori* J99

Sequences presented here are as they were cloned into the pET28a system.

#### 9.2.3.1 BabA 210

```

      10           20           30           40           50           60
MGEDDGFYTS VGYQIGEEAAQ MVTNTKGIQD LSDRYESLNN LLNRYSTLNT LIKLSADPSA

      70           80           90           100          110          120
INAVRENLGA SAKNLIGDKA NSPAYQAVLL AINAAVGFWN VVGYVTQCGG NANGQKSISS

      130          140          150          160          170          180
KTIFNNEPGY RSTSITCSLN GHSPGYYGPM SIENFKKLNE AYQILQTALK RGLPALKENN

      190          200          210
GKVNVTYTYT CSGDGNNNCS SQVTGVNNQK DGEHHHHHHH

```

#### 9.2.3.2 BabA235

```

      10           20           30           40           50           60
MGEDDGFYTS VGYQIGEEAAQ MVTNTKGIQD LSDRYESLNN LLNRYSTLNT LIKLSADPSA

      70           80           90           100          110          120
INAVRENLGA SAKNLIGDKA NSPAYQAVLL AINAAVGFWN VVGYVTQCGG NANGQKSISS

      130          140          150          160          170          180
KTIFNNEPGY RSTSITCSLN GHSPGYYGPM SIENFKKLNE AYQILQTALK RGLPALKENN

      190          200          210          220          230          240
GKVNVTYTYT CSGDGNNNCS SQVTGVNNQK DGTKTKIQTI DGKSVTTTIS SKVVDREHH

HHHH

```

#### 9.2.3.3 BabA 255

```

      10           20           30           40           50           60
MGEDDGFYTS VGYQIGEEAAQ MVTNTKGIQD LSDRYESLNN LLNRYSTLNT LIKLSADPSA

      70           80           90           100          110          120
INAVRENLGA SAKNLIGDKA NSPAYQAVLL AINAAVGFWN VVGYVTQCGG NANGQKSISS

      130          140          150          160          170          180
KTIFNNEPGY RSTSITCSLN GHSPGYYGPM SIENFKKLNE AYQILQTALK RGLPALKENN

      190          200          210          220          230          240
GKVNVTYTYT CSGDGNNNCS SQVTGVNNQK DGTKTKIQTI DGKSVTTTIS SKVVDSRADG

```

250                      260  
 NTTGVSYTEI TNKLEGVPEH HHHHH

Primer name	Primer sequence	T <sub>m</sub> (°C)
Primers yielding constructs with C-terminal His-tags		
Forward universal	GTT GT CCATGGGGGAAGACGACGGCTTTTACAC	66
Reverse 210	GGGGGCTCGAG TCCGTCTTTTTGATTATTTACACCTG	64.7
Reverse 235	GGGGGCTCGAG ACGACTATCAACCACTTTTGAAC	65.3
Reverse 255	GGGGCCTC GAG CACA CCT TCT AAT TTG TTG GTG	64.7
Primers yielding constructs with N-terminal His-tags		
Forward universal	GACTAGCATATGGAAGACGACGGCTTTTACAC	61.5
Reverse 210	GGCGGCCTCGAGCTATCCGTCTTTTTGATTATTTACACC	65.6
Reverse 235	GGCGTGCTCGAGCTAACGACTATCAACCACTTTTGAAC	65.9
Reverse 255	GGGGGCCTCGAGCTACACACCTTCTAATTTGTTGGT	67.1

Table 9.3: Primers used for the cloning of BabA extracellular fragments from *H. pylori* J99. All constructs, including fusion constructs, could be generated with these primers. Blue: Stop codon. Green: Restriction site. Red: required additional base pairs.

#### 9.2.4 Sequences of BabA extracellular fragments from *H. pylori* P12

All sequences presented as a result from cloning into the pET28 system.

##### 9.2.4.1 BabA 210

10                      20                      30                      40                      50                      60  
 MGEDDGFYTS VGYQIGEEAQ MVTNTKGIQE LSDNYEKLNN LLNNYSTLNT LIKLSADPSA

70                      80                      90                      100                      110                      120  
 VNAARENLGA SAKNLIGDTK NSPAYQAVLL AINAAVGFWN VVG YVTQCGG NANGQTSTSS

130                      140                      150                      160                      170                      180  
 TTIFNNEPGY RSTSITCSLN GYTPGYGPM SIENFKKLE AYQILQTALK QGLPALKENN

190                      200                      210  
 GTVSVTYTYT CSGKGNDCS PQTGTVNQON DGEHHHHHH

**9.2.4.2 *Baba 235***

10                    20                    30                    40                    50                    60  
 MGEDDGFYTS VGYQIGEEAAQ MVTNTKGIQE LSDNYEKLNN LLNNYSTLNT LIKLSADPSA  
  
70                    80                    90                    100                    110                    120  
 VNAARENLGA SAKNLIGDTK NSPAYQAVLL AINAAVGFWN VVGYVTQCGG NANGQTSTSS  
  
130                    140                    150                    160                    170                    180  
 TTIFNNEPGY RSTSITCSLN GYTPGYYGPM SIENFKKLNE AYQILQTALK QGLPALKENN  
  
190                    200                    210                    220                    230                    240  
 GTVSVTYTYT CSGKGNDCS PQTTGVNQON DGTKTETQTI DGKQVNTTIS SKVVDSKLEH  
  
 HHHH

**9.2.4.3 *Baba 255***

10                    20                    30                    40                    50                    60  
 MGEDDGFYTS VGYQIGEEAAQ MVTNTKGIQE LSDNYEKLNN LLNNYSTLNT LIKLSADPSA  
  
70                    80                    90                    100                    110                    120  
 VNAARENLGA SAKNLIGDTK NSPAYQAVLL AINAAVGFWN VVGYVTQCGG NANGQTSTSS  
  
130                    140                    150                    160                    170                    180  
 TTIFNNEPGY RSTSITCSLN GYTPGYYGPM SIENFKKLNE AYQILQTALK QGLPALKENN  
  
190                    200                    210                    220                    230                    240  
 GTVSVTYTYT CSGKGNDCS PQTTGVNQON DGTKTETQTI DGKQVNTTIS SKVVDSKASG  
  
250                    260  
 NTSHVSYTEI TNKLEGVEHH HHHH

Primer name	Primer sequence	T <sub>m</sub> (°C)
Primers yielding constructs with C-terminal His-tags		
Forward universal	GACTAGCATGGTGAAGACGACGGCTTTTACACAAGC	67
Reverse 210	GGGGCGCTCGAGTCCGTCGTTTTGTTGGTTTACAC	68.2
Reverse 235	GGGGCGCTCGAGTTTGCTATCAACCACTTTTGAAGT	66.6
Reverse 255	CGGGCGCTCGAGCACTCCTTCTAATTTGTTAGTGATTCGG	67.2
Primers yielding constructs with N-terminal His-tags		
Forward universal	GACTAGCATATGGAAGACGACGGCTTTTACACAAGC	63.7
Reverse 210	GTGGTGCTCGAGCTATCCGTCGTTTTGTTGGTTTAC	65.3
Reverse 235	GTGGTGCTCGAGCTATTTGCTATCAACCACTTTTGAAGT	64.7
Reverse 255	CTGGTGCTCGAGCTACACTCCTTCTAATTTGTTAGTGATTCGG	65.2

Table 9.4: Primers used for the cloning of BabA extracellular fragments from *H. pylori* P12. All constructs, including fusion constructs, could be generated with these primers. Blue: Stop codon. Green: Restriction site. Red: required additional base pairs.

### 9.2.5 Sequences of BabA extracellular fragments from *H. pylori* G27

All sequences presented as a result from cloning into the pET28 system.

#### 9.2.5.1 BabA 210

```

10      20      30      40      50      60
MGEDDGFYTS VGYQIGEEAQ MVTNTKGIQE LSDRYESLNN LLTRYSTLNT LIKLSADPSA

70      80      90      100     110     120
VNAARENLGM SAKNLIGDKA NSPAYQAVLL AINAAVGFWN VVGYITQCGG NANGQKSTSS

130     140     150     160     170     180
TTIFNNEPGY RSTSITCSLN GYNPGYYGPM SIENFKKLNK AYQILQTALK QGLPALKENN

190     200     210
GTVSVTYTYT CSGNGNDNCS KEATGVDNQN SGEHHHHHH

```

**9.2.5.2 *BabA 235***

10                    20                    30                    40                    50                    60  
 MGEDDGFYTS VGYQIGEEAAQ MVTNTKGIQE LSDRYESLNN LLTRYSTLNT LIKLSADPSA  
  
70                    80                    90                    100                    110                    120  
 VNAARENLGM SAKNLIGDKA NSPAYQAVLL AINAAVGFWN VVGYITQCGG NANGQKSTSS  
  
130                    140                    150                    160                    170                    180  
 TTIFNNEPGY RSTSITCSLN GYNPGYYGPM SIENFKKLNK AYQILQTALK QGLPALKENN  
  
190                    200                    210                    220                    230                    240  
 GTVSVTYTYT CSGNGNDNCS KEATGVDNQN SGTKTETQII DGKSVTTTIS SKVVDSEHH

HHHH

**9.2.5.3 *BabA 255***

10                    20                    30                    40                    50                    60  
 MGEDDGFYTS VGYQIGEEAAQ MVTNTKGIQE LSDRYESLNN LLTRYSTLNT LIKLSADPSA  
  
70                    80                    90                    100                    110                    120  
 VNAARENLGM SAKNLIGDKA NSPAYQAVLL AINAAVGFWN VVGYITQCGG NANGQKSTSS  
  
130                    140                    150                    160                    170                    180  
 TTIFNNEPGY RSTSITCSLN GYNPGYYGPM SIENFKKLNK AYQILQTALK QGLPALKENN  
  
190                    200                    210                    220                    230                    240  
 GTVSVTYTYT CSGNGNDNCS KEATGVDNQN SGTKTETQII DGKSVTTTIS SKVVDSSASG  
  
250                    260  
 NTRGVSYTEI TNKLDGVEHH HHHH

Primer name	Primer sequence	T <sub>m</sub> (°C)
Primers yielding constructs with C-terminal His-tags		
Forward universal	GAGGTGCCATGGGGAAGACGACGGCTTTTAC	67.7
Reverse 210	CGGGCGCTCGAGTCCGCTGTTTTGATTATCTAC	65.6
Reverse 235	CGGGCGCTCGAGACTAGAATCAACGACTTTTGAAC	65.1
Reverse 255	CGGGCGCTCGAGCACACCATCTAATTTGTTGG	66.7
Primers yielding constructs with N-terminal His-tags		
Forward universal	GACTAGCATATGGAAGACGACGGCTTTTACACAAGC	63.7
Reverse 210	GTCCTGCTCGAGCTATCCGCTGTTTTGATTATCTAC	63.3
Reverse 235	GTCCTGCTCGAGCTAACTAGAATCAACGACTTTTGAAC	63.2
Reverse 255	GTCCTGCTCGAGCTACACACCATCTAATTTGTTGG	64.1

Table 9.5: Primers used for the cloning of BabA extracellular fragments from *H. pylori* G27. All constructs, including fusion constructs, could be generated with these primers. Blue: Stop codon. Green: Restriction site. Red: required additional base pairs.

### ***9.3 Appendix III: fluorescence spectroscopy data***

The following plots show the results of all fluorescence titrations for Epa1A and all variants. The results are either normally plotted, with concentration on the X-axis and quench on the Y-axis, or as Scatchard plots. On Scatchard plots, the X-axis shows the Quench, while the Y axis shows Quench divided by ligand concentration. A Scatchard plot should result in a straight line, which should be fitted by a single linear equation. If this is not the case, cooperativity, ligand or binding site heterogeneity could have been present during measurements. Thus, Scatchard plots are very useful in quality control of titration experiments.

As some of the Scatchard plots presented deviations from linearity, a binding site heterogeneity model with two different kinds of binding sites was tested (green) versus a more simple total binding model (red) (see section 5.3.5 for more information on the two models). The results on Figures 9.1 and 9.2 show that, even though the fits with the two binding sites are marginally better, one of the two resulting dissociation constants is always very error prone, while the other one is very similar to the one obtained with the total binding model. Therefore, it was decided that the total binding model represented, at least for comparison purposes, a better balance between fit quality and physical significance of the obtained values.



9.3.1.1 Lactose titrations

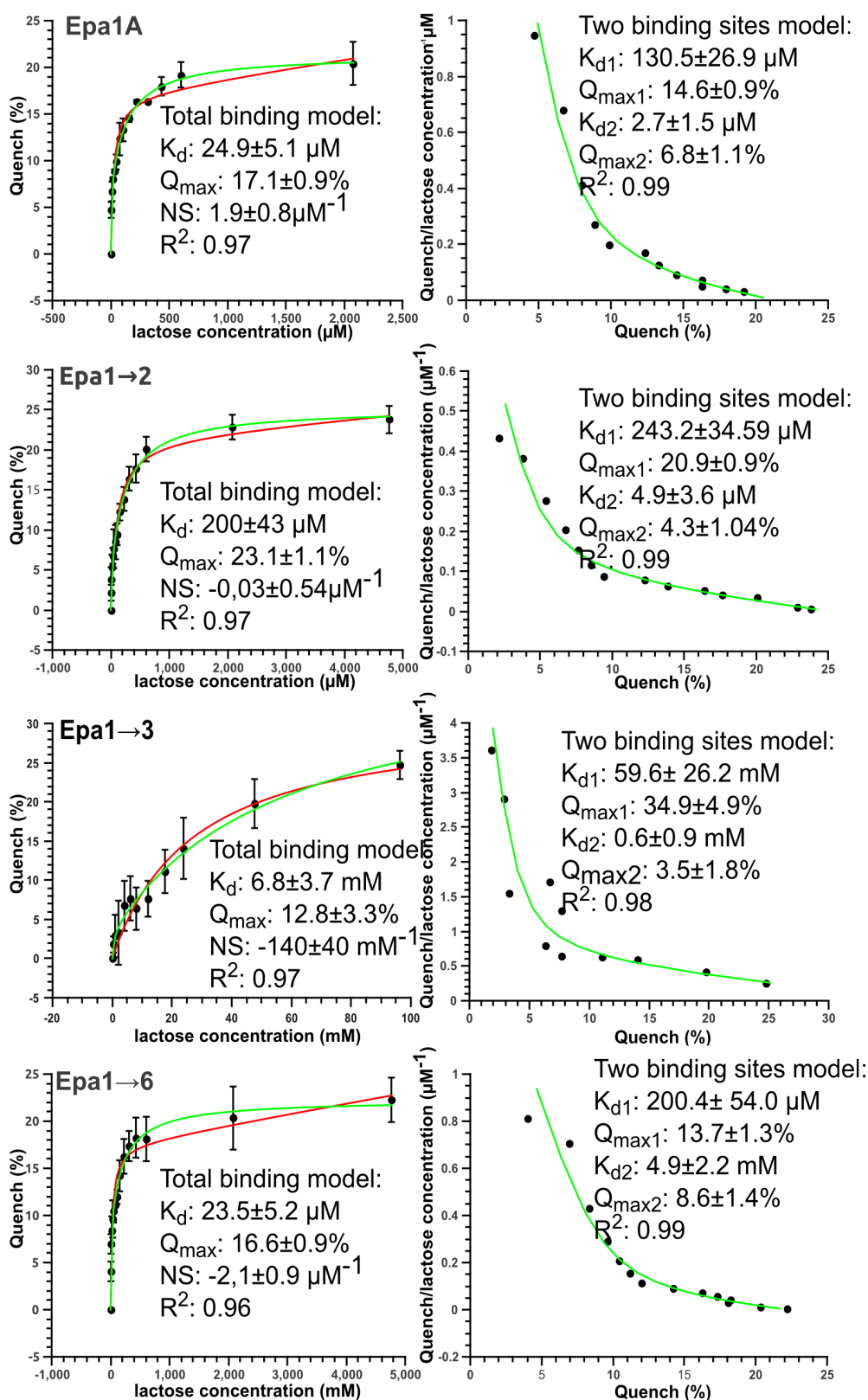


Figure 9.1: fluorescence titrations for Epa1A and variants against lactose.

9.3.1.2 T-antigen titrations

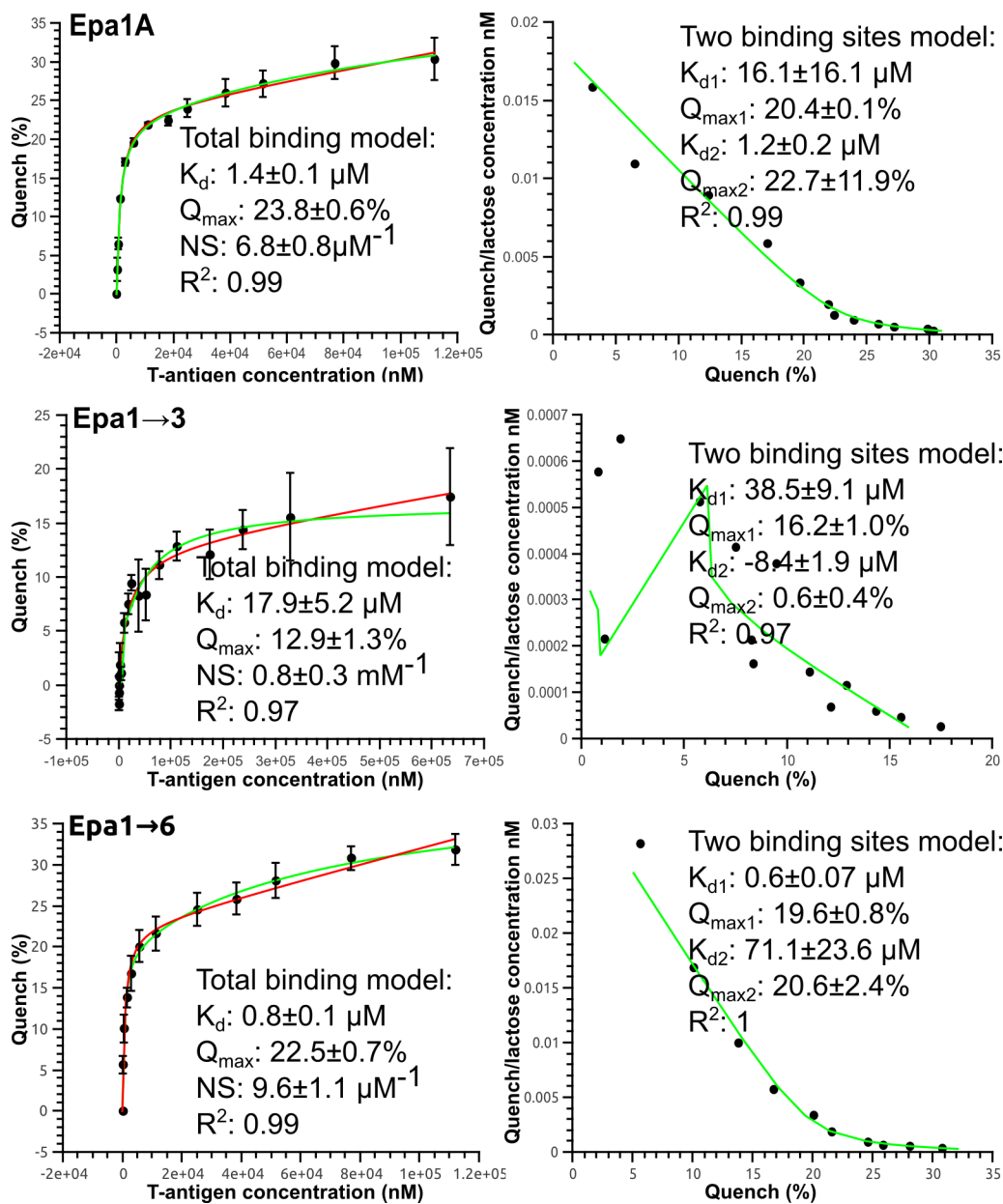


Figure 9.2: Fluorescence titrations for Epa1A and variants against the T-antigen