biochemical and structural studies of lectines from opportunistic filamentous fungi

Dörte Hündling

▶ **To cite this version:**

Dörte Hündling. biochemical and structural studies of lectines from opportunistic filamentous fungi. Biochemistry, Molecular Biology. Université Grenoble Alpes, 2015. English. NNT : 2015GREAV055 . tel-01684241

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 7 août 2006

Présentée par

Dörte HÜNDLING

Thèse dirigée par **Annabelle VARROT**
codirigée par **Anne IMBERTY**

préparée au sein du **Laboratoire CERMAV (Centre de Recherche sur les Macromolécules Végétales)**
dans **l'École Doctorale de Chimie et Sciences du Vivant**

Caractérisation biochimique et structurale de lectines d'*Aspergillus fumigatus*

Thèse soutenue publiquement le **15.06.2015**
devant le jury composé de :

M Markus KÜNZLER
Directeur de Recherche, ETH Zürich (Rapporteur)
Mme Anne BEAUVAIS
Directeur de Recherche, Institut Pasteur Paris (Rapporteur)
M Michel CHIGNARD
Directeur de Recherche, Hôpital St Antoine Paris (Membre)
Mme Murielle CORNET
Directeur de Recherche TIMC-IMAG (Présidente)
Mme Anne IMBERTY
Directeur de Recherche, CERMAV (Membre)
Mme Annabelle VARROT
Chargée de Recherche,CERMAV (Membre)

Acknowledgements

Table of contents

Abbreviations

ABPA:	Allergic Bronchopulmonary Aspergillosis
AFL:	*Aspergillus fumigatu*s Lectin
BLAST:	Basic Local Alignment Search Tool
CAPS:	*N*-cyclohexyl-3-aminopropanesulfonic acid
CBM:	Carbohydrate Binding Module
CF:	Cystic Fibrosis
CFG:	Consortium for Functional Glycomics
CFTR:	Cystic Fibrosis Transmembrane conductance Regulator
CV-N:	Cyanovirin-N
CVNH:	Cyanovirin Homolog
DHN:	dihydroxynaphtalene
DLS:	Dynamic Light Scattering
DMF:	Dimethylaformamide
EDC:	1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide
ESRF	European Synchrotron Radiation Facility
Fuc:	Fucose
Gal:	Galactose
GalNAc:	N- acetylgalactosamine
Glc:	Glucose
GlcNAc:	N-acteylglucosamine
HEPES:	*(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)*
HU:	Hemagglutination Unit
IA:	Invasive Aspergillosis
IMAG:	Immobilized Metal Affinity Chromatography
IPTG:	Isopropyl β-D-1-thiogalactopyranoside
ITC:	Isothermal Titration calorimetry
kDa;	kilo Dalton
LacNAc:	Lactosamine
MAD	multiple- wavelength Anomalous Diffraction

Man:	Mannose
MeOH:	Methanol
MES:	2-(*N*-morpholino)ethanesulfonic acid
NCBI:	National Center for Biotechnology Information
NeuAc:	N-acetylneuraminic acid
NHS:	*N*-hydroxysuccinimide
NMR:	Nuclear Magnetic resonance
ORF:	Open Reading Frame
PAGE:	Polyacrylamide gel electrophoresis
PAMP:	Pathogen-Associated Molecular Patterns
PBS:	Phosphate Buffered Saline
PCR:	Polymerase Chain Reaction
PDB:	Protein Data Bank
PEG:	Polyethylene glycol
PRR:	Pattern Recognition Receptors
PVDF:	Polyvinylidene fluorid
RBC:	Red Blood Cells
SAD:	Single- wavelength Anomalous Diffraction
SDS:	Sodium Dodecyl Sulfate
SPR:	Surface Plasmon Resonance
TBS:	TRIS Buffered Saline
TBST:	TRIS Buffered Saline Tween
TCA:	Trichloracetic acid
TCEP:	*tris*(2-carboxyethyl)phosphine
TEV:	Tobacco Etch Virus
TLR:	Toll-Like Receptor
TRIS:	tris(hydroxymethyl)aminomethane
TSA:	Thermal Shift Assay

List of Tables

List of Figures

The french cystic fibrosis association " Vaincre la Mucoviscidose" was founded in 1965 and has since then greatly contributed to the care of cystic fibrosis patients and their families in France. While cystic fibrosis (CF) was long considered a children`s disease, continued research and advancements in medical care has led to a significant increase in life expectancy and quality of life.

The greatest challenge however, remains that of pulmonary infection, which is still the leading cause of morbidity and mortality in CF patients.

While a variety of antibiotics are available to keep bacterial infections under control, fungal infections have become more prominent in recent years. Infections with filamentous fungi are often fatal and have generally a poor outcome due to difficulty in diagnosis and lack of efficient treatment strategies. *Aspergillus fumigatus* is the leading pathogenic filamentous fungi that infects not only cystic fibrosis patients but presents a general nosocomial risk factor and is subject of this thesis.

We focused our research on carbohydrate recognizing proteins (lectins) of *A. fumigatus*, which are believed to present an interface between host- tissue and pathogen. Investigating lectins of pathogenic agents may therefore contribute to the general understanding of infection strategies and provide novel treatment approaches.

In this thesis I will give an overview of *A. fumigatus* as an opportunistic pathogen and its role in the cystic fibrosis context together with an outline of protein glycosylation and their changes in the cystic fibrosis lung. Lectins are briefly introduced outlining major lectin families, their specification and biological function, emphasizing on structural features. An overview of fungal lectins is given, their structure, function and potential application are outlined.

This thesis addresses then the investigation of lectins from *A. fumigatus*, their identification and biochemical and structural characterization.

Being concerned with infectious agents that play a vital role in the cystic fibrosis context, our work was financially supported by "Vaincre la Mucoviscidose".

Summary

The aim of this thesis was to contribute to the understanding of infection strategies of the opportunistic pathogen *Aspergillus fumigatus*. This pathogenic mould is an emerging cause of morbidity and mortality in immuno-compromised patients and hospital environments. An infection with *Aspergillus* is generally referred to as Aspergillosis; it can develop in a variety of organs but the most common sites are the respiratory apparatus i.e. lungs and sinuses. Besides infections (invasive aspergillosis), colonization with the fungus can cause allergic reactions (allergic broncho pulmonary aspergillosis) and asthma.

The number of immuno-suppressed patients is steadily increasing due to advancement in the HIV, cancer and cystic fibrosis medical care, as well as an increasing number of organ transplantations. Needless to say that new antifungal drugs and preventive medication is desperately needed to support medical care for those patients. Even though several fungicides already exist on the market, invasive aspergillosis remains to be often fatal. On one hand, this is due to difficulties in diagnosis and on the other hand, resistances are emerging rapidly.

The motivation behind this thesis is to understand the underlying mechanisms that are involved in the first contact between conidial spores and host tissues. Initial adhesion steps often involve carbohydrate binding proteins, called lectins. They recognize glycoconjugates such as glycoproteins, glycolipids and glycosaminoglycans which cover the epithelial tissue and mucosal surface of the respiratory tract. Identification and characterization of the lectins from *A. fumigatus* will therefore contribute to the understanding of the glycostrategy of this opportunistic pathogen and of the mechanisms involved in adhesion and infection. Detailed structural analysis of the carbohydrate-protein interactions will allow ascertaining the lectins role in virulence and guide the design of glycomimetics, as adhesion inhibitors. With this novel approach of targeting the pathogen adhesion rather than its proliferation, resistances are believed to be less frequent due to the lack of evolutionary pressure. In this work, two different strategies were employed to obtain novel lectins. Firstly, lectins were purified from crude fungal extracts and secondly the *A. fumigatus* genome was screened for encoded proteins showing sequence similarity with known fungal lectins. While lectin purification from the crude extracts was inconclusive due to low lectin activity in the starting material, genome screening showed that several putative lectins were present. One of these lectins, named AFL6, belonged to the cyanovirin-N homolog (CVNH) family and it was

recombinantly expressed and purified. Glycan array and micro calorimetry techniques were carried out to investigate its carbohydrate binding specificity and the three dimensional structure was determined using X-ray crystallography.

The structure showed an overall similarity with other CVNHs with slight differences in the presumed carbohydrate binding sites. Unlike other family members, it shows a low affinity for mannosides and an apparent affinity for lactosamine containing glycan structures.

Résumé

L'objectif de cette thèse a été de contribuer à la compréhension des stratégies d'infection du pathogène opportuniste *Aspergillus fumigatus*. Ce pathogène est une cause émergente de morbidité et de mortalité chez les patients dont l'immunité est compromise et dans les milieux hospitaliers. Une infection avec *Aspergillus* est en général appelée une Aspergillose et elle peut se développer dans un certain nombre d'organes, le plus fréquemment dans l'appareil respiratoire (poumons et sinus). Outre les infections (Aspergillosis invasive), la colonisation par ce champignon peut causer des réactions allergiques (Aspergilloses broncho pulmonaire allergique) et de l'asthme.

Le nombre de patients immunodéprimés augmente régulièrement à cause des avancées des traitements du SIDA, du cancer, de la mucoviscidose ainsi que par le nombre grandissant de transplantation d'organes. De nouveaux médicaments antifongiques et des médicaments préventifs sont nécessaires pour venir en soutien aux soins médicaux des patients. Bien que plusieurs fongicides existent déjà sur le marché, l'Aspergillosis invasive reste souvent fatale. Ceci est lié d'une part à la difficulté d'établir le diagnostic, et d'autre part au fait que des résistances émergent rapidement.

La motivation de cette thèse est de comprendre les mécanismes impliqués dans le premier contact entre les conidies et le tissu de l'hôte. Ces mécanismes d'adhésion initiaux sont souvent réalisés par des liaisons entre les lectines et les carbohydrates. Le tissu épithélial et la surface muqueuse du système respiratoire sont couverts de structures possédants des carbohydratestels que les glycoprotéines, les glycolipides et les glycosaminoglycanes. L'identification des lectines d'*A. fumigatus* et leurs caractérisations devraient dorénavant contribuer à la compréhension de la glycostratégie de ce pathogène opportuniste ainsi que des mécanismes impliqués dans l'adhésion et l'infection. L'analyse détaillée de la structure des lectines permettra d'établir le rôle de ces protéines dans la virulence et de guider la conception de glycomimétiques, afin d'inhiber le phénomène d'adhésion. Cette nouvelle approche consistant à bloquer l'adhésion de l'agent pathogène plutôt que sa prolifération, vise à diminuer les résistances par une réduction de la pression évolutive. Deux stratégies différentes ont été utilisées pour identifier de nouvelles lectines. Tout d'abord une purification des lectines à partir d'extraits fongiques brut a été tentée et d'autre part un criblage pour

trouver des séquences similaires avec les protéines codées par A. *fumigatus* a été réalisé parmi une banque de lectines fongiques connues.

Alors que l'étude d'une lectine extraite d'extraits bruts n'a pas été concluante en raison de la faible activité de la lectine dans le matériau de départ, le criblage génomique identifié plusieurs lectines potentielles chez *Aspergillus fumigatus*. L'une de ces lectines, nommée AFL6 appartient à la famille des homologues à la cyanovirine-N (CVNH) et elle a été exprimée de façon recombinante et purifiée. Sa spécificité a été étudiée par puce à sucres et par microcalorimétrie de titration. Sa structure tridimensionnelle a été déterminée par cristallographie aux rayons X. La structure a montré un repliement proche d'autres CVNHs avec de légères différences dans les sites présumés de liaison des sucres. Contrairement à d'autres membres de la famille, AFL6 montre une faible affinité pour les mannosides et une affinité apparente pour des glycanes contenant le motif lactosamine.

Chapter I

General Introduction

1. Fungi: mushroom, yeast and mold

Figure 1: mushrooms, baker's yeast and mold (*Aspergillus fumigatus*)

The vast kingdom of fungi includes more than 70000 known species and can be subcategorized into macrofungi (mushrooms, rusts, truffles and puffballs), yeasts and molds. Fungi are eukaryotic heterotrophs and absorb nutrition from their environment. Mushrooms are macrofungi with distinctive fruiting bodies which are either epigeous or hypogeous. Their habitat is decaying organic matter and/or establishing symbiosis with plants such as trees. Many kinds of mushrooms are edible and represent an excellent source of proteins. Various mushrooms possess interesting biological activity and have been used as medicine, predominantly in Asian countries for thousands of years.

Yeasts are microfungi, usually in the form of single cellular organisms. They play an important role in fermentation and biotechnology. Some yeasts like *Candida* species are human pathogens that may infect the oral and vaginal mucosa, and can also cause systemic infections in severely immunocompromised patients.

More often systemic infections and allergic reactions are caused by filamentous fungi and the most predominant among them are species of the *Aspergillus* genus, especially *Aspergillus fumigatus*.

2. *Aspergillus fumigatus*: an emerging opportunistic fungal pathogen
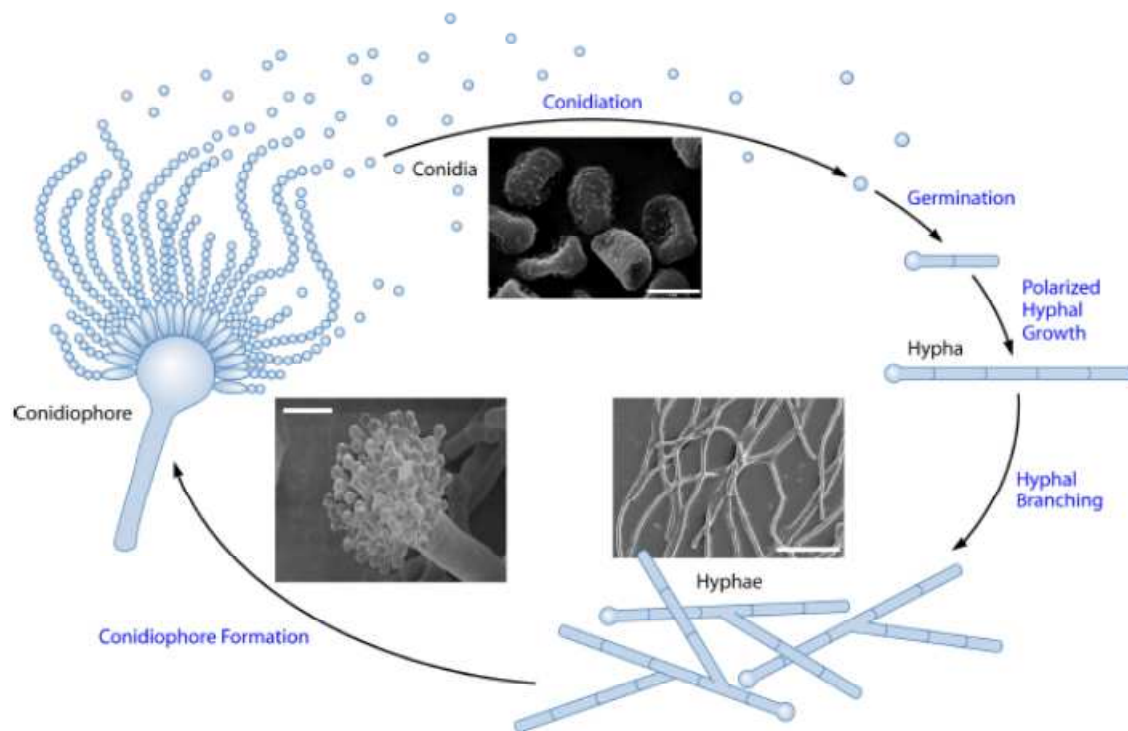
Figure 2: Life cycle and distinct morphological states of *A .fumigatus*. Conidiophore structures produce conidial spores which can germinate, and develop into hyphae. Hyphae cells continue to grow elongate and branch and eventually go on to form conidiophores. Picture reproduced from (Shapiro et al. 2011)

Aspergillus species are ubiquitous in nature and are found in air, food water and soil. They exist in distinct morphological stages and their asexual life cycle is depicted in figure 2. In decomposing vegetation, like other fungi, they play an important role in the recycling of nitrogen and carbon. Furthermore, some species are used in industry such as *A. oryzae, A. niger* and *A. terrus*. Few *Aspergillus* species can cause aspergillosis in humans and other animals and *A. fumigatus* is the most common opportunistic pathogen among them. It causes allergic bronchopulmonary aspergillosis (ABPA), aspergilloma and invasive aspergillosis (IA). ABPA is an allergic reaction due to the presence of conidia in the lung and it occurs most commonly in patients with extrinsic asthma or cystic fibrosis for whom it can lead to pulmonary fibrosis and eventually to respiratory failure (Knutsen & Slavin 2011). Aspergilloma is a spheroid mass of hyphae that grow in preexisting pulmonary cavities which may have been caused by tuberculosis or other cavitating lung diseases. It leads to expectoration of blood due to disruption of blood vessels which is often fatal (Latgé 1999).

Invasive aspergillosis involves mycelia growth into the host tissue. It occurs in severely immunocompromised patients, mainly in hematology patients and has a very high mortality rate (Lamoth et al. 2011). IA is difficult to diagnose and to differentiate from fungal colonization and histopathological evidence is often found only post-mortem (Lazarescu & Vinelli 2014). In cystic fibrosis patients, IA occurs predominantly after lung transplantation and is often fatal in these circumstances.

2.1 Infection with *Aspergillus fumigatus*

The first reported human case of aspergillosis dates back to 1845 and was detected by microscopic analysis of sputum samples (Bennett 1842). Recognition of aspergillosis as an opportunistic infection was mentioned 100 years later in a patient with systemic mycosis of *A. fumigatus* and *Candida* (Rankin 1953).While both *A. flavus* and *A. niger* are reported to cause aspergillosis in humans, *A. fumigatus* causes approximately 90% of this disease (Sales-Campos et al. 2013). Infection or colonization with *A. fumigatus* usually starts with inhalation of conidia, which are widespread in the environment. Due to their small size (3-5 µm), they can enter deep into the respiratory tract where they may find appropriate conditions to develop (figure 3). In the immune competent host, conidia are usually trapped in the mucus and actively transported by mechanical beating of cilia present on the respiratory epithelium. This however requires healthy mucus production, which is why cystic fibrosis patients are at a high risk for chronic pulmonary infections. Indeed, *A. fumigatus* is frequently found in expectorations from cystic fibrosis patients (Pihet et al. 2009).

2.2 Immune Response

The interaction between *A. fumigatus* and the innate immune system has been the subject of extensive research to understand the molecular mechanisms and identify targets for medical intervention. On a cellular level, conidia are taken up by phagocytosis of alveolar macrophages and polymorphonuclear neutrophils which are subsequently recruited by the production of pro- inflammatory mediators (Balloy & Chignard 2009). Swollen conidia are then eliminated within the phagosomal compartment by reactive oxidant species (ROS) (Philippe & Ibrahim-Granet 2003). The fungal defence mechanism is further supported by the production of anti-microbial peptides and proteins such as lactoferrin, which hinders fungal growth by trapping iron and proteases (Balloy & Chignard 2009).
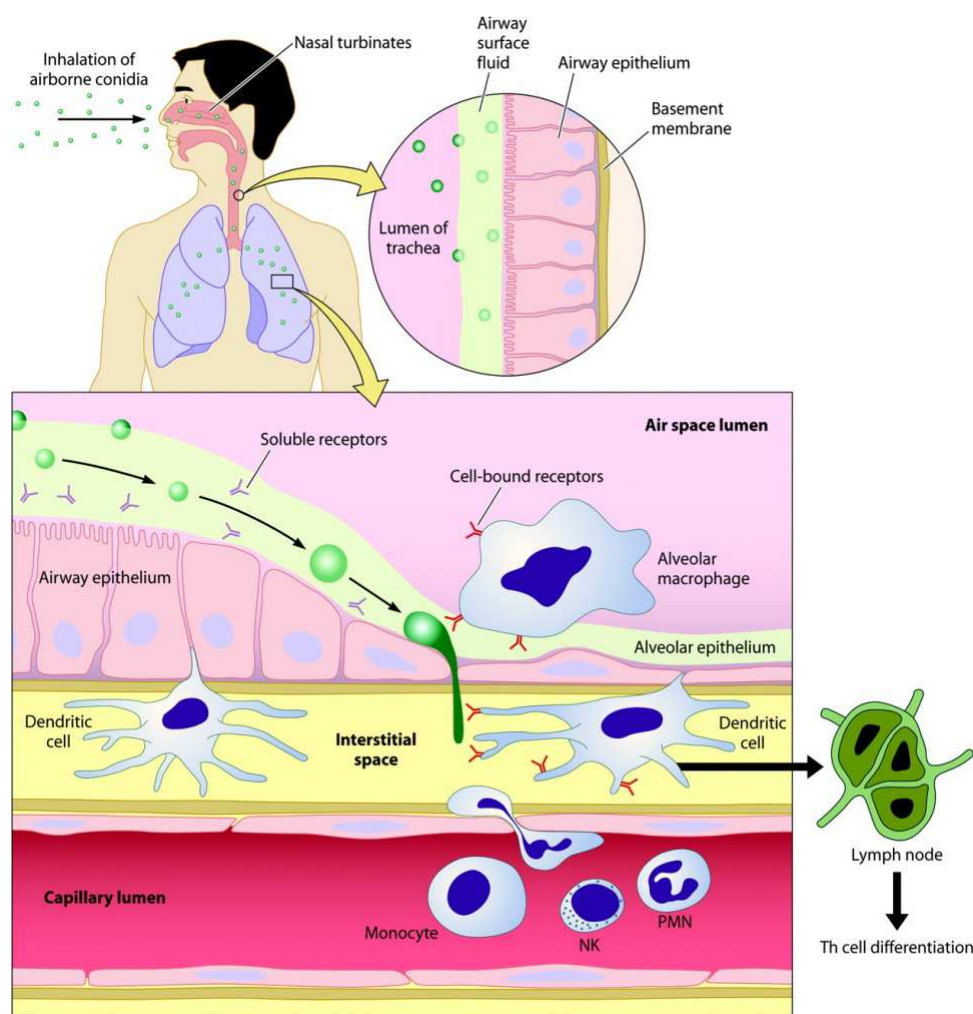
Figure 3: Schematic representation of components of the host immune response to inhaled *Aspergillus* conidia. Picture reproduced from (Park & Mehrad 2009)

At the molecular level the innate recognition of *A. fumigatus* is mediated by pattern recognition receptors (PRR), such as C-type lectins and toll-like receptors which recognize pathogen-associated molecular patterns (PAMP) like components of the fungal cell wall. C-type lectins include the secreted collectin-like surfactant proteins A and D (SP-A, SP-D) and mannose binding lectins that act as opsonins facilitating phagocytosis (Willment & Brown 2008). Furthermore, membrane bound receptors such as dectin-1 and 2 and DC-SIGN belong to the C-type lectin family and induce the production of chemokines (Lamoth et al. 2011). Dectin 1 is an innate immune receptor of mammalian cells and recognizes the fungal cell wall and initiates immune response. It is a C-type lectin that binds β-glucans that constitutes a major part of the fungal cell wall (Willment & Brown 2008). Toll like receptors (TLRs) were among the first described PRRs and several TLRs were identified to play an important role in the fungal defence mechanism. Activation of surface TLRs leads to homo- (TLR4) or hetero-dimerization (TLR2/TLR1 or TLR6) which leads to signalling cascades including MyD88, NFκB and MAP kinases and the subsequent expression of chemokines and cytokines (IL12, IL10, IL23, TGFβ, TNFα) (Bourgeois & Kuchler 2012). PAMPs that are recognized by the immune system consist mainly of carbohydrates, which are the major constituents of the fungal cell wall. These are especially α1,3 glucans, which were shown to activate T helper cells (Th1)-cells and to play an important role in activation of the host immune defences (Bozza et al. 2009).

2.3 Cell Wall and Virulence

The cell wall is essential to fungal survival, growth and defense against the host immune system. It is constantly adapted and changes in environment or external stress can lead to significant modifications in the composition and structural arrangement of the cell wall constituents. The cell wall is mainly composed of polysaccharides (figure 4).
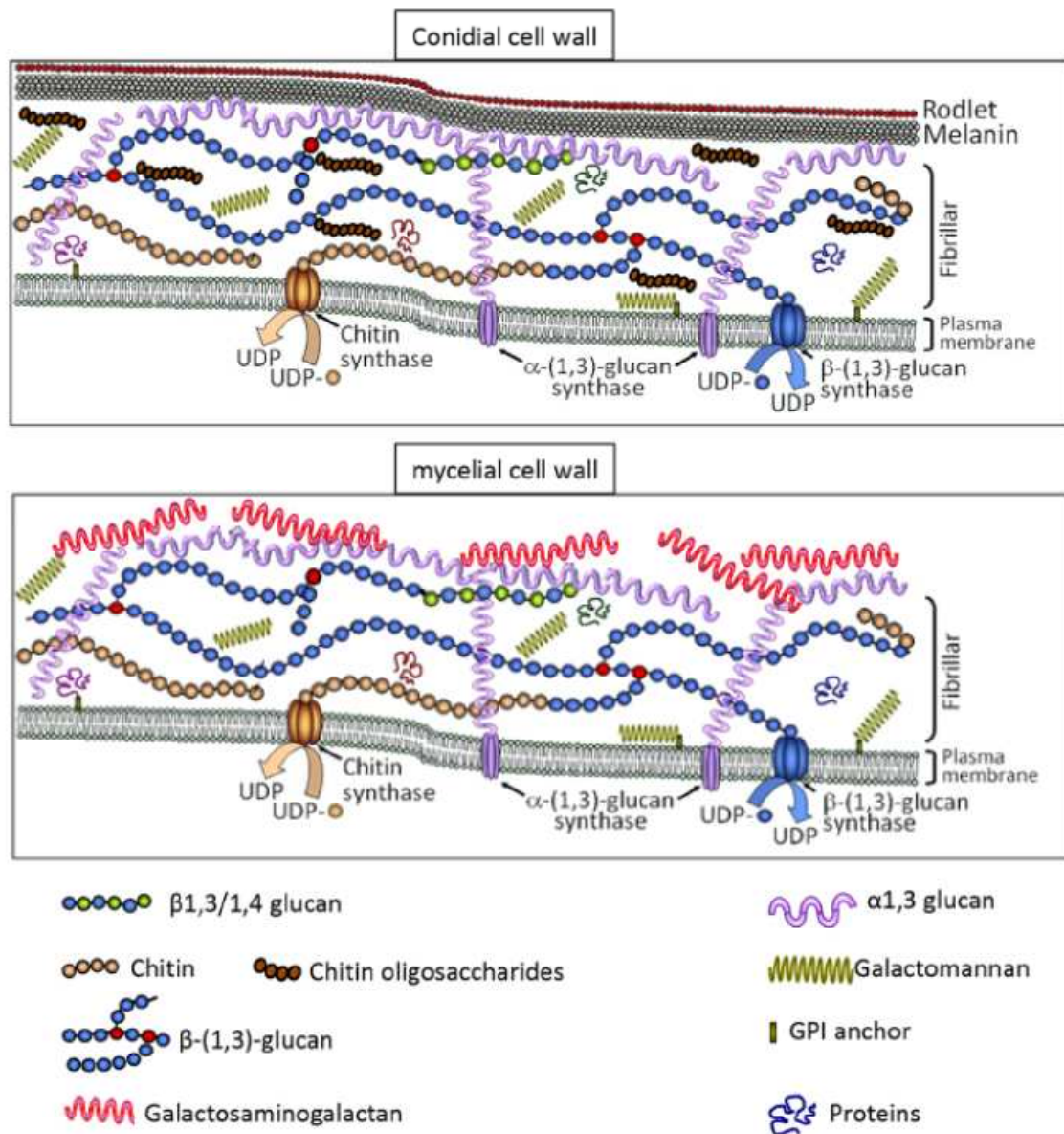
Figure 4: schematic representation of *Aspergillus fumigatus* conidial and mycelial cell wall. Picture reproduced from (Beauvais et al. 2014)

The core structure is made up of interlinked β1,3/β1,4 linked-glucan with branched chitin and chitinosan, as well as α1,2 galacto- (and α1,3mannan). An amorphous fraction is composed of linear chains of α1,3glucan with interlinked α1,3glucose and galactomannan (Free 2013) The conidial cell wall is additionally covered with dihydroxynaphtalene (DHN)-Melanin, which helps to maintain the structural integrity. Furthermore, it contains a hydrophobic rodletlayer which protects the fungal spore from immunological recognition by limiting the exposure of PAMPs (Aimanianda et al. 2009; Beauvais et al. 2014). It was shown that proteins play an important role in maintaining structural organization of the cell wall. For example, the deletion of the cell surface protein CspA, which is a GPI anchored cell wall protein, results in reduction of adhesion of conidia and increased disorganization of the cell wall, while its over-expression enhances resistance to degradation (Levdansky et al. 2010). The role of α1,3 linked glucan, the major polysaccharide component of the cell wall, was elucidated by the creation of a triple mutant strain which was devoid of α1,3glucan. It was shown that the absence of α1,3glucan does not affect growth or germination (Henry et al. 2012) but results in reduced pathogenicity, indicating the importance of α1,3glucan in virulence. Macrophage production was elevated in the triple mutant, due to increased exposure of PAMPs by cell wall reorganization (Beauvais et al. 2013).

2.4 Biofilm formation

In established colonization/ infection by *A. fumigatus,* exposure to the hostile environment of the host immune system is overcome by biofilm formation. The biofilm is a polymeric extracellular matrix (ECM) that covers the fungal colony and glues hyphae together, resulting in a reduced susceptibility to antifungal drugs. The existence of biofilm formation in *A. fumigatus* was recently described when mycelium was grown on agar based surfaces (Beauvais et al. 2007). Electron microscopy of aspergilloma and invasive aspergillosis in murine lungs showed that biofilm formation was present *in vivo* (Loussert et al. 2010). The biofilm composition was found to contain polysaccharides such as galactomannan, galactosaminogalactanandα1,3glucan, as well as pigments and proteins (Loussert et al. 2010; Beauvais et al. 2007). Furthermore, extensive research has been carried out using RNA sequencing, microarray and proteomics approaches to investigate the global understanding of *A. fumigatus* biofilm formation (Muszkieta et al. 2013).

2.5 Available Treatments and emerging resistance

Treatments that are currently available to fight against fungal infections can be categorized in three classes of chemicals: Polyenes, echinocandins and azoles. The most common polyene used is Amphotericin B (AmB). It binds to membrane sterols, creates transmembrane channels and leads to an increase permeability for cations. It inhibits also proton ATPase pumps depleting cellular energy reserves and increasing membrane fragility. AmB is highly effective *in vitro* but efficacy *in vivo* remains low and it shows toxicity towards the host. It binds to ergosterol in the fungal membrane, but also to cholesterol in the host (Latgé 1999). Furthermore some strains show intrinsic resistance to AmB (Shapiro et al. 2011). Echinocandins inhibit the synthesis of β1,3-glucan in a competitive manner and resistance is not yet emergered, possibly due to the relatively short duration of use (Shapiro et al. 2011).

Among the azoles intraconazole is the most commonly used antifungal against *A. fumigatus*. It binds to cytochrome P-450 and prevents synthesis of ergosterol. Resistance to azoles is an emerging problem that can be acquired during treatment or may be driven by agricultural use of azoles. The resistance appears by alteration of the drug target and upregulation of drug pumps. Furthermore, azole resistant strains have been isolated from patients independently of their azole treatment and from the environment across Europe (Howard & Arendrup 2011). Indeed, azole resistance is now considered a major public health concern as antifungal resistance is emerging in *Candida* and *Aspergillus* strains (Vermeulen et al. 2013; Alcazar-Fuoli & Mellado 2014).

3. Cystic fibrosis

Cystic fibrosis is an inherited disease that was considered a children's disease due to its early fatality. In the past decades however, medical care for those patients has drastically improved and with it life expectancy and quality. Medical research, especially research about the basis of infection, new treatment and diagnosistc tools, plays an important role in the advancements of medical care for cystic fibrosis patients.

3.1 Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)

Cystic fibrosis (CF) is a monogenic disorder caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. CFTR is a cAMP regulated chloride channel and a regulator of other ion channels expressed on most epithelial cells. The anion channel is needed on epithelial cells to control anion flow and maintain hydration of epithelial cells in the airways, the intestinal tract and pancreatic ducts as well as sweat glands and testes. The gene is located on chromosome seven and more than 1500 mutations are identified to date. However, the most common mutation results in the deletion of the phenylalanine residue at position 508 (*F508del*) of CFTR. The mutations are illustrated in figure 5 and can be categorised into 5 classes (Ferec & Cutting 2012). Class I mutations alter the stability of mRNA by introducing premature termination codons whereas class II mutations alter the stability of the protein. F508del is an example of such a mutation where protein maturation and trafficking to the Golgi apparatus is interrupted (Kopito 1999). In class III mutations CFTR is properly folded and located in the membrane but its activation is compromised. In the mutations of classes I-III protein activity is absent as opposed to classes IV and V where CFTR carries out residual function with altered conductivity (class IV) or reduced amounts of functional CFTR (class V). Diagnosis of cystic fibrosis is nowadays confirmed by genetic screening in newborns or pre-natal. However, in young children sweat tests are still used as reliable diagnostic tools. The sweat of cystic fibrosis patients contains a higher chloride concentration due to the malfunction of the chloride channel CFTR.
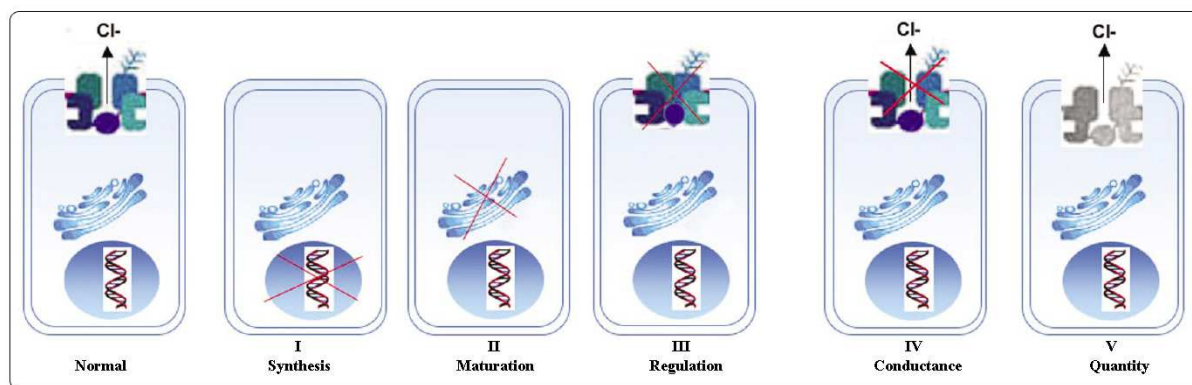
Figure 5: schematic overview of the five classes of CFTR mutations that lead to cystic fibrosis.
Picture reproduced from
The CFTR mutations database -the CFTR gene (http://www.umd.be/CFTR/W_CFTR/gene.html)

3.2 Clinical Symptoms of Cystic Fibrosis

The most significant symptoms of cystic fibrosis include the production of thick and sticky mucus in the intestines and airways which is why the term mucoviscidosis is also used for this disorder. Clinical characteristics are meconium ileus in newborns, liver fibrosis and male infertility due to congenital bilateral absence of the vas deferens. In the pancreas CFTR on the ductal epithelium of the exocrine system maintains an alkaline pH to provide optimal enzyme activity in the duodenum. In CF patients the idiopathic chronic pancreatitis leads to obstruction of the pancreatic duct due to lower pH in the secretions. Proteins clog and obstruct the pancreatic duct leading to fibrotic tissue in the pancreas and malnutrition due to the inability to absorb fats (Witt 2003). Patients are therefore often underweight and require nutritional supplements. In fact, the term cystic fibrosis was first used in 1938 in autopsy studies of mal-nourished children as "cystic fibrosis of the pancreas" (reviewed in (Davis 2006)). Since then life expectancy has greatly improved and CF is no longer only a children's disease. However, CF is still considered lethal and the main contributing factor to premature deaths is the chronic infection of the lungs and airways. Airway epithelium of the CF lung is covered by a thick layer of mucus that cannot be cleared. This leads to chronic obstruction andprovides a perfect environment for colonisation by inhaled bacteria and fungal spores (Delhaes et al. 2012). The link between the CFTR malfunction and the phenotype of thick and sticky mucus was reviewed recently. It was suggested that mucus hyper-production is not directly related to the malfunction of the chloride channel but rather by downstream

consequences that lead to secondary effects such as overproduction of mucins which may be the cause for the altered mucus properties (Kreda et al. 2012).

3.3 Infections in Cystic Fibrosis patients

Infection of the airways is the most important complication in CF patients. The infections are mediated by the inability to clear the thick and sticky mucus produced in the respiratory tract (figure 6). Furthermore, the cellular innate immune defences may be directly affected in CF. Once initiated, the inflammatory response appears to be dysregulated. Macrophages were shown to have defects in phagocytosis and showed an impaired killing mechanism by insufficient acidification of the lysosomes (Brennan 2008). It was recently shown that lung function in 3 month old infants, diagnosed with CF already showed abnormal lung function and hyperinflammation (Hoo et al. 2012). Further indications are given by the fact that low levels of CFTR in myeloid cells result in dysregulated inflammatory function (Bonfield et al. 2012)
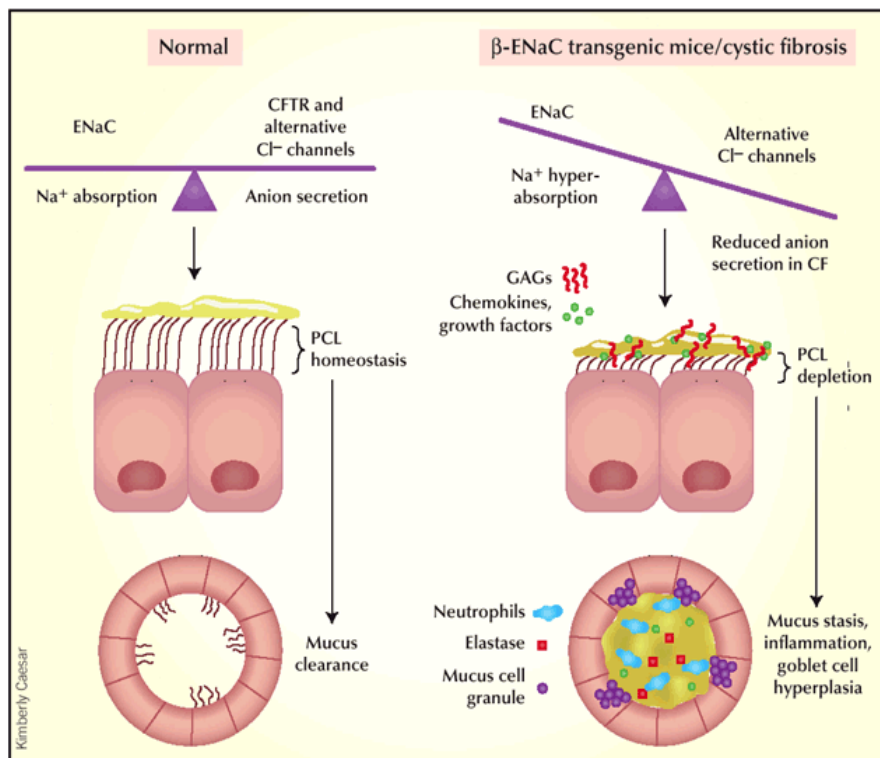
Figure 6: Schematic representation of the CF lung phenotype that leads to increased inflammation and infection. Picture reproduced from (Frizzell & Pilewski 2004)

3.3.1 Bacterial infections

The most prevalent bacterial infections in CF patients are mediated by the opportunistic pathogens *Pseudomonas aeruginosa*, *Burkholderia cepacia*, *Staphylococcus aureus* and *Stenotrophomonas maltophilia* (Coutinho et al. 2008) but recent new methods for detection and diagnosis suggested a much larger bacterial community in the cystic fibrosis lungs (Guss et al. 2011). Bacterial infection is the main cause of morbidity and mortality in patients with CF. In early life *Haemophilus influenza* and *Staphylococcus aureus* are the most prevalent bacterial pathogens whereas *P. aeruginosa* infection usually occurs in adolescence and adulthood. Acquired bacterial infections can become chronic and last for years and co-infections with several bacterial pathogens are often observed. On the one hand, *P. aeruginosa* was found to secrete factors that inhibit the growth of *S. aureus*, on the other hand *S. aureus* has also been observed to grow in proximity to *P. aeruginosa* resulting in antibiotic resistant *S. aureus* variants (Hoffman et al. 2006).

P. aeruginosa is the most common bacterial pathogen found in the context of CF and phenotypic conversion is observed during chronic colonization. It diversifies in motility, virulence, antibiotic susceptibility and the production of adhesins, polysaccharides and secondary metabolites in the CF lungs. Moreover, the formation of biofilms is often observed, making the bacterial infection difficult to eradicate (Tang et al. 2014).

3.3.2 Fungal Infections

Although CF is classically associated with bacterial infections, new detection methods of fungal pathogens suggest a much larger prevalence than expected (Delhaes et al. 2012).Pathogenic fungi can be divided into yeasts and moulds, also known as filamentous fungi. Among the yeasts, *Candida albicans* and other *Candida* species are most commonly infecting the oral and genital mucosa which results in localized infections. Although *Candida* are frequently isolated from sputum samples, their contribution to the reduction of pulmonary function remains unclear (Chotirmall et al. 2010). Besides the frequent presence of *Candida* spp, a case report of a CF child described pulmonary infection with the novel yeast *Blastobotrys rhaffinosifermentans*, with strong indication that the infection with this yeast highly contributed to the airway inflammation (Wong et al. 2014).

Among the filamentous fungi that colonize the CF airways, *A. fumigatus* is the most prevalent. It is frequently isolated from sputum samples from adolescent and adult CF patients. It causes predominantly ABPA and invasive forms occur occasionally. An extensive review by the Cystic Fibrosis Foundation has recently highlighted the difficulty in diagnosing

ABPA due to overlapping clinical, radiographic, microbiologic and immunologic features of ABPA and CF so that the clinical indication for treatment is often given when antibiotics fail (Stevens et al. 2003).The occurrence of ABPA has been increasing in the last years and a recent review of data from 30 different countries suggests that it is present in almost 50% of the adult CF population (Armstead et al. 2014). *A. fumigatus* is furthermore found to colonize the CF lungs without causing ABPA and its contribution to the decline in lung function is still debated in the scientific community. Some studies found no association between *A. fumigatus* colonization and a decline in respiratory function (de Vrankrijker et al. 2011),while another study found that it increases cytokine production fromTh2 cells and down regulation of the vitamin D receptor. Antifungal treatment resulted in significant clinical improvement (Coughlan et al. 2012). Conflicting results were also found with *A. fumigatus* colonization in context with antibiotic treatment and *Pseudomonas* infection (reviewed in (Liu et al. 2013). The heterogeneity of the study design and duration may account for these conflicting results and the impact of fungal colonization in the CF lung remains unclear (Chotirmall & McElvaney 2014).

Besides *A. fumigatus*, other *Aspergillus* species are isolated from patients such as *A. flavus*, *A. niger* , *A. nidulans* and *A. terreus* (Pihet et al. 2009). Some rare species like *A. lentulus* and *Neosartorya pseudofisheri* have been isolated from CF patients as well (Symoens et al. 2010). Furthermore *Penicillium* and *Scedosporium* species have been found to infect or colonize CF lungs. In a recent study investigating the seroprevalence of *Scedosporium,* it was found that 10% of sera were positive for *Scedosporium* with a significant correlation with *A. fumigatus* seropositivity. However, an association with decline in lung function could not be observed (Parize et al. 2014). Moreover, *Exophilia dermatitis* frequently colonizes the airway. In a study of 98 patients, *E. dermatitis* was recovered with the same frequency as *A. fumigatus* and triggered IgG production. Culture positive patients showed a decline in respiratory function and four patients were diagnosed with symptomatic infection and were treated with broad-spectrum azoles. Co-colonization occurred with non-tuberculosis *Mycobacteria* and *S. aureus* (Kondori et al. 2014).

It remains however, that *A. fumigatus* contributes significantly to the morbidity and mortality to cystic fibrosis patients. Novel treatment and diagnostic methods are therefore desperately needed. Investigating the initial steps in infection, the first contact between *A. fumigatus* with host constituents includes the understanding of the glycocalyx of the host tissue as well as adherence mechanisms of the pathogen through lectins. Protein glycosylation and the changes in the CF lung are described in the next section before lectins are introduced.

4. Glycosylation

4.1 Sugars: About the Complexity of Glycosciences

Besides the nucleic acids that build the DNA and amino acids that form proteins, carbohydrates are the third major building block of life. Carbohydrates or saccharides (from Latin *saccharum*: sugar) may be categorized in monosaccharides (single sugars), oligosaccharides in which two to ten monosaccharides are covalently linked and polysaccharides. Oligo and polysaccharides may be linked in linear or branched manner. Monosaccharides can exist in linear form as aldoses, containing an aldehyde function or as ketoses with a keto-group. Ring formation results in the formation of an anomeric center which leads to the α or β anomers as shown in figure 7.

Figure 7: cyclization of D-glucose in which the carbonyl group forms a covalent bond with the oxygen of a hydroxyl group. Cyclic presentations are shown in Haworth projection. The linear from is shown in Fischer projection.

Linkage of simple sugars to form oligo- and polysaccharides (glycans) can occur not only at the anomeric center, resulting in α and β linkages but on any hydroxyl group, resulting in a variety of possibilities as opposed to the linear formation of proteins and DNA, where coded information is restricted to their sequence (figure 8).
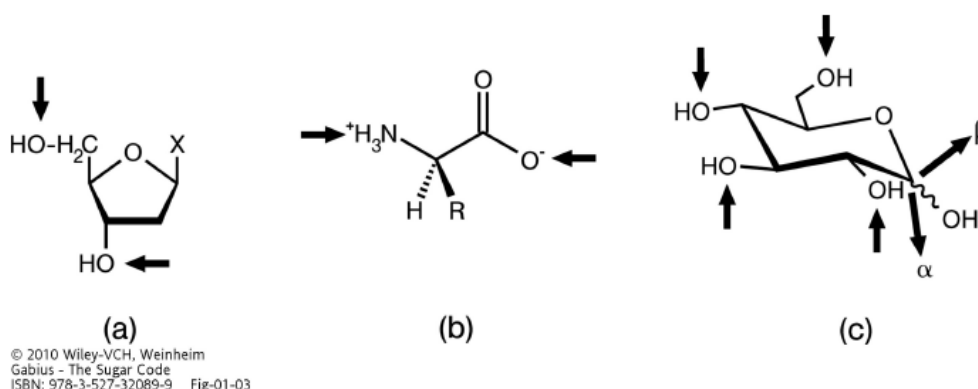
Figure 8: Illustration of linkages to form linear oligomers of (a) nucleic acids, (b) amino acids. The oligomerization of (c) carbohydrates is much more complex and can yield a variety of branched and linear oligo-and polysaccharides.Picture reproduced from (Rüdiger & Gabius 2011)

Glycans that are linked to lipids or proteins are referred to as glycoconjugates and are major components of membranes and cell walls in all living kingdoms. They serve not only as structural components but are also involved in a variety of biological processes, such as cell-cell communication, maturation and differentiation as well as host-pathogen recognition (Varki et al. 2009).

4.2 Glycosylation of Proteins

Other than proteins and nucleic acids, the glycan biosynthesis is not driven by a template. Its composition depends upon an orchestration of different glycosyltransferases and glycosidases. In eukaryotes several core glycan structures are assembled in the endoplasmatic reticulum (ER) and mature in the Golgi network and the different compartments of the secretory pathway (Van Breedam et al. 2014).Due to this complex biosynthesis and the lack of proof-reading machinery, glycan structures present large diversity and strong heterogeneity (Raman et al. 2005). Glycosylation is a common post translational modification of secreted protein and every cell is covered by a complex layer of glycans (Varki et al. 2009). Protein glycosylation can be divided into two types depending on the glycosidic bond type: N-glycans and O-glycans. Changes in protein glycosylation can have drastic effects of protein folding, stability and trafficking. An excellent example for the complex interconnections between glycosylation, protein function and phenotype is given by the

altered glycosylation of CFTR in cystic fibrosis. In some mutations, the glycosylation sites are altered leading to mis-glycosylation, protein instability and this is even thought to affect the overall glycosylation pattern of the cell (Rhim et al. 2001)
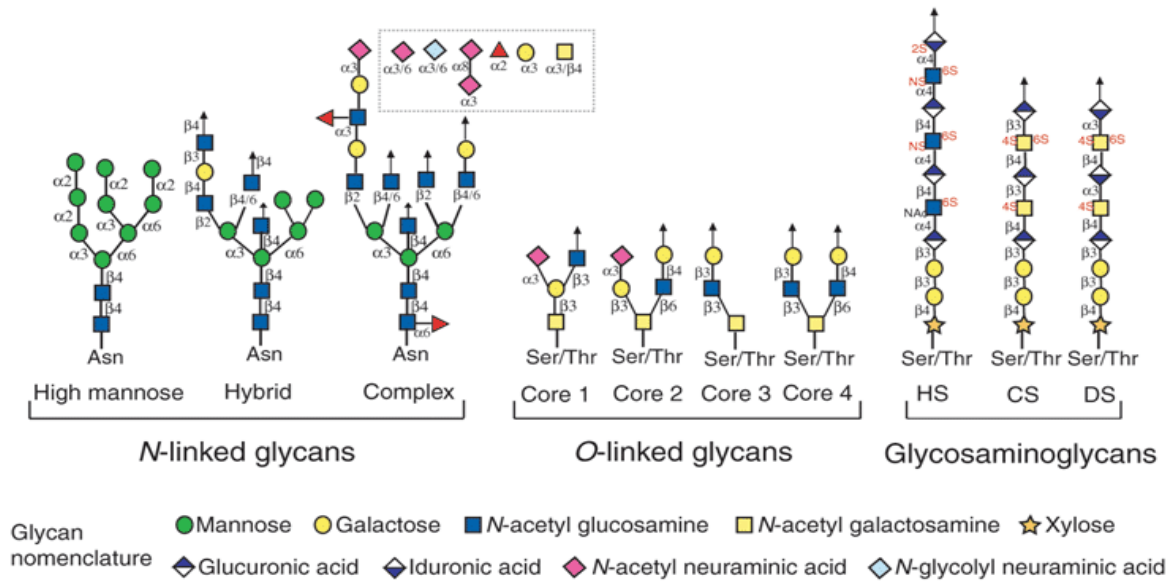
Figure 9: schematic representation of different types of protein glycosylation
Picture reproduced from (Raman et al. 2005).

N-linked glycans are covalently linked through a GlcNAc residue to the nitrogen atom of the side chain of asparagines in the Asn-X-Ser or Asn-X-Thr consensus sequence, where X can be any amino acid except for proline (GlcNAcβ-Asn) (Van Breedam et al. 2014; Varki et al. 2009). *N*-glycosylation is initiated in the endoplasmic reticulum during translation, and can be divided into three major steps. The synthesis of the common core glycan structure starts with the formation of dolichol-linked GlcNAc and subsequent addition of GlcNAc and three mannose residues. The core glycan structure is then extended to create the precursor glycan. The precursor is then transferred to the asparagines side chain of the protein and further processing modification by adding and trimming takes place in the Golgi apparatus. *O*-glycans are covalently attached to the oxygen atom of the side chain of Serine and Threonine residues and their biosynthesis takes place in the Golgi apparatus.

4.3 Glycosylation changes in the Cystic fibrosis lungs

4.3.1 Airway epithelium

To target the initial steps of infection, it is important to understand changes in the environment in predisposed patients. The glycosylation pattern in the CF lung was show to differ from healthy individuals. The airway epithelium of the CF lung contains *N*-glycoproteins with a decreased amount of terminal sialic acids and increased fucosylation (Scanlin & Glick 1999). These changes in protein glycosylation appeared to be directly linked to the CFTR mutation and normal glycosylation patterns could be restored by expression of wild- type CFTR *in vitro* (Scanlin & Glick 2001). Furthermore the enzyme activity and mRNA level of three fucosyltransferases that are responsible for fucosylation were not changed in the CF cells (Glick et al. 2001). It was demonstrated that the pH in the Golgi of CF cells is higher than in normal cells and it has been suggested that abnormalities in sialylation is directly linked to the malfunction of CFTR. This chloride channel may be responsible for acidification of the Golgi to provide optimal conditions for a pH sensitive sialyltransferase in normal cells. Defective acidification in CF may explain the decrease in sialic acids found in the CF phenotype (Barasch & Al-awqati 1993). However, direct measurements of the Golgi pH by fluorescent pH indicators could not confirm the change in Golgi acidification in CF patients (Seksek et al. 1996). It is now believed that CFTR plays a role in the disruption of compartmentalization of the trans Golgi network (Rhim et al. 2001; Rhim et al. 2004)

3.4.2 Mucins:

Mucins are high molecular weight and heavily glycosylated glycoproteins that are produced by goblet cells and bronchial glands in the human airway mucosa. O-linked glycosylation accounts for the major part of oligosaccharides in mucins. In contrast to the normal airway epithelium, mucins (MUC5AC and MUC5B) of the CF mucosa appeared to contain a higher amount of sialic acids and this is dependent of the inflammation state of the lung. Sputum samples from CF patients that were infected with *Pseudomonas* contained a higher amount of sialic acids than uninfected CF patients (Davril et al. 1999). Furthermore, mucins appeared to show higher sulfation, which seemed to also increase with the severity of the infection (Roussel et al. 1975; Davril et al. 1999; Xia et al. 2005). Since the sulfate content of mucins secreted by the submucosal gland was not different in CF and non-CF patients (Schulz et al. 2005), it is believed that the increase in sialic acids and sulfates is a secondary effect of

inflammation rather than a direct consequence of the absence of functioning CFTR in the Golgi (Kreda et al. 2012).

Furthermore, the mucus of CF patients is highly adhesive, which may be due to the secretion of proteases by pathogens in infections (Rubin 2007). In the intestines, mucins were found to be present in higher density in mice lacking functional CFTR (CftrΔ508), which could be normalized by the addition of bicarbonate (Gustafsson et al. 2012).

5. Lectins

This section comprises a general introduction to lectins. An overview of different lectin families, their structural features and biological function are summarized in table xx. Focussing on structural characteristics, some important examples of the families are mentioned emphasising on fungal lectins. Further information about the known lectins in A. fumigatus can be found in Chapter III as an introduction to identifying and characterizing novel lectins in this opportunistic pathogen.

5.1 Overview

Lectins are carbohydrate binding proteins of non immune origin, without enzymatic activity. They are ubiquitous in nature from viruses and bacteria to higher organisms, like fungi, plants and animals. Lectins have been first identified by their ability to agglutinate erythrocytes and where originally called agglutinins. With the finding that some of these agglutinins had a different specificity towards the ABO blood types the term lectin was coined from the latin word "legere" meaning "to select". The biological functions of lectins are very diverse and depend greatly on the origin of the lectin but their common ground is the involvement in cell-cell communication through the glycan-structure on cell membranes (Varki et al. 2009). Lectins often possess low affinity towards monosaccharides and stronger interactions are usually the result of multivalent binding. Multivalency can be achieved by the presence of tandem repeats in the sequence leading to several binding sites or by oligomerization.
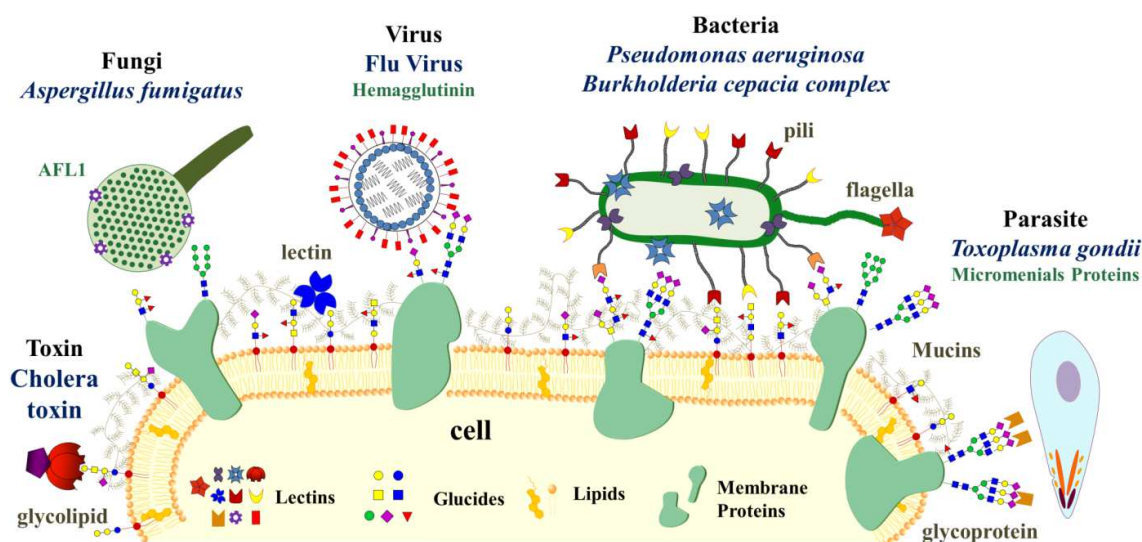
Figure 10: Schematic representation of involvement of lectins in pathogenesis. Adapted from (Imberty & Varrot 2008)

Figure 10 illustrates how lectins are used by many pathogens, be it viruses, bacteria, parasites or fungi to recognize their host target cell and establish colonization and/or infection. microbial lectins have therefore gained more interest in recent years as therapeutical targets for anti adhesive therapy. However lectin research began in plants and animals.

In plants, lectins are particularly abundant in the seeds of legumes and play an important role in the defense against pathogens (Lannoo & Van Damme 2014). Furthermore, these L-type lectins are thought to play an important role in the symbiosis between nitrogen-fixing bacteria and plant roots (Díaz et al. 1989; Sharon & Lis 2004). Although discovered in plants, L-type lectins are found in most eukaryotic organisms and are further discussed below. R-type lectins are abundant in plants as well and often contain toxins. In animals, lectins are known to be involved in proliferation, glycoprotein trafficking and clearance, immune defenses, fertility and many more. The most important lectin families, their specificity and biological functions are summarized below (table 1); among them many plant and animal lectins.

The role of lectins in the interaction between animal host and microbial pathogens is subject of intense research since lectins are often at the interface of host-pathogen recognition. They are involved in the activation of the innate and adapted immune system of the host but also in the adhesion and toxicity of the pathogen (Brummer & Stevens 2010; Sundstrom 2002; Imberty & Varrot 2008; Van Breedam et al. 2014). In viral infections, binding of the viral glycoconjugates to membrane associated host lectins can accumulate virions on the cell surface and facilitate infection. Furthermore, host glycans can be recognized by lectins on the viral envelope and initiate infection. Virus binding can also directly lead to internalization

into cell compartments where the next stage of infection can be initiated. Furthermore, the activation of signaling pathways through lectin binding may modulate both viral infection and the immune response (Van Breedam et al. 2014).

Bacterial lectins may be categorized into two major groups. Fimbrial and flagelin lectins are adhesins found on the surface, or surface organelles (fimbriae and pili) of many bacteria and are involved in attachment to the host surface, colonization and biofilm formation. the second group comprises soluble lectins and secreted toxins. Secreted toxins often have a lectin domain like the neurotoxin from *Clostridium botulinum* or the cholera toxin from *Vibrio cholerae.* Some cytoplasmic soluble proteins like LecA and LecB from *Pseudomonas aeroginosa* recognize galactose and fucose and are implicated in host recognition and biofilm formation (Imberty et al. 2005).

Given the diverse and ubiquitous nature of lectins, no unified classification is available yet. They are often classified by their origin, carbohydrate specificity or structural fold. Table 1 shows some important examples of lectin families, their origin and specificity, as well as biological functions and structural information.

Family	Origin	Structural features and Glycosylation	Ligands	Functions
Calnexin	vertebrates	N-terminal Ca-binding luminal domain, transmembrane helix	GlcNAc2Man 9Glc1 (N-glycoprotein)	protein folding assistance
C-type lectins	animal	Ca^{2+} dependent ligand binding, diverse structures	Divers	Adhesion, immune system, inflammation
P-type lectins	vertebrates	9 stranded flattened β-barrel	Man-6-Phosphate	Trafficking of hydrolases to the lysosome
R-type lectins	Invertebrates Plants	Structural similarity to Ricin B, β-trefoil domain, N-glycosylation of high mannose type	B-Gal, GalNAc, Man and others, depending on origin and subgroup	Adhesion and defense (toxicity)
L-type lectins	Discovered in plants, other eukaryotic organisms	Dome like structure comprised of antiparallel β-sheets, short loops and β-bends; jelly roll fold; most are N-glycosylated. Calcium depended binding in plants.	divers	Storage, symbiosis, defense, innate immunity
Galectins	All organisms	Antiparallel β-sheets arranged in a β-sandwich or jelly roll, no α- helices present	β- galacose-containing glycoconjugates	Cell-cell and cell-matrix interaction, signaling, immune response and inflammation, metastasis
I-type lectins (Siglecs)	vertebrates	Immunoglobin like fold, antiparallel β-sheets organized into β-sandwich	Sialic acid (for siglecs), divers	Cell-cell and cell-matrix interaction, immune system, neuronal development
H-type lectins	invertebrates	Six β-chains, β-sandwich fold	GalNAc	Self/ non-self recognition, cell adhesion, innate immunity
Ficolins	verterates	Contain subunits of fibrinogen like (globular) and collagen-like (stretched) domains	GlcNAc	Innate immunity

Table 1: Examples of lectin families

5.2 Structural diversity of lectins

As mentioned above, lectins may be categorized by their structural fold and table 1 already indicates the structural diversity found in lectins. Structural investigation of lectins is an excellent tool to elucidate their carbohydrate specificity and yield detailed knowledge about the protein architecture. Furthermore, it gives the opportunity to identify the most important protein-carbohydrate contacts. Nearly 250 different lectins have been structurally determined to date by X-ray crystallography or NMR resulting in more than 1200 three-dimensional structures. These are gathered in the comprehensive 3D-lectin database (http://glyco3d.cermav.cnrs.fr/). Figure 11 gives an overview of the structural diversity of lectins.Even though representatives are chosen from animal lectins, some of the presented lectin folds can be found in plant or microbial lectins.
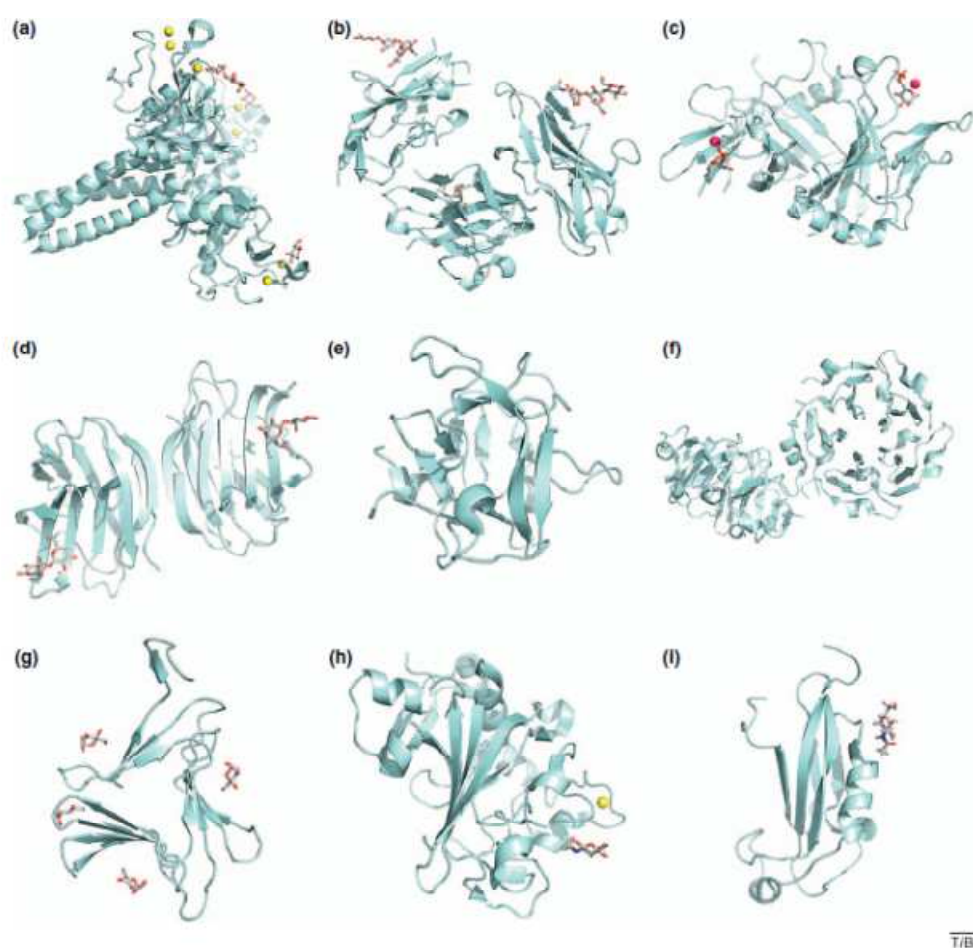
Figure 11: examples for the diversity of lectin folds. (a) C-type collectin (PDB: 1PWB); (b) immunoglobulin fold/ I-type lectin (PDB: 1QFO); (c) P-type lectin fold (PDB: 1M6P); (d) β-sandwich/jelly-roll fold of galectins (PDB: 1GZW); (e) β-trefoil fold (PDB: 1DQO); (f) β-propeller fold (PDB: 3KIH);(g) β-prism II fold (PDB: 1KJ1);(h) fibrinogen-like fold (PDB: 1JC9); (i) α/β-fold (actinoporin like) (PDB: 2JX9). picture reproduced from (Gabius et al. 2011)

C-type lectins are a major part of the innate immune system in animals and are characterized by a conserved binding site containing Ca^{2+}. Within the family of C-type lectins, several subgroups are known that differ in fold, specificity and function. Collagen-containing C-type lectins (collectins), like lung surfactant protein D (SP-D, Figure 11a) is an example of a well characterized C-type lectin. SP-D binds directly to carbohydrate residues on pathogens and allergens and mediates opsonisation. The trimeric lectin consists of an α- helical coiled coil region and the globular carbohydrate recognition domain, where Ca^{2+} is involved in carbohydrate binding with the maltose ligand (1PWD).

I-type lectins are named after their immunoglobulin like fold, belonging to the immunoglobulin superfamily but are neither antibodies nor T-cell receptors. Like c-type lectins they are exclusively found in animals. Sialic-acid binding I-type lectins (Siglecs) are a well characterized group of this family. Figure 11b shows the structure of the N-terminal carbohydrate binding domain of sialoadhesin in complex with 3` sialyllactose (1QFO). Siglecs are membrane proteins containing and N-terminal sialic acid binding region (V-set) at the extracellular lumen. Their cytoplasmic C-terminus contains a tyrosin signaling motive, down or upregulating signaling pathways in immune cells.

P-type lectins bind mannose- 6- phosphate (M6P) and are involved in the trafficking of hydrolases from the golgi apparatus to the lysosomal cell compartments. Figure 11c shows the endoplasmatic domain of the cation dependent M6P- receptor complexed with its ligand. P-type lectins are transmembrane proteins with a large N-terminal, glycoylated endoplasmatic domain that contains the carbohydrate binding site and a short C-terminal cytoplasmatc tail. They form homo-dimers with each monomer binding one M6P residue. The fold can be described as a nine stranded flattened β-barrel.

Galectins were first discovered in animals, but can be found in all organisms and they bind β-galactose containing glycoconjugates (table 1). Their common fold is composed of five and six standed antiparallel β-sheets that are arranged in a so called jelly-roll conformation (figure 11d, human galectin-1).This jelly-roll fold, or β-sandwich is also found in L-type lectins, which are present in plants, fungi and animals where it is sometimes called Concanavalin-A like β-sandwich, named after the first L-type lectin that was discovered from jack beans (Fujimoto et al. 2014).

The β-trefoil domain as shown in figure 11e is also known as Ricin-B like fold or R-type lectin. In plants R-type lectins often contain a toxin- subunit and play a role in defense. However, the β-trefoil domain is found in a variety of lectins like in fungi from *Clitocybe nebularis* (figure 12c) and the carbohydrate binding domain of the murine mannose receptor which belongs to the C-type lectins(figure 11e) and it structural features are described in more detail below. The structural diversity of the large family of C-type lectins is made clear in comparison to figure 12h which shows a C-type lectin from the Japanese horse shoe crab exhibiting a fibrinogen-like fold.

The β-propeller fold, as found in the *A. fumigatus* lectin AFL (figure 12f) is a well known member of fungal lectins and is therefore described below. However, it is also found in the Tachylectin from the Japanese horseshoe crab as shown in figure 11f where it is composed of a 5 bladed β-propeller whereas fungal lectins like AFL, AAL from the orange peel fungus *Aleuria aurantia* and PVL from P*sathyrella velutina* comprises 7 blades.

The β-prism II fold is common in mannose specific agglutinins from plants such as the lectin from garlic shown in figure 11g. However, they are also found in fishes, fungi and bacteria. It can be described as a β-barrel with a pseudo-3-fold symmetry which is made up of three four stranded β-sheets.

The rhamnose-binding like domain (RBL) from the murine lathophilin, a G protein coupled receptor in the brain of vertebrates (figure 11h) is a rather rare carbohydrate binding domain. It was first discovered in sea urchin eggs but has been discovered in other animals since. The α/β fold in figure 11h shows elongated loops that are involved in rhamnose binding (Vakonakis et al. 2008).

5.3 Fungal Lectins

Fungal lectins have been identified in mushroom fruiting bodies, in mycelium and conidia from filamentous fungi as well as in yeasts. Mushroom lectins gained growing interest due to their potential application in diagnostics and biotechnology as well as their antitumor, antiproliferative and immunomodulatory activities as summarized in table 2 (Varrot et al. 2013).

Origin	Biological function	Application in research and biotechnology

Mushroom
 	
Microfungi
 	
Yeast
 	
	• Storage
	• growth
	• morphogenesis
	• parasitism
	• infection
	• molecular recognition
	• defense
	• cell flocculation (yeast only)
		• purification of glycans and glycoproteins
		• Studies of glycomics
		• Biomarkers
		• Cancer research
		   ○ diagnostics
		   ○ immune stimulation
		   ○ antiproliferative
		• antiviral
		• Insecticide/ vermicide
		• Targeted drug delivery

Table 2: Origin, biological function and potential applications of fungal lectins

The study of lectins in pathogenic microfungi, such as yeasts and filamentous fungi is often motivated by their role in adhesion and virulence (Gallegos et al. 2014).

As described above, galectins are β-galactosides binding lectins that are well characterized in animals where they are involved in a variety of processes including differentiation, maturation, migration and apoptosis (Viguier et al. 2014). In fungi, lectins with a galectin-like fold have been characterized from *Agrocybe aegerita* (Jiang et al. 2012; Ban et al. 2005) and *Coprinopsis cinerea* (Wälti et al. 2008).They represent prototype galectins that are composed of a β-sandwich with two antiparallel β-sheets that contain each six β-stands.

Lectins with an actinoporin-like fold were characterized from the fruiting bodies of *Boletus edulis* (Bovi et al. 2011), *Sclerotium rolfsii* (Leonidas et al. 2007), *Xerocomus chrysenteron* (Birck et al. 2004) and *Agaricus bisporus* (Carrizo et al. 2005). They consist of a β-sandwich where the two β-sheets contain six and four β-strands respectively. The β-strands are connected by a helix-loop-helix motif which contains the carbohydrate binding site. Two

binding sites are present on each monomer with a primary site being specific for GalNAc and the secondary for GlcNAc. These lectins can form dimers or tetramers.
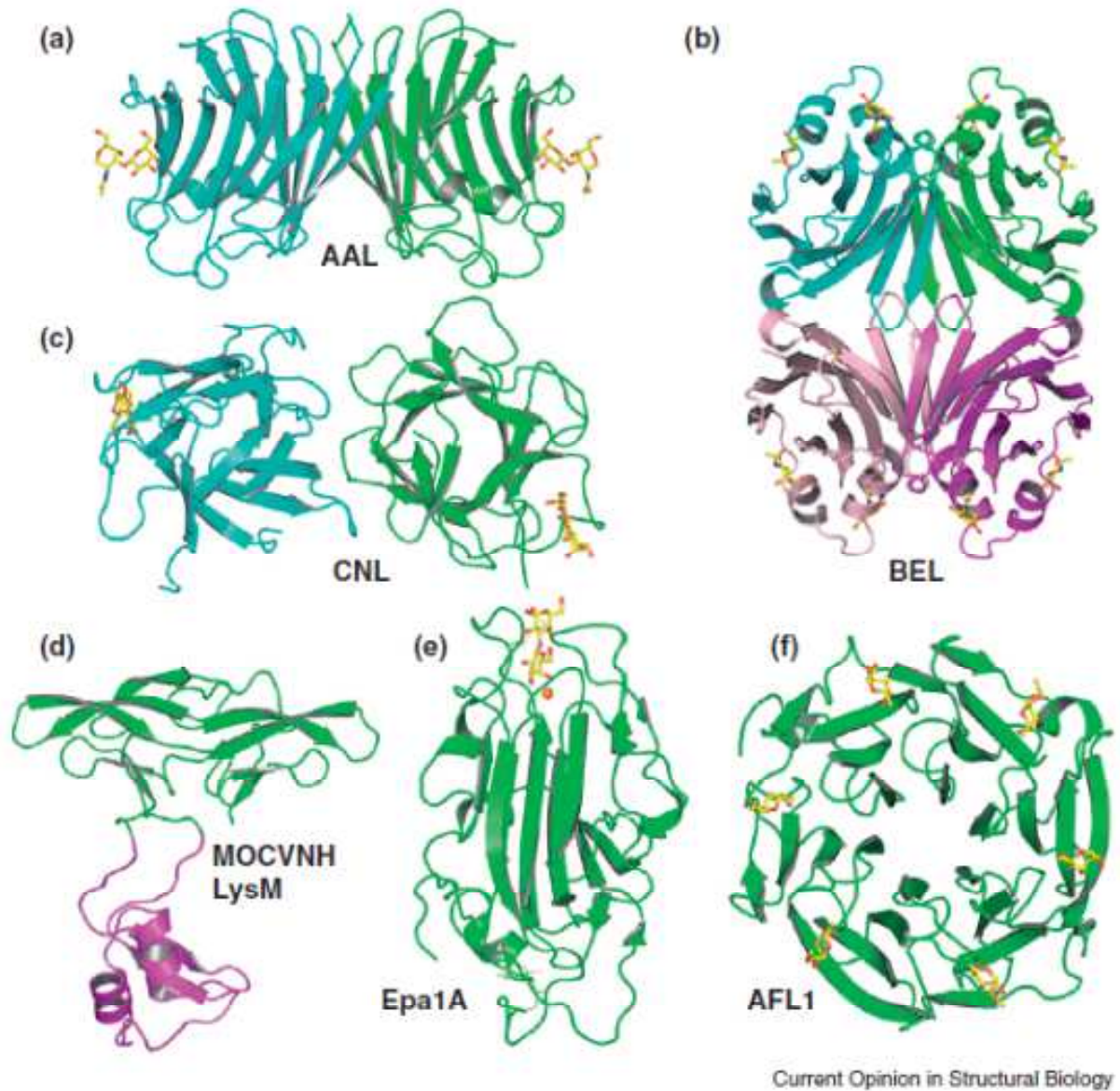
Figure 12: most common folds of fungal lectins. (a) Galectin fold of AAL *Agrocybe* aegerita lectin (PDB; 3AFK);(b) Actinoprotin-like fold of BEL *Boletus edulis* lectin (PDB: 3QDV); (c) β-trefoil fold of CNL *Clitocybe nebularis* ricin-B like lectin/ R-type (PDB: 3NBD); (d) Cyanorivin-N like fold with LysM domain of MOCVNH LysM *Magnaporthe oryzae* lectin (PDB: 2L9Y); (e) yeast adhesion fold of Candida galbrata epithelial adhesin 1A domain Epa1A (PDB: 4AF9); (f) β-propeller fold of *Aspergillus fumigatus* lectin AFL (PDB; 4AGI); carbohydrate ligands are shown as stick representation with yellow carbon atoms. reproduced from (Varrot et al. 2013)

Lectins with a β-trefoil fold have been identified from *Coprinopsis cinerea* (Schubert et al. 2012) and *Macrolepiota procera* (Žurga et al. 2014) where they were found to be involved in the fungal defense mechanisms. Furthermore, lectins with this fold were found in *Clitocybe nebularis* (Pohleven et al. 2012), *Laetiporus sulphureus* (Mancheño et al. 2005), *Marasmius oreades* (Grahn et al. 2007), *Polyporus squamosus* (Kadirvelraj et al. 2011) and *Sclerotinia sclerotiorum* (Sulzenbacher et al. 2010). The β-trefoil fold is characterized by three subdomains, α, β and γ each comprising a four stranded β-sheet and the domains are arranged around a pseudo-3-fold axis. The lectins vary greatly in their C-terminal domain architecture, oligomerization states and carbohydrate specificities (Varrot et al. 2013).

The cyanovirin-N homologs (CVNH) occurring in fungi are named according to their similarity with the virucidal lectin cyanovirin-N from the cyanobacterium *Nostoc ellipsosporum* and their structural fold, containing two domains A and B is further introduced in Chapter IV. In fungi, CVNH are predominantly found in filamentous ascomycetes and structural characterization is available for *Gibberella zeae* (Matei et al. 2011), *Neurospora crassa* and the truffle *Tuber borchii* (Koharudin et al. 2008). In addition, the CVNH from *Magnaporthe oryzae* contains a LysM domain inserted in between domains A and B (Koharudin et al. 2011).

In yeasts lectins are involved in adhesion processes. In *Saccharomyces cerevisiae,* the adhesin Flo5 mediates flocculation (Veelders et al. 2010) while in the pathogenic *Candida galbrata,* the adhesin Epa1A mediates adhesion to the host epithelium (Ielasi et al. 2012). Their lectin domain is formed by a β-sandwich with contains a calcium ion, that is involved in carbohydrate binding.

The β-propeller fold is present in the lectin from *Aleuria aurantia* (Wimmerova et al. 2003) and in AFL1 from *Aspergillus fumigatus* (Houser et al. 2013). Each of the six blades of the β-propeller is formed by a four stranded β-sheet. Both lectins are fucose specific and their fucose binding sites are located at the interface between two blades. The binding mode of AFL is further elaborated in Chapter III.

Table 3 summarizes selected fungal lectins and their carbohydrate specificity for a comprehensive overview.

Family	Name/ Origin	Specificity
Galectin	AAL/ *Agrocybe aegerita*	Galβ 1-4GlcNAc
	AGC/ *Agrocybe cylindracea*	Siaα2-3Galβ1-4GlcNAc
	CCG *Coprimus cinereus*	Galβ1-4GlcNAc
R- type	CNL/ *Clitocybe nebularis*	Galβ1-4GlcNAc
	PPL/ *Pleurocybella porrigens*	GalNAc
	PSL/ *Polyporus squamosus*	Siaα2-6Gal
	MOA/ *Marasmius oreades*	Galα1-3Gal
	MPL/ *Macrolepiota procera*	Galα1-3Galβ1-4GlcNAc
Fruiting body Actinoporin like	ABA/ *Agaricus bisporus*	Galβ1-3GalNAc, GlcNAc
	BEL/ *Boletus edulis*	Galβ1-3GalNAc
	SRL/ *Sclerotium rolfsii*	Galβ1-3GalNAc
	XCL/ *Xercomus chrysenteron*	Galβ1-3GalNAc
Fruiting body β-propeller	PVL/ *Psathyrella velutina*	Glcβ1-4GalNAc
	AAL/ *Aleuria aurantia*	Fucα1-6 GlcNAc
	AFL/ *Aspergillus fumigatus*	Fucα1-3(Galβ1-4)GlcNac
	AOL/ *Aspergillus oryzae*	Fucα6 GlcNAc

Table 3: Family, origin and carbohydrate specificity of selected fungal lectins.

AFL from *A. fumigatus* and Epa1A from *Candida glabrata* are the only lectins from fungal human pathogens that have been structurally characterized, underlining the need for more knowledge about the lectin- carbohydrate interactions in fungal infection. Indeed, lectins from filamentous fungi have been reported such as a sialic acid specific lectin has been purified from the phytopathogenic fungus *Macrophomina phaseolina*, but structural data is not yet available (Bhowal et al. 2005). Furthermore, a mannose specific lectin from *Penicillium chrysogenum* has been identified and its use as a pest control has been proposed (Francis et al. 2011).

The use of lectins as therapeutical targets, as well as their possible uses in research and biotechnology, as mentioned in table 2, underlines the importance of identification and characterization of novel lectins, especially from pathogenic microfungi like *A. fumigatus*.

6. Glycostrategy of *A. fumigatus* and aim of the thesis

A. fumigatus infection is an increasingly urgent issue to address in cystic fibrosis patients, hospital environments and immune suppressed patients. Diagnostic is lengthy and treatment often inefficient. It is therefore of utmost importance to find new approaches to address this problem. Our tactic involves the understanding of the so-called glycostrategy. It describes the role of glycan-lectin interactions between host and pathogen. These interactions are thought crucial to establish colonization and infection and may therefore present a putative target for antifungal therapy. Given the hypothesis that the adhesion of pathogens through lectins represents the initial and essential first step in colonization and infection, inhibition of those interactions may abolish colonization and infection. The inhibition may take place though adhesion competition using high affinity glycomimetics. These molecules have structures that resemble the carbohydrate ligand that bind to the pathogenic lectins with higher affinity than the natural ligand thus preventing adhesion, colonization and infection (Chabre et al. 2011; Imberty et al. 2008).

In addition, other protein-carbohydrate interactions occur in the process of infection, involving either fungal cell wall polysaccharides or human glycoconjugates. Their characterization is of great interest to shed light into their role in fungal infection and to pave the way towards novel approaches to antifungal treatment.

6.1 The fucose-strategy

Since fucosylation of glycoproteins in cystic fibrosis is increased, it is likely that opportunistic pathogens like *Aspergillus fumigatus* show increased binding to such fucosylated glycans. Indeed, a fucose specific lectin (AFL1) was identified and its expression in conidia was confirmed. Furthermore, it was shown that AFL1 simulated bronchial cells and led to the production of interleukin-8 indicating a role in virulence and in early infection stages (Houser et al. 2013).

6.2 The sialic acid-strategy

The conidial surface is covered by a layer of rodlets that are involved in hydrophobic binding to albumin and collagen, however a rodA mutant showed no difference in binding to the highly glycosylated fibrinogen and laminin, indicating that glycans may not be involved rodlet binding (Thau et al. 1994). A more important role in the conidial binding is that of sialic acid. Sialic acids are negatively charged nine carbon sugars that are often positioned as terminal monosaccharides on glycoproteins and play an important role in pathogenesis (Varki

& Gagneux 2012). At the surface of *A. fumigatus* conidia, sialic acids are bound to galactose and show predominantly a α2-6 linkage. Removal of sialic acids via neuraminidase treatment decreased their ability to bind fibronectin and caused the conidia to agglutinate. Furthermore, the removal of sialic acids reduced the phagocytosis of conidia by mouse macrophages *in vitro* (Warwas et al. 2007). These results indicate that sialic acids on the conidial surface are involved in adhesion and immune response. Furthermore, it was shown that conidia from non- pathogenic *Aspergillus* species contain a smaller amount of sialic acids and showed reduced binding to laminin (Wasylnka & Moore 2000; Wasylnka et al. 2001).

In addition to the role of sialic acids on the conidial surface, the matrix proteins of the human airway mucosa and the basal lamina are rich in sialic acids. Fibrinogen and fibronectin are major plasma glycoproteins that deposit at damaged epithelia and epithelial damage often predisposes fungal infection. Laminin is a major structural component of basement membranes and the role of these glycoproteins in adhesion has been investigated. Several studies were carried out using conidia binding and inhibition assays to immobilized glycoproteins (Peñalver et al. 1996; Coulot et al. 1994; Bouchara et al. 1997; Tiralongo et al. 2009; Tronchin et al. 1997; Gil et al. 1996). The binding was inhibited by soluble glycoproteins and sialic acid. However, it was debated whether the adherence of conidia was mediated by protein-protein interactions (Peñalver et al. 1996), ionic interactions (Wasylnka & Moore 2000), lectin carbohydrate binding (Bouchara et al. 1997) or could be due to a pH shift caused by the addition of sialic acid in *in vitro* assays (Tiralongo et al. 2009). With the identification of a sialic acid specific lectin in *A. fumigatus* conidia (Tronchin et al. 2002) the role of lectin–carbohydrate interaction in the adherence was confirmed. Besides this, other proteins have been identified to bind laminin independently of carbohydrate interaction contributing to the adherence of conidia (Upadhyay et al. 2009; Tronchin et al. 1997).

6.3 Aim of the thesis

The aim of the thesis is to contribute to the elucidation of the glycostrategy of *A. fumigatus* and its role in infection and virulence. Since the role of sialic acids in virulence and adhesion is debated, we attempted to reproduce the purification of the sialic acid-specific lectin from crude fungal extract and identify the underlying gene for recombinant expression. Structural and biochemical characterization of the lectin was aimed to identify the binding parameters. The importance of sialic acid binding- lectins in fungal pathogenesis has been demonstrated in the phytopathogenic fungus *Macophomina phaseolina* and the human pathogenic fungus *Penicillium marneffei* (Bhowal et al. 2005; Hamilton & Jeavons 1998).

Furthermore, it was the aim of this thesis to explore the existence of putative lectin- coding ORFs in the *Aspergillus fumigatus* genome. One of the identified lectins, a cyanovirin-N homolog, was recombinantly expressed, purified and crystallized. The carbohydrate interactions were assessed by microcalorimetry and glycan array.

With the difficulty in diagnosis, emerging of resistant strains and the assessment whether *A. fumigatus* contributes to the decline in respiratory function, glycomimetic therapy is aimed at circumventing this dilemma by providing a prophylactic treatment method, that aims to inhibit adhesion and colonization. Such a treatment strategy may possibly provide a standard protocol for medication of cystic fibrosis patients with isolated *Aspergillus* in sputum samples.

Chapter II
Methodology

1 Thermal Shift Assay (TSA)

TSA is a method to investigate the thermal stability of proteins by determination of their melting temperature Tm. The Tm of a protein describes the point of denaturation under the given conditions. During the experiment, the temperature is gradually increasing from 20-100 ℃ in steps of 1 ℃. The denaturation of the protein exposes hydrophobic patches which are recognized by an added fluorescent dye such as Sypro-Orange. The dye shows only weak fluorescencein a hydrophilic surrounding and its increase in fluorescence it directly related to the denaturation of the protein. With increasing temperature and denaturation of the protein, the monitored fluorescence will intensify. A peak is reached at the complete denaturation of the protein and further aggregation of denatured protein results in dissociation of the dye and decrease of the fluorescence. The melting temperature is defined by the slope of the resulting plot of fluorescence against temperature (figure 13)
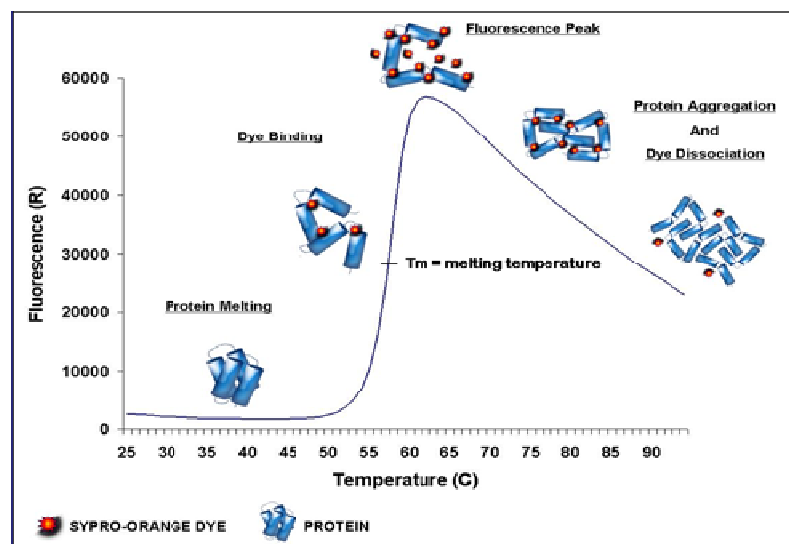
Figure 13: Schematic overview of the thermal shift assay. With increasing temperature, the protein denatures and the dye binds to exposed hydrophobic regions, resulting in fluorescence. A peak is reached when the protein is completely denatured and further aggregation results in dissociation of the dye and a decrease in fluorescence. The melting temperature is defined by the slope of the resulting curve.

The thermal shift assay can be carried out by using an RT-PCR machine and several conditions can be screened simultaneously. It requires low amount of native protein and the use of Sypro-Orange guarantees protein-independent detection, with wavelength for excitation and emission at 490 nm and 575 nm, respectively. The determination of thermal stability is not only an important characterization method for biologically and medically important proteins but can greatly improve crystallization success (Dupeux et al. 2011).

2 Dynamic Light Scattering (DLS)

Dynamic light scattering (DLS) can be used to determine the distribution profile and the size of macromolecules such as proteins or polymers. It is a non-invasive technique that measures the Brownian motion of particles in solution and calculates their size. The underlying mechanism is that small particles move faster in solution than larger particles. By shining laser light on the sample, the particles scatter the light in all directions and the intensity fluctuations of the scattered light beam are measured. From this, the diffusion coefficient is calculated and the size distribution profile holds information about the homogeneity of the sample. As such this technique is well suited to study the oligomerization or aggregation state of proteins.

3. Hemagglutination assay

Lectins were first identified and recognized for their ability to agglutinate red blood cells. They bind to the glycan structures present on the cell surface and thus form bridges between the red blood cells. The presence of a lectin can be visualized by incubation with a suspension of red blood cells in u- or v-shaped wells as shown in figure 14.
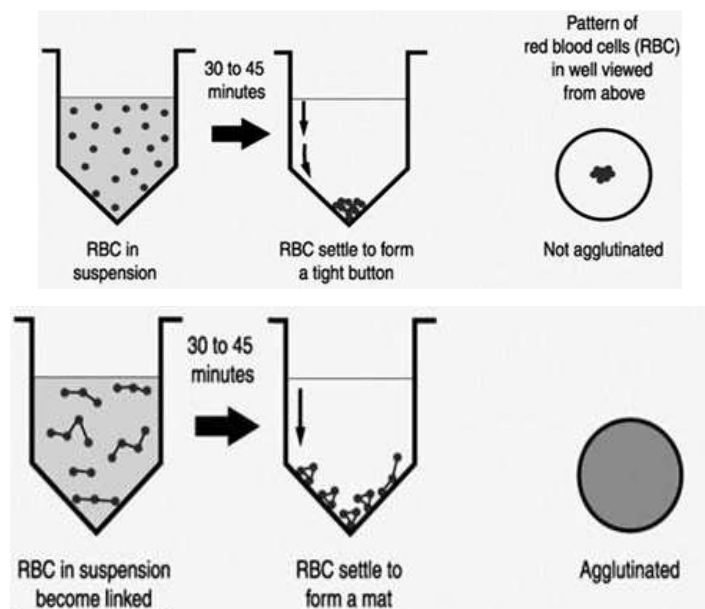
Figure 14: Overview of a classical hemagglutination assay. Suspension of RBC settles in form of a button in negative hemagglutination assay (top). In the presence of carbohydrate binding molecules RBCs become linked and form a mat. The hemagglutination assay is positive (bottom).

4 Isothermal Titration Calorimetry (ITC)

Calorimetry is one of the oldest techniques in experimental science, that measures the changes or transfer of heatand to date two major calorimetry methods are widely used. Differential Scanning Calorimetry (DSC) allows studying the thermodynamics of thermal denaturation, refolding and phase transitions of proteins, whereas Isothermal Titration Calorimetry (ITC) can be used to obtain thermodynamical parameters of protein-ligand, protein-protein interactions or enzymatic reactions. Indeed, ITC has been widely used to elucidate the thermodynamics of lectin- carbohydrate interactions (Dam & Brewer 2002). It is particularly well suited for such as it uses native protein without the need for labelling or immobilisation of the interaction partners.

The general principle of ITC is depicted in figure 15. The calorimeter (figure 15A) consists of two cells in an adiabatic environment. One of the cells is the reference cell, which contains only buffer, the other one is the sample cell, containing the protein solution. The syringe

contains the ligand, in this case the carbohydrate, diluted in the same buffer as the protein. When small amounts of ligand are injected into the protein solution, the chemical interaction between protein and ligand releases or absorbs small amounts of energy in the form of heat. This change in temperature is monitored by the machine (Microcal™ iTC200, Malvern Instruments), which measures the energy that is necessary to readjust the two cells to the same temperature. A series of regular injections allows plotting the electric energy over time (μcal sec^{-1}). After integration of the peaks, the binding isotherm is plotted as a function of the ligand/protein molar ratio (Figure 15 B).

In the first injections, the number of available binding sites exceeds the number of ligands, so that the heat release (or absorption) is maximal. During the following injections, the number of available binding site diminishes; hence the signal is getting smaller, resulting in a sigmoid curve after integration (Figure 15 C). The program (Origine®) calculates the theoretical curve of that superimposes with the experimental data to obtain the thermodynamic parameters of the interaction (Wiseman et al. 1989). The point of inflection corresponds to the stoichiometry; the amplitude equals the enthalpy of binding ΔH and the association constant K_a can be calculated from the slope. From these basic values the free Gibbs energy of binding ΔG and the entropy contribution ΔS can be calculated as described below.
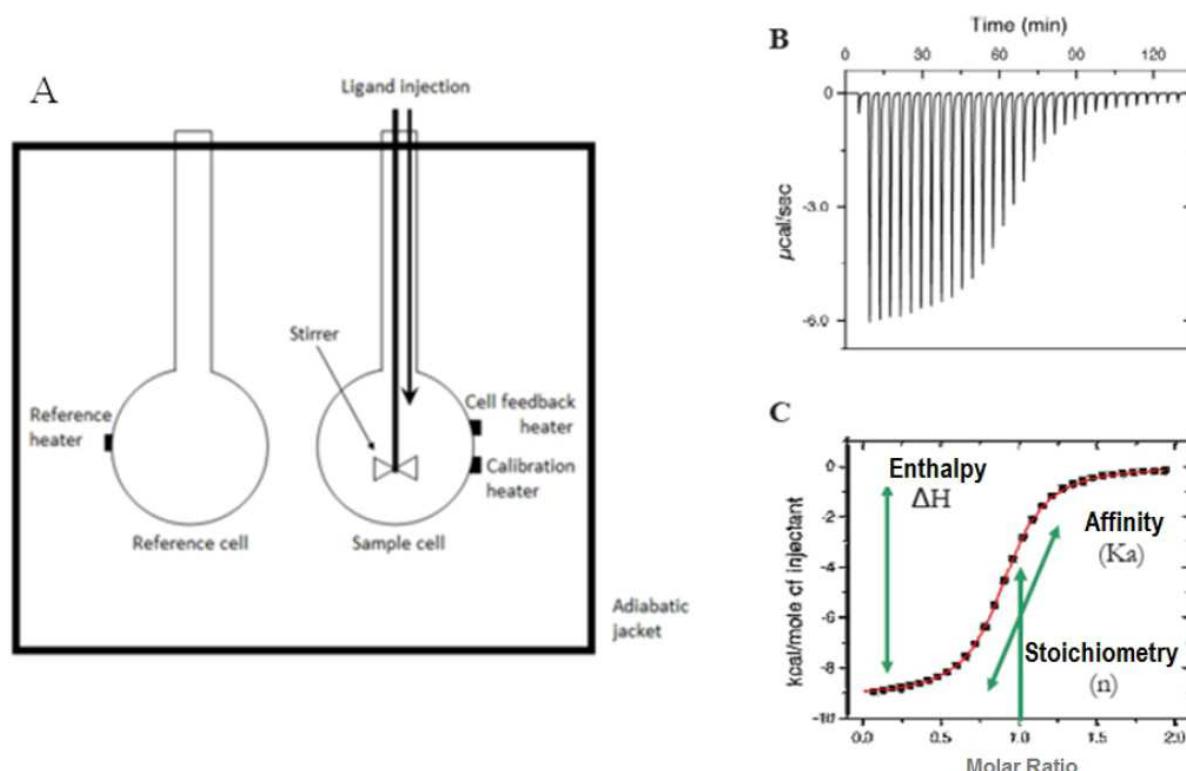
Figure 15: general principle of isothermal titration calorimetry. A) Schematic representation of the microcalorimeter. B) Representative of experimental data C) Integrated data and theoretical isotherm that allows calculating thermodynamical parameters of the binding interaction.

4.1 Affinity constant K_a and dissociation constant K_d

When a protein (P) and its ligand (L) create a non-covalent protein-ligand complex (PL), its interaction can be described as follows:

$$P + L \leftrightarrow PL$$
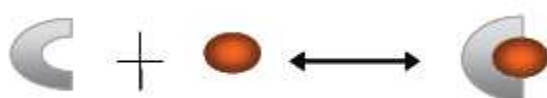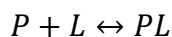
The equilibrium depends on the strength of the interaction and is defined by its association constant K_a in M^{-1}. The larger the K_a value the higher is the affinity of the protein for the ligand. Often the affinity is described by the dissociation constant K_d, which is the inverse K_a, that means K_d gets smaller with increasing affinity

$$Ka = \frac{1}{Kd} = \frac{[PL]}{[P][L]}$$

4.2 Free energy (Gibbs energy) ΔG, enthalpy ΔH and entropy ΔS of binding

The Gibbs energy, or free energy of binding, results from the total energy that is generated by the reversible, non covalent interaction. A negative ΔG describes a spontaneous interaction. The relation between K_a and the free energy ΔG is described by the following equation, where R is the gas constant (8.314 J/ mol*K) and T is the temperature in Kelvin (K).

$$\Delta G = -RT \ln Ka$$

The energy of binding ΔG is a thermodynamic value that has two contributors: enthalpy and entropy of binding. The enthalpy of binding ΔH is negative is the interaction releases energy (exothermic) and positive if it consumes energy (endothermic). Hydrogen bonds generally result in strong negative enthalpy contribution. The entropy (ΔS) is commonly referred to as a measure of disorder. It is defined in the second law of thermodynamics, which constitutes that the entropy of a closed system may never decrease. Calculating the entropic contribution (TΔS) in ITC can give information about the nature of the protein-ligand interaction and can be calculated as follows.

$$\Delta G = \Delta H - T\Delta S$$

The entropic contribution to the protein- ligand interaction can be due to the phenomenon of the dislocation of solvent (water) molecules around the ligand when it binds to the protein. If the binding is dominated by hydrophobic interactions, the entropic contribution is favourable. On the other hand, the free carbohydrate ligand may take different conformations whereas it may be forced into one configuration upon binding. In this case, the entropic contribution is unfavorable.

4.3 Experimental considerations

To setup an ITC experiment some general questions need to be addressed. To insure an accurate estimate of the enthalpy, affinity constant and the stoichiometry the concentrations of both protein and ligand need to be calculated. For high affinity ligands, the protein concentration must be low, otherwise the slope will be too steep to calculate the affinity (K_a). On the other hand, protein concentrations must be higher, in low affinity setups or the slope will be too flat. However, the protein concentration is limited by the solubility and often it is not possible to attain protein concentrations higher than 1 mM, corresponding to a K_d of 1 mM. For lower affinity the curve is no longer sigmoid, but becomes a hyperbole. From this it is still possible to calculate the affinity and the enthalpy, but the stoichiometry cannot be obtained and needs to be preset during calculations.

5 Surface Plasmon Resonance

Surface Plasmon Resonance (SPR) occurs when polarized light interacts with free electrons of a metal at a metal/dielectric interface. An electro-magnetic wave is created at the interface that propagates parallel to the metal surface. The wave influences the angle of the reflected light and is directly dependent on the changes on the metal surface. Sensor chips are available where a ligand can be attached to gold surface. Interactions of the immobilized ligand and a second component (analyte) passing over the surface in solution are monitored in real time. Upon binding of the analyte to the ligand, the refractive index changes and thus the angle of the reflected beam. This change is directly related to the analyte concentration, its molecular weight and the rate of association and dissociation to the ligand. Figure 16 gives a schematic overview of the principle of SPR (Cooper 2002)
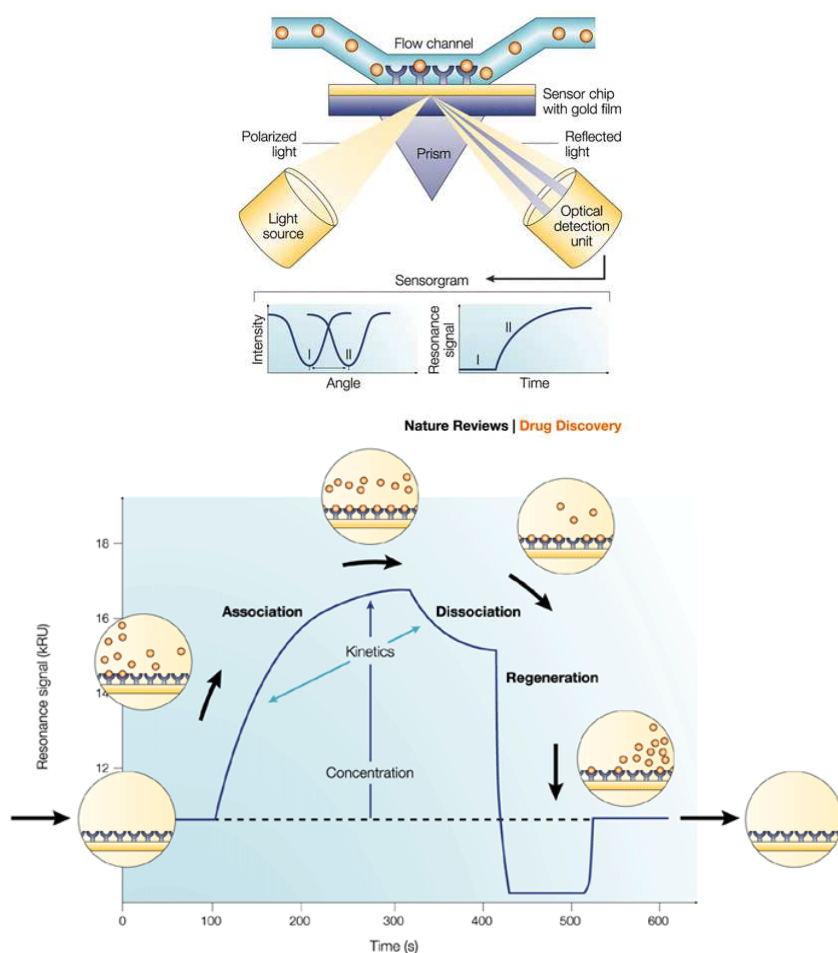
Figure 16: schematic overview of SPR. The ligand is covalently bound to the gold surface of the sensorchip and the analyte is passing in the flow channel in solution. In the first step association of the analyte to the ligand results in an increase of the response unit (RU) until the surface is saturated. In the second step the binding partners are in dynamic equilibrium and the free dissociation depends on their affinity. At last the chip is regenerated where all interactions are disrupted and the initial state of chip is restored. Pictures are reproduced from (Cooper 2002)

Measurements are usually repeated with different concentrations and varying conditions to produce a valuable dataset to calculate a binding algorithm. The affinity and kinetics of the interactions can be calculated from the ratio of the rate constants.

$$K_D = \frac{1}{K_A} = \frac{k_{diss}}{k_{ass}}$$

Furthermore thermodynamic analysis is also possible and the main advantage of SPR is the high sensitivity, the large selection of molecules that can be attached to the surface and the real-time evaluation of interaction without a need for labeling. It requires low amount of analyte which is our case is purified protein or cytoplasmic extracts. In this study a Biacore X100 (GE Healthcare) was used.

6 Glycan array

Glycan structures in nature show a huge diversity not only due to their monosaccharide composition and sequence but also because of different possible linkages and anomery. To find specific lectin carbohydrate interactions, it is therefore often necessary to use high throughput tools that allow finding natural glycan ligands for carbohydrate binding proteins. The Consortium for Functional Glycomics provides such a high throughput service on the base of glycan microarrays which allow today to screen more than 600 human glycan structures at once using as little as 200 µg ml^{-1} labeled protein. This technique has many advantages. The often low affinity of carbohydrate binding proteins to mono- and disaccharides that may hamper the investigation is circumvented. The dense presentation of oligosaccharides provides the additional advantage of a multivalent presentation, mimicking the glycoconjugates on the cell surface.

The general principle of a printed glycan microarray, as performed at the CFG is explained in figure 17 below. Glycans are covalently attached by a free amine group to an *N*-hydroxysuccinimide (NHS) or epoxide activated surface on a glass slide. The resulting glycan chip is then incubated with a labelled carbohydrate binding protein. A fluorescent label like Alexa Fluor 488 or Cyanine 5 is most commonly used, as shown in figure 17 below. However, it is also possible to detect biotinylated proteins with fluorescently labelled streptavidin, or to use a fluorescently labelled antibody for detection.
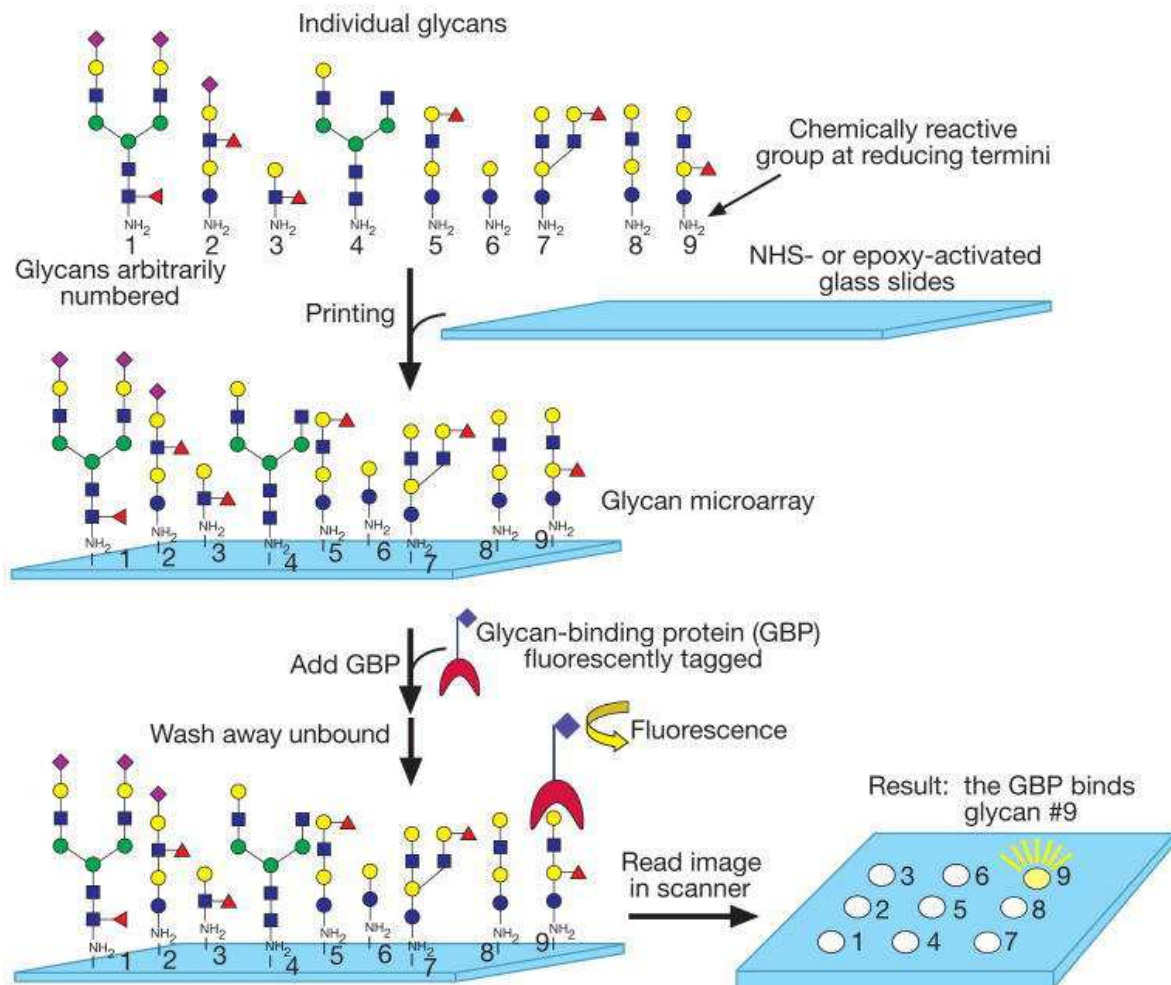
Figure 17: schematic overview of printed glycan microarrays. The glycans are covalently bound to NHS activated glass slides via a free amine at their reducing end. The glycan binding protein (GBP) is detected by fluorescence. It may be labelled directly or detected using fluorescently labelled secondary reagents. Picture reproduced from (Varki et al. 2009)

While printed microarrays are the latest development using piezoelectric or contact printers, glycan arrays where produced by attaching biotinylated glycans to a streptavidin coated microtiter plates. The major advantage of the newly developed printed arrays lies in the higher number of glycans and the minimal amount of glycan solution, making it more cost and time effective.

7 X-ray crystallography

X- ray crystallography is especially well suited to solve the three dimensional structure of biolgical macromolecules, such as proteins because hard X-rays have a wavelength of several Ångström (0.1 nm) which is within the range of atomic bonding. Interaction of the X-ray with any object will cause diffraction, however single macromolecules will result in weak diffraction. This problem is circumvented by using crystal arrangements of macromolecules. In crystal form the protein is arranged in a periodic manner, which results in an increase in signal intensity.

The resolution of the crystal structure of a protein can be divided into three steps: firstly pure and monodisperse protein solution is required. Secondly a monocrystal of the protein has to be obtained that diffracts well. At last, the phase problem has to be addressed either by experimental phasing methods or molecular replacement.

7.1 Crystallogenesis

A crystal can be described as an orderly three-dimensional periodical repetition of a motif. To crystallize a protein, means to change its arrangement from a soluble disordered phase to an ordered crystalline state. Protein crystallization can therefore be described as a phase transition. This phase transition is depended on many parameters and takes place in two steps: nucleation and crystal growth. To induce nucleation the protein has to be in a supersaturated state. By slowly diminishing the protein solubility, the interactions between protein molecules become more favourable than protein-solvent interactions. While non-specific interactions result in amorphous precipitations, specific protein-protein interactions may lead to crystal nuclei.
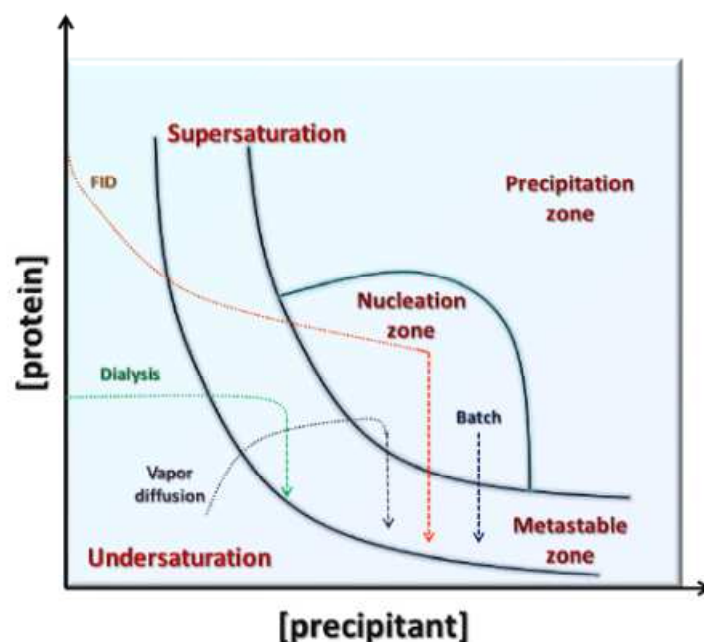
Figure 18: Schematic representation of a protein phase diagram. Under the influence of the concentration of the protein and the precipitant, the system reaches the nucleation zone. With the growing of microcrystals, the protein concentration diminishes and the system reaches the metastable zone where crystal growth is favourable.

To bring the protein solution into a supersaturated state, many different parameters have to be taken into consideration, such as the purity, composition and concentration of the protein solution, the pH and the nature and concentration of precipitants and additives used. Furthermore, the temperature may be varied to alter the protein solubility.

The precipitant plays an important role in the production of protein crystals by directly affecting solubility. Commonly used precipitants can be categorized into three groups which may be used separately or in combination. Polymers act by excluding solvent molecules from their interaction with the protein, whereas salts modify the ionic strength of the solution and promote hydrophobic interactions between the protein molecules. Organic solvents change the dielectric constant of the solvent and play a role in the charges of protein and solvent. Besides these major precipitants, it may be necessary to include further additives that perturb and manipulate protein-protein and protein-solvent interactions.

Crystals that were obtained in this work were carried out by hanging drop vapour diffusion. Even though different approaches to obtain crystals exist, vapour diffusion is the most commonly used among them. Dialysis, batch technique and free interface diffusion will not be explained here.
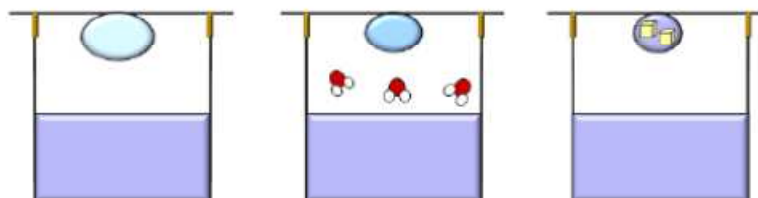
Figure 19: hanging drop vapor diffusion. The hanging drop contains a mixture of protein solution and reservoir solvent. Water diffuses from the drop into resulting in an increase in protein concentration in the drop.

In vapour diffusion (figure 19) 1-2 µl of protein solution is mixed with 1-2 µl of the reservoir solution. The reservoir contains solvent including precipitant(s) and/or buffer and/or other additives. The ratio of protein and reservoir solution may be varied. Initial crystal screening conditions often result in micro-crystals, too small for data collection, or crystals that do not diffract sufficiently to solve the structure. In these cases, optimization is needed and it can be carried out with many different methods. The easiest method to try to increase the crystal size is to optimise the initial conditions by varying the concentration of the protein or precipitant reagents, vary the precipitant nature, screen the pH range or include additives. These additives can be non-ionic detergents, chaotropic or metallic salts or redox-agents, such as DTT. Correct and efficient optimization is often a matter of experience and has to be evaluated for each individual experiment. A commonly used optimization method is crystal seeding where microcrystals are used as nucleation points for crystal growth. Once crystals of sufficient quality are obtained, X-ray data can be collected.

Since X-rays are high in energy and may cause severe damage to the biological sample, the crystal is frozen in a loop in liquid nitrogen and the collection is carried out under constant nitrogen flux. The addition of cryoprotective substances can be necessary to avoid the formation of ice crystals that could destroy the crystal arrangement. X-ray data were collected at the European Synchrotron Radiation Facility (ESRF) in Grenoble.

During the collection the turns perpendicular to the beamline, by several degrees depending on the symmetry and the diffraction spots are collected by a couple charge detector (CCD).

Depending on the beamlines (insertion device or bending magnet) and the crystal symmetry, a complete dataset may be obtained in no more than several minutes.

7.2 Symmetry elements

A crystal is the result of the translation of the unit cell in all three dimensions. The unit cell is the smallest volume that is a complete representative and contains all the information of the crystal. Each unit cell is defined in terms of lattice points and by its lattice parameters which are the length of the cell edges a, b and c and the angles between them α, β and γ while the positions of the atoms inside the unit cells are given by the set of spatial coordinates (x, y and z) measured from the lattice points. Only 14 different types of crystal lattices, called Bravais lattices can be obtained in three dimensions. Symmetry operations can be applied to the unit cell such as rotation, inversion or reflection. 32 possible symmetry operations, combined with the 14 Bravais lattices yield in 230 possible space groups. However, the number of space groups is limited to 65 in protein crystallography due to the fact that biological macromolecules are chiral molecules which can not contain inversion centre or mirror plane. The asymmetric unit is defined as the smallest molecular unit that yields to the unit cell by application of the symmetry operations, described by the space group. Figure 20 shows the 14 Bravais lattices with their possible space groups in protein crystallography.

Figure 20: crystal systems and space groups. There are 14 possible lattice arrangements in three dimensions, named Bravais lattices and 65 possible space groups for biological macromolecules.

7.3 X-ray diffraction

Diffraction of X- rays is possible due to their wave-like behaviour. When X-rays are scattered by a crystal lattice, peaks can be observed when the waves combine to constructive interference. This phenomenon can be described by Bragg`s law. Bragg`s law is defined by the following equation:

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

Where λ is the wavelength of the incident X-ray beam, d is the distance of the atomic planes in the crystal, θ is the angle of incidence and n is an integer. In other words, peaks of scattered intensity are observed, when the angle of incidence is equal to the angle of scattering and the pathlength difference is equal to an integer number of the wavelength. The conditions for maximum intensity, contained in Bragg`s law, allow to calculate the details about the crystal structure.
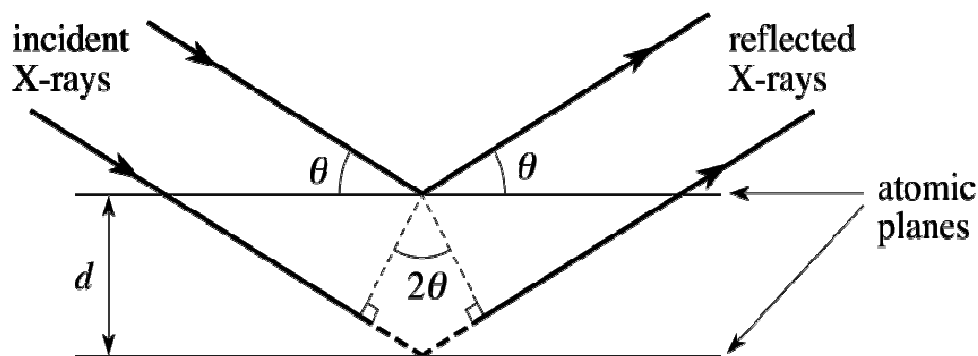
Figure 21: schematic representation of Bragg`s law

7.4 Data processing

During data collection, a set of images with the position and intensities of reflections are recorded. The position of the reflection points provides information about the space group and unit cell geometry, while their intensities are related to the distribution of electron density in the asymmetric unit. The dataset is analysed using computer programs such as MOSFLM (Leslie, 2007) or XDS (Kabsch 2010) which assign *Miller indices* (*h,k,l*) and an intensitiy *I* associated with an experimental errorσI to each reflection point. The Miller indices describe the coordinates of the atomic planes in reciprocal space. All reflections are then integrated by

the software and the intensities *I* are scaled and merged before conversion to structure factors F_{hkl} using the CCP4 program suite (Winn et al. 2011).

To evaluate the quality of the experimental data set, the parameters of completeness, multiplicity and intensity error should be examined. Due to symmetry relations, some reflections are equivalent and should have the same intensity. Because only *unique* reflections are necessary to obtain the distribution of electron density in the unit cell, the calculated *completeness* is a percentage of the unique reflections measured. The ratio of total reflections and unique reflections gives the value of multiplicity. The third important parameter is the ratio between the Intensity of a reflection and its error *I/σI*.

7.5 Structure factors and electron density

Each reflection is characterized by its amplitude and phase. To obtain the structure of a diffracting motif, in this case the distribution of electron density in the asymmetric unit it is necessary to calculate the Fourier transform of the structure factors. The structure factors represent the reflection amplitudes and phases and are described as follows:

$$F_{hkl} = \int_V \rho\left(x, y, z\right) \times e^{2\pi i \left(hx + ky + lz\right)} dV$$

The electron density *ρ(x,y,z)* is therefore the inverse Fourier transform of the structure factor F_{hkl}. The calculation of the electron density at the point *(x, y, z)* of a given cell with the volume *V* must be a summary of all waves diffracted by the planes *hkl*. Their amplitudes depend on the number of electrons in each plane.

$$\rho\left(x, y, z\right) = \frac{1}{V} \sum_{hkl} F_{hkl}\, e^{-2\pi i(hx + ky + lz)}$$

The electron density can be represented as a propagating wave with its amplitude $|F_{hkl}|$ which is proportional to the measured intensity and a phase φ_{hkl}. While the amplitude can be calculated from the measured intensities, the phases are not known. The phases have to be obtained experimentally by multiple isomophous replacement (MIR), single-wavelength anomalous diffraction (SAD) or multi-wavelength anomalous diffraction (MAD). If a homologous structure of the protein is known, the phases may be obtained by molecular

replacement, where the phases of the known protein structure are used to calculate the structure factors for the unknown protein.

7.6 Molecular replacement

To use molecular replacement a structure of a homologous protein is needed which shares at least 30% of sequence identity that can be used as a model to estimate a first set of phases and to calculate the structure factors. However, sequence identity is not sufficient, since structural similarity is the most important feature for successful molecular replacement.

The position of the model protein in the unit cell is needs to be determined. The Patterson function is an inter-atomic vector map (Patterson-map) that can be calculated independently from phases with the following equation:

$$P(u, v, w) = \frac{1}{V} \sum_{hkl} |F_{hkl}|^2 e^{-2\pi i(hu+kv+lw)}$$

The resulting Patterson map contains a peak for each atom related to every other atom and comparison of the Patterson map of the model and the unknown protein allows correlating their orientation in the asymmetric unit by determination of the Euler angles ($\theta_1, \theta_2, \theta_3$). Having placed the model by rotation and translation into the correct position within the new asymmetric unit, its known phases can be used to calculate the structure factors for the unknown protein and an initial electron density map. In this work, the program PHASER (McCoy et al. 2007) was used.

7.7 Refinement and Validation

Once phases are obtained and the initial electron density map is calculated, the atomic positions have to be refined. While the initial map may not be sufficient to precisely position all atoms, it usually allows for initial modelling. The new model is then used to calculate a new set of phases and to obtain a new and improved density map. The goal of the refinement is to eliminate errors, which means to minimize the differences between the calculated structure factor from the model and the observed structure factors. Refinement was carried out using REFMAC (Murshudov et al. 2011), while the inspection of the electron density map and manual model building where done using Coot (Emsley & Cowtan 2004).

The correspondence between the model and the experimental data is described in the R-factor with the following equation

$$R = \frac{\sum_{hkl}|F_{obs}(hkl)| - |F_{calc}(hkl)|}{\sum_{hkl}|F_{obs}(hkl)|}$$

Besides the R-factor, a related parameter is used to validate the accuracy of the refinement. The R_{free}-factor is calculated the same way but is used for a set of reflections (5-10%) that are omitted in the refinement process. For successful refinement, R- and R_{free}-factors should drop or converge in parallel.

Chapter III

Identification of novel lectins in

Aspergillus fumigatus

Summary

This chapter presents two approaches to identify novel lectins in *A. fumigatus*. New lectins may be identified either by stepwise purification from natural extraction or by searching the genome for proteins with sequence similarity with other known lectins. The first approach was tested using mycellial extracts from *A. fumigatus* that were provided from our collaborators in Angers, where the extracts were prepared and frozen, before being sent to Grenoble. The extracts exhibited low hemagglutination activity, which hampered the identification and purification of lectins. Nonetheless, a putative lectin, named AFL5 was identified, recombinantly expressed, and purified. AFL5 showed positive hemagglutination activity in PBS and HEPES buffers but not in TRIS-containing buffers and only a fraction of the purified protein was active, as determined using gel-filtration followed by hemagglutination assays. Furthermore, hemagglutination activity showed drastic variations in between expression batches, even though expression and purification were carried out using the same protocol. The carbohydrate specificity of the active fractions was further analyzed by hemagglutination inhibition assays. The glycoprotein fetuin was able to inhibit agglutination at a minimal concentration of 625 µg ml^{-1}, but none of the tested monosaccharides (glucose, mannose, fructose, galactose, *N*-acetylglucosamine, *N*-acetylgalactosamine) showed inhibition at concentrations lower than 100 mM.

The second approach to identify new lectins was carried out by screening the translated genome of *A. fumigatus* for proteins that show sequence similarity with other known fungal lectins. The BLAST algorithm (Altschul & Madden 1997) was used to search the genome of *A. fumigatus* strains 1163 and af293. The resulting hits were analyzed by their bit score and Expect (E) value; significant matches were further investigated by multiple alignments using ClustalW (Thompson et al. 1994). Using this procedure, four novel putative lectins were identified, namely one R-type lectin and three cyanovirin homologs (CVNH) of type I and

one of type II, whereas one of the type I CVNH was only present in the 1163 strain. The other type I CVNH was named AFL6 and its characterization will be described in chapter IV.

Résumé

Ce chapitre décrit deux approches visant à identifier de nouvelles lectines chez *A. fumigatus*. Les nouvelles lectines peuvent être identifiées soit par purification à partir d'extraits bruts naturels soit par analyse du génome combinée à la recherche de similarité de séquence avec des lectines connues. La première approche a été utilisée à partir d'extraits de mycelium d'*A. fumigatus* fournis par nos collaborateurs à Angers. Les extraits bruts ont été préparés et congelés à Angers, puis envoyés à notre laboratoire à Grenoble. Ils présentent une faible activité d'hémagglutination, ce qui a entravé l'identification et la purification. Néanmoins, une lectine potentielle, nommée AFL5 a été identifiée et exprimée sous forme recombinante puis a été purifiée. AFL5 a montré une activité d'hémagglutination dans le PBS et les tampons HEPES mais pas dans ceux contenant du TRIS. Seule une fraction de la protéine purifiée semble active, d'après les expériences de filtration sur gel suivies de tests d'hémagglutination. En outre, l'activité d'hémagglutination semble varier considérablement entre les lots, alors que l'expression et la purification ont été effectuées de façon identique. La spécificité des fractions actives pour les monosacharides (glucose, mannose, fucose, galactose, N-acétylglucosamine et N-acétylgalactosamine) et pour les glycoprotéines a été évaluée par des tests d'inhibition de l'hémagglutination. La fétuine a été capable d'inhiber l'agglutination à une concentration minimale de 625 µg ml^{-1}, mais aucun des monosaccharides testés n'a montré d'inhibition pour des concentrations inférieures à 100 mM.

La seconde approche d'identification de nouvelles lectines chez *A. fumigatus*, a été réalisée en recherchant dans le génome traduit du champignon des protéines présentant une similarité de séquence avec des lectines fongiques connues. L'algorithme BLAST a été utilisé pour une recherche sur le génome des souches 1163 et af293 d'*A. fumigatus*. Les résultats obtenus ont été analysés d'après leur score et les valeurs attendues associées (E-value); des alignements multiples utilisant ClustalW (Thompson et al. 1994) ont été réalisés sur les séquences les plus intéressantes. En utilisant cette procédure, quatre nouvelles lectines putatives ont été identifiées, dont une lectine de type ricin-B et trois homologues de la cyanovirine (CVNH) de type I et de type II. La lectine CVNH de type I a été nommée AFL6 et sa caractérisation fait l'objet du chapitre IV.

1. Introduction

1.1 Lectin identification methods

1.1.1 Purification from natural extracts

Identification of novel lectins from natural extracts has been traditionally carried out using hemagglutination tests. Indeed, lectins were first discovered by their ability to clot red blood cells and described as agglutinins (Boyd & Shapleigh 1954; Stillmark 1888). Hemagglutination tests are therefore the basis for any new discovery of lectins from nature. Using red blood cells from different animals (e.g. mouse, rat, rabbit, chicken) and different human blood types, offers an interesting tool for an initial characterization. Subsequently, hemagglutination inhibition tests are performed using saccharides or glycoproteins to provide information about their binding patterns.

In mushrooms, lectins are often found in the fruiting body. Crude fungal extracts are obtained by grinding and homogenizing the fruiting body prior to re-suspension of the material in phosphate buffered saline (PBS) to extract lectins (Kawagishi et al. 1994). For filamentous fungi, such as *Aspergillus*, screening for lectins was carried out in culture supernatant to detect secreted lectins, and also in extracts from mycelium, hyphae and conidia. Furthermore, the physiological stages of the fungi were taken into account by analyzing resting and swollen conidia as well as germ tubes (Singh et al. 2008; Singh et al. 2010; Tronchin et al. 2002).

The most effective way of purification can be carried out by affinity chromatography using sugar based gel matrix, which was first described by Goldstein and coworkers (Goldstein & Agrawal 1965). Since then, most published purification methods include a step of affinity chromatography. Most commonly used affinity ligands are glycoproteins such as fetuin, avidin and mucin, but saccharides such as mannose, chitin and lactose are used as well (Nascimento et al. 2012). Typically the ligands are selected based on the lectin specificity, as derived from hemagglutination inhibition assays.

Other purification protocols include an initial step of concentration by salting out proteins with ammonium sulfate, followed by chromatography using ion-exchange, hydrophobic interaction or gel filtration. A systematic review of purification methods for 46 lectins from plants, fungi, algae, marine sponges, mushrooms and fish showed that affinity chromatography and ammonium sulfate precipitation were most commonly used as a first purification step. These first steps were then followed by ion-exchange and hydrophobic

interaction chromatography and the total number of purification steps varied between one and five (Nascimento et al. 2012). Gel filtration was the least frequently used chromatography method and was never used as a first purification step in the analyzed publications. Overall recovery yields were shown to be greatly dependent on the source of the lectin, but also the number of purification steps appeared to be a determining factor. It could be observed that the yields of recovered lectin diminished with increasing purification steps.

Besides chromatographic methods, recent advancements have been made using magnetic nanoparticles, coated membranes and affinity precipitation. Magnetic nanoparticles were successfully used to purify the well known lectin ConA from castor bean, using glucose or mannose coated iron oxide or maltose coated magnetic nanoparticles (Slesiona et al. 2012; Hubbuch et al. 2001; Banerjee & Chen 2007). Different galactose specific lectins were used to develop a new technique using cellulose- based affinity membranes which were functionalized with different affinity ligands (Boi et al. 2006). Furthermore affinity precipitation using macroligands combined with partitioning methods, such as two phase affinity extraction was successfully used to recover lectins from tomato, wheat germ and potato (Teotia et al. 2006).

In this work we used classical chromatographic methods as described above to recover a previously described sialic-acid specific lectin from *A. fumigatus* with a molecular mass of 32 kDa. This lectin was discovered in fungal extracts prepared from mycelia, resting or swollen conidia and germ tubes. Even though the highest specific hemagglutination activity was found in resting conidia, the purification was carried out from frozen mycelium because relatively large amounts of starting material were necessary for purification. Successful purification of the lectin was achieved by initial gelfiltration, followed by ion-exchange chromatography and hydrophobic interaction chromatography, Hemagglutination inhibition tests using rabbit erythrocytes showed specificity for sialic acid-rich glycoproteins and Neu5Ac (Tronchin et al. 2002).

Our aim was to reproduce the purification from mycelia extracts and identify the gene for recombinant expression to provide biochemical and structural characterization of this sialic-acid specific lectin from *A .fumigatus*.

1.1.2 Identification using bioinformatic tools

Recent advances in genome sequencing have made way for extensive bioinformatical research concerning gene identification, annotation and the search for homologous proteins. With the availability of whole genome sequences from important pathogens, several lectins were identified by this approach. In bacteria, lectins from the pathogen Mycobacterium tuberculosis had been successfully identified through bioinformatical approaches (Desh Deepak Singh et al. 2007). In an extensive screening of 30 fully sequenced mycobacterial genomes of 20 distinct species, 64 sequences containing predicted lectin domains have been identified (Abhinav 2013). In their effort to find novel lectins in mycobacteria, the authors firstly annotated conserved domains within the open reading frames (ORF) using the conserved domain database (CDD) web server at NCBI (Marchler-Bauer et al. 2005) to identify ORFs that contain carbohydrate binding domains, followed by a template based homology modeling using PHYRE and I-TASSER. This structural based homology search was completed by a sequence based homology search (Abhinav 2013). Bioinformatical approaches were not only used to identify novel lectins but also to investigate origin, evolution and biological role of carbohydrate binding modules (CBM) in fungi and oomycetes (Larroque et al. 2012). By screening the complete genomes of 60 fungi from different lineages and 7 oomycetes the authors identified 518 CBM1-containing proteins, most of them associated with cellulose degradation (Larroque et al. 2012). Phylogenic trees were created based on bioinfomatical analysis of 22 known fungal lectins and they were categorized into three groups based on the carbohydrate binding domains (Kahn et al. 2011). Furthermore, a siglec-like sialic acid binding domain was identified in an adenovirus capsid protein, which showed no apparent sequence similarity with sialic acid binding immunoglobulin-like lectins (Siglecs). By identifying common recognition features in the local area of the sialic acid recognition site in crystal structures, a database screening process found not only the Siglec like proteins but also the fiber knob protein of canine adenovirus-2 (Rademacher et al. 2012).

For genome screening in *Aspergillus* species, the comprehensive web server Aspergillus Genome Database (AspGD) is available with extensive annotations of ORFs, conserved domains and identification of orthologs (Arnaud et al. 2012). BLAST searches can be carried out in the translated genomes of *Aspergillus* species and the genomes of *A. fumigatus*, *A. nidulans*, *A. niger* and *A. oryzae* can be browsed by landmarks or regions of the chromosomes. Using the BLAST tool of this web server, hits can be found with alignment,

annotation and literature references, if available. Furthermore, the presence of orthologs in other *Aspergillus* species is indicated. This *Aspergillus* specific website and its servers and databases were used in this thesis for bioinformatical search of lectins sharing sequence similarity with known fungal lectins. The search was designed with the guide of the comprehensive CERMAV 3D Lectin database under the category "yeast and fungal lectins" in association with PDB and NCBI databases to obtain the query sequences.

1.2 Lectins in *Aspergillus*

The vast *Aspergillu*s genus includes more than 180 species. While some of them are medically or commercially important, some of them are known to be opportunistic pathogens that can infect humans and other animals. Among the *Aspergillus* species that are commonly implicated in human or animal infections are *A. fumigatus*, and *A. flavus.* Studying lectins from *Aspergillus* spp is therefore not only useful to understand host- pathogen interaction, but also to identify possible tools for diagnostics and biotechnology. So far, 25 *Aspergillus* species other than *A. fumigatus* have been screened for lectin activity using hemagglutination assays, and among these, 13 exhibited lectin activity in mycelia grown in liquid and solid cultures (Singh et al. 2008; Singh et al. 2010). In these screening attempts, classical methods using hemagglutination assays were applied to identify the presence of lectins in culture supernatant and mycelia from liquid and solid culture. Using red blood cells (RBCs) from different animals as well as different human blood group combined with hemagglutination inhibition assays, the apparent specificity of these lectins was determined. In the first screening of 10 different *Aspergillus* species four species exhibited intracellular lectin activity namely *A. niger*, *A. versicolor*, *A. rugulosus* and *A. nidulans*. All of the identified lectins were found to agglutinate human erythrocytes from all blood types as well as porcine erythrocytes. Hemagglutination was inhibited by glycoporteins such as inulin, mucin and asialofetuin as well as the carbohydrates *N*-acetylglucosamine, melibiose, D-ribose, L-fucose, D-arabinose, D-sucrose and D-mannitol (Singh et al. 2008). In the second screening, 15 different *Aspergillus* species were analyzed and nine of them showed lectin activity in mycelia with the ability to agglutinate rat, rabbit, pig and human red blood cells with no preference for human blood groups. The species exhibiting lectin activity were *A. sydowii*, *A. candidus*, *A. allahabadi, A. terricola, A. ficuum, A. sparsus, A. carneus, A. pulvinus* and *A. aculeatus*. The lectins agglutinated all human RBC with the exeption of *A. pulvinus* , which showed specificity for type A and O RBCs. Lectins from all analyzed species showed specificity to the glycoproteins mucin and asialofetuin. The use of various sugars in

hemagglutination inhibition tests showed that L-arabinose was recognized by all lectins except that from *A. carneus* and that D-xylose was recognized by lectins from *A. sydowii, A. ficuum, A. sparsus* and *A. carneu*s. (Singh et al. 2010).

In *A. fumigatus,* two lectins have been identified so far (Tronchin et al. 2002; Kuboi et al. 2013). As mentioned above , the first lectin was discovered in mycelia, resting and swollen conidia and germ tubes and was partially characterized in 2002. It showed an apparent specificity for sialic acid containing glycoproteins like mucin, fetuin, thyroglobin and fibrinogen as well as sialic acid (Neu5Ac) (Tronchin et al. 2002).

A fucose specific lectin named FleA was initially identified in cell lysate from *A. oryzae* (Ishida et al. 2002) and its ortholog was discovered later in native culture of *A. fumigatus* and named AfuFleA (Kuboi et al. 2013; Houser et al. 2013). These fucose specific lectins, show sequence and structural homology with the lectin AAL from the fungus *Aleuria aurantia* (Wimmerova et al. 2003).

Structural and biochemical analysis of AfuFleA, now named AFL, were carried out to elucidate its possible role in host recognition and to investigate the carbohydrate recognition.

AFL was found to be localized in *A. fumigatus* conidia, most probably on the surface. Thus it may provide an interface with the host tissue for recognition of fucose- containing glycoconjugates. Furthermore, it was shown that AFL stimulation of human bronchial cells had an immunostimulating effect, suggesting a possible role in virulence (Houser et al. 2013)

Like its homolog in the mushroom *Aleuria aurantia* AAL, AFL is made up of six tandem repeats, that form a six-bladed β-propeller. Both lectins build homo-dimers. The six binding sites of AFL are located between two neighboring blades of the monomer. The binding motif is comprised of conserved Arg and Glu/Gln residues, establishing polar contacts with the hydroxyl O-3,O-4 and O-5 of fucose and a conserved Trp/Tyr residue that mediates a CH-π interaction with the ring (C-3,C-4 and C5).

Interestingly, minor changes in the amino acid composition of the binding sites cause differences in the binding pattern as shown by co-crystallization with Lewis[Y]. While binding sites 1, 2 and 6 predominantly showed binding to the α1-2 linked fucose, site 3 favored the α1-3 linked fucose and binding site 5 binds both. Binding site 4 was not big enough to accommodate the whole tetrasaccharide. These changes in the binding pattern were only observed in for large epitopes. The studies suggest, that AFL is able to bind fucose is all possible linkages, which can be found not only in decomposing organic matter but also on human tissues highlighting the possible role of AFL in host recognition and early stages of infection (Houser et al. 2015).

The studies on AFL represent the first detailed structural characterization of a lectin from a pathogenic mold and are an important example of the importance of such investigations to understand host-pathogen interactions for future inhibitor development and treatment strategies.

2. Experimental Details

2.1 Purification from fungal extracts

Fungal extracts from *Aspergillus fumigatus* CBS 113.26-1 were kindly provided from the group of Jean-Philippe Bouchara at the University of Angers. They were sent on dry ice to Grenoble where they were stored at -80° C until purification.

2.1.1 Purification by affinity:

In a first attempt to purify the putative sialic acid-specific lectin from *A. fumigatus,* 1 ml fetuin-resin (EY laboratories) was packed into a column. Fetuin is a glycoprotein that contains glycans with terminal sialic acids both 2,3 and 2,6 linked. The resin was equilibrated with PBS and 50 ml of the crude fungal extract containing 4-5 mg total protein was slowly thawed in an ice cold water bath. It was centrifuged at 15000 x g and the supernatant was applied to the column. The flow through was collected and the resin was washed with 10 ml PBS. Elution was carried out in three steps. In a first step, the salt was increased to 1 M NaCl and then to 1.4 M. The last elution step was carried out using 0.1 M *N*-acetylneuraminic acid (NeuAc). 10 µl of each elution fraction was tested for protein content with Bradford reagent.

2.1.2 Purification by separation

The purification was carried out by adjusting the published purification method that led to the discovery of a sialic acid specific lectin in 2002 by the above mentioned collaboration partners in Angers (Tronchin et al. 2002). The protocol comprises a stepwise purification using ammonium sulfate precipitation, ion-exchange and hydrophobic interaction. 50 ml of frozen fungal extract from *Aspergillus fumigatus* CBS 113.26 containing 4-5 mg total protein was slowly thawed in an ice cold water bath and the samples were kept on ice throughout the purification procedure to minimize protein degradation. A sample of 50 µl was taken for hemagglutination tests.

Ammonium sulfate was slowly added with thorough agitation to 80% (0.52 g ml^{-1}) saturation and then incubated at 4 ℃ for 1h. The precipitated proteins were collected by centrifugation

at 20000 x g for 20 min. The pellet was resuspended in 3-5 ml 20 mM TRIS buffer (pH 8.5) and dialyzed against the same buffer for 24 h to ensure the removal of ammonium sulfate for the following ion exchange chromatography. A sample of 50 µl was taken for hemagglutination assay.

Chromatography

Ion exchange chromatography was carried out with the strong anion exchanger HiTrap ® QSepharose FF 1 ml (GE Healthcare) in 20 mM TRIS (pH 8.5) and the elution was carried out with a stepwise increase in NaCl to a final concentration of 1 M; fractions of 1 ml were collected and 50 µl of each fraction were tested for hemagglutination activity.

Peak fractions were pooled and further purified by hydrophobic interaction chromatography. Ammonium sulfate was slowly added to a concentration of 1.4 M and insoluble proteins were removed by centrifugation after 1 h incubation at 4 ℃. The supernatant was loaded on HiTrap® PhenylHP 1 ml column (GE Healthcare) equilibrated with 50 mM TRIS pH 8.5 containing 1.4 M ammonium sulfate. Elution was carried out by gradually decreasing the concentration of ammonium sulfate to none. Fractions of 1 ml were collected and a sample of 50 µl was taken for hemagglutination assay. Samples were furthermore analyzed by SDS PAGE to visualize the purity and the molecular weight.

Hemagglutination assay

Hemagglutination assays were done using rabbit erythrocytes (RBC) purchased from Biomérieux, Lyon. A 2% solution in 150 mM NaCl was freshly prepared from the 50% stock solution before each hemagglutination assay.

Inhibition tests were carried out with the monosaccharides glucose, fucose, mannose, galactose, GlcNAc and GalNAc and the glycoprotein fetuin. The putative inhibitors were diluted serially in two-fold manner with an initial concentration of 400 mM and incubated with the minimal concentration of rAFL5 that caused RBCs to agglutinate.

N-terminal sequencing

After purification from fungal extract, the first amino acids of the purified putative lectin were determined by N-terminal sequencing. To be concentrated, the proteins were precipitated by addition of 1 volume 100% trichloroacetic acid (TCA) per 4 volumes of protein solution. The pellet was resuspended in 20 mM TRIS buffer pH 8.5 and SDS PAGE loading buffer. Proteins were then subjected to SDS PAGE and subsequently transferred onto a membrane following the protocol provided by the *Institutde Biologie Structurale Jean-*

Pierre Ebel (IBS, Grenoble). Briefly, the PVDF membrane was activated in 100% methanol for 1 min. and the semi dry western blot was carried out in 10 mM CAPS pH 11 containing 10 % methanol. Protein bands were revealed with a staining solution containing 0.1% Coomassie R250, 40% methanol and 1% acetic acid. The membrane was then sent to the IBS and the first amino acids of the desired band were determined using standard Edman degradation techniques.

The gene corresponding to the purified protein was identified using tBLASTn (protein sequence against translated nucleotide sequence) on the translated genome of *A. fumigatus*.

2.2 Recombinant expression and purification of the putative lectin

The gene *Afu8g00550*, henceforth called *AFL5* was generated at GenScript with suitable restriction sites and codon optimization for recombinant expression in *E. coli* BL21 (DE3). Molecular weight, theoretical PI and extinction coefficients of rAFL5 were calculated using the ExPasy tool ProtParam (http://web.expasy.org/protparam/).

Upon arrival of the plasmid, the lyophilized DNA was recovered in 20 µl nuclease free water and 1 µl was transformed via heat shock in calcium competent *E. coli* XL-1 for plasmid amplification. Purification of the plasmid was carried out from 10 ml liquid LB culture with Wizard ® SVMinipreps (Promega) according to their standard protocol. 5-10 µg DNA were digested with 1 U µg⁻¹ DNA of each *NdeI* and *XhoI* fast digest enzymes from Thermo Scientific in the universal green buffer. After 30 min. incubation at 37˚C the samples were loaded on 0.8% agarose gel containing 3 µl of the intercalating dye GelRed (Thermo Scientific). Electrophoresis was carried out in TBE buffer at 80V for approximately 1 h until the bands were separated. The band with the size 825 bp corresponding to the rAFL5 insert was cut out with a scalpel and the DNA was extracted from the gel using Wizard ® SV Gel and PCR Clean-Up System from Promega following their standard protocol. The concentration of the recovered DNA was measured using NanoDrop ND-1000 (Thermo Scientific). The vector pET-TEV derived from pET28a where the thrombin cleavage site was replaced by a TEV cleavage site, was digested and purified the same way (Houben et al. 2007).

Vector and insert were combined in a 1:10 molar ratio and the total volume did not exceed 10 µl. Ligation was performed by adding an equal volume of TaKaRa Ligation Kit Mighty Mix (Ozyme). After 1 h incubation at room temperature, 10 µl were transformed into

Calcium competent *E.coli* XL-1 for plasmid amplification. Cells were grown over night at 37˚C on LB-Agar plates containing 30 µg ml^{-1} kanamycin for selective growth. The next day 10 ml of LB-broth with 30 µg ml^{-1} kanamycin were inoculated with single colonies and cells were grown at 37˚C with agitation overnight. The plasmid was then purified with Wizard ® SV Minipreps (Promega) and digested with the restriction enzymes *NdeI* and *XhoI* as described above to verify the presence of the insert.

The plasmid pET-TEV containing *AFL5* was then transformed into *E. coli* BL21 (DE3) strain. Cells were grown until OD$_{600}$ of 0.6 - 0.8 and protein expression was induced with 0.1 mM IPTG. The temperature was reduced from 37℃ to 16℃ and the culture was incubated overnight. rAFL5 containing bacteria were harvested by centrifugation and the pellet was frozen at -20℃ until protein purification.

The frozen pellet was resuspended in 15-20 ml Buffer A containing 20 mM TRIS/ HCl pH 7.5, 500 mM NaCl and 10 mM imidazole. The cells were disrupted at 1.9 kbar with the One Shot Cell Disrupter (Constant Systems Ltd.) and centrifugation at 45000 g separated the cell debris, so that the supernatant contained the recombinant His-tagged rAFL5. Affinity chromatography was carried out using Fast Protein Liquid Chromatography (FPLC, GE Healthcare) where the clear cell lysate was loaded onto a 1ml His-Trap™ column (GE Healthcare) for immobilized metal affinity chromatography (IMAG). His-rAFL6 was eluted with a linear gradient of 10–500 mM imidazole. Peak fractions were pooled and concentrated in a centrifugation- ultrafiltration unit (Vivaspin, Sartorius) to a concentration higher than 1 mg ml^{-1} for subsequent TEV cleavage. The His-tag was cleaved off using the protease from Tobacco Etch Virus (TEV) in Buffer A with the addition of 0.5 mM EDTA, 0.25 mM TCEP and 1:50 TEV protease: His-rAFL5 ratio. After incubation at 10 ℃ overnight the TEV protease and uncleaved His-rAFL5 were separated by IMAG and cleaved rAFL5 was collected in the flow through. For crystallization studies, AFL5 was further purified by gel filtration with a S75 16/60 column (GE Healthcare) in 20 mM HEPES pH 7.5, 100 mM NaCl. For other studies, gel filtration was carried out in PBS.

To verify the expression and purification of rAFL5 and to monitor its purity, each step was analyzed by SDS-PAGE using 12% gel. Samples of 16 µl were mixed with 4 µl 5x loading buffer and denatured at 95˚C for 5 min before loading. As a reference for the molecular weight Precision Plus Protein™ Standards Unstained (Bio-Rad) was loaded alongside and electrophoresis was carried out at 200V. The gel was then immersed in a Coomassie based staining solution (InstantBlue, Expedeon) to reveal protein bands.

2.3 Surface Plasmon Resonance

All SPR experiments were carried out with a BIAcore X100 and a CM5 chip from GE Healthcare by immobilizing glycoprotein fetuin. rAFL5 or fungal extracts was injected into the flow channels. To immobilize the glycoprotein on the CM5 chip, the surface was activated by washing with 10 mM buffered sodium acetate at pH 4.5 at a flow rate of $5 \mu l\ min^{-1}$ followed by injection with 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) and N-hydroxysuccinimide (NHS) at the same flow rate. Fetuin was diluted in the same buffer to a concentration of $50\ \mu g\ ml^{-1}$ and injected into flow channel 2 (fc2). Flow channel 1 (fc1) serves as a reference channel to compensate for unspecific binding to the surface. The response was monitored and injection was stopped when 3838.8 RU were reached. Both flow channels were then washed with ethanolamine to deactivate the residual NHS groups on the chip surface.

The chip was first tested with a known lectin from the mushroom *Psathyrella velutina* (PVL) which binds to NeuAc and GlcNAc residues, which are present on fetuin. Then, the chip was used to investigate AFL5 binding. In all experiments, PVL was used as a control.

2.4 Stability assessment (TSA)

To test the thermal stability of rAFL5, a range of different buffers with pH ranging from 4.5 to 8.5 were tested. Buffer stocks were prepared at 100 mM. To screen the conditions, a mixture was prepared containing 70 µl rAFL5 at $3.17\ mg\ ml^{-1}$, 7 µl 500x Sypro Orange (Sigma) and 63 µl H_2O. 5 µl of this mix were combined with 12.5 µl 100 mM buffer and 7.5 µl H_2O in each well. This leads to a final concentration of 50 mM buffer and $0.31\ mg\ ml^{-1}$ rAFL5. The thermal shift assay was then performed with Bio-Rad Real-time PCR system MJMini using the Bio-Rad CFX Manager software. The temperature was gradually raised from 20˚C to 100˚C. After the initial testing to determine the optimal pH, further additives were screened to test their influence of rAFL5 stability. Table 4 shows a list of tested additives, most of them derived from the MD1-11 additive kit from Molecular Dimensions Ltd. Tests were carried out in HEPES buffer at pH 7.0 and 7.5 respectively.

2.5 Labelling with Alexa Fluor 488 for glycan array

To identify glycan structures that are recognized by rAFL5, the lysine residues of the protein were labeled with the succinimidyl ester (NHS) of Alexa fluor 488 (invitrogen). Labeling was carried out according to the manufacturer`s protocol and labeled rAFL5 was sent to the Consortium for Functional Glycomics (CFG) in the USA.

2.6 Purity assessment by dynamic light scattering (DLS)

Experiments of dynamic light scattering were carried out using the Zetasizer from Malvern Instruments and a 40 µl quartz cuvette. The purified protein solution was centrifuged at high speed for 20 min to remove bigger particles like dust contaminants or aggregated proteins. The supernatant was then transferred into the clean quartz cuvette which was carefully placed into the machine. The Zetasizer software was used to setup the parameters and start the measurements.

2.7 Selenomethionine- containing rAFL5 for experimental phasing

For experimental phasing using multiwavelength anomalous dispersion (MAD), rAFL5 was expressed in minimal M9 medium supplied with100 mg D-L-selenomethionine per litre culture. Feedback inhibition of the methionine pathway in the expression strain *E. coli* BL21 (DE3) was achieved by addition of 100 mg lysine, phenylalanine and threonine and 50 mg isoleucin, leucine and valine per litre culture before induction with 0.1 mM IPTG. Purification and crystallisation was carried out as described above.

2.8 Crystallization

Crystallization was carried out by sitting drop vapor diffusion on 24 well plates at 20˚C. Initial crystallization screens were done with 10 and 15 mg ml^{-1} rAFL5, using Structure Screen 2 and Morpheus from Molecular Dimensions Ltd. When most of the drops were clear after one week incubation, the concentration was augmented to 20 mg ml^{-1} rAFL5 and Crystal Screen 1 from Hampton research was used as well. Large crystals grew in TRIS pH 8.5 containing 2.0 M ammonium sulfate which were collected and cryoprotected in 80% saturated lithium sulfate solution. In addition, needles were obtained, in TRIS buffer pH 8.5 containing 30 % PEG 4000, with either 0.2 M magnesium sulfate or 0.2 M lithium sulfate.

Single needles were collected and frozen in liquid nitrogen. Cryoprotection was given by the presence of 30% PEG 4000 in the mother liquor.

2.9 X-data collection and processing

X-ray data was collected at the beamline BM30A at the European Synchrotron Research Facility (Grenoble) with a ADSC Q4 CCD. Indexing was carried using Mosflm (Leslie & Powell 2007).

2.10 Identification of novel lectins by sequence similarities

Besides the direct purification attempts from fungal extracts, the available genome of As*pergillus fumigatus* (Nierman et al. 2005) was searched for sequence similarities with other known fungal lectins. The CERMAV lectin 3D database (http://lectin3d.cermav.cnrs.fr) was used to design the search. The lectin 3D database contains an overview of all available X-ray structures of lectins, and links to the PDB website to obtain the structures in pdb format (Bernstein et al. 1977). The amino acid sequences from all lectins summarized under the category "yeast and fungal" lectins were retrieved from NCBI and were used to search the *A. fumigatus* genome. For the bioinformatical search the publicly available Aspergillus Genome Database (AspGD) website (http://www.aspergillusgenome.org/cgi-bin/compute/blast_clade.pl) and the integrated Basic Local Alignment Search Tool (BLAST) was used. The search space was restricted to the *Aspergillus fumigatus* genomes from the stains Af293 and A11630 to obtain a highly sensitive search.

Hits were further analyzed when their bit score was higher than 30 and the E-value was lower than 0.01. Furthermore, annotations were taken into account: annotations as lectins or carbohydrate binding domains were considered highly significant whereas annotations in carbohydrate metabolism and those of unknown function were considered to be of low significance. Hits with bit scores lower than 30 and E-values and/or annotations other than carbohydrate binding domains were deemed insignificant.

Hits with bit scores between 30 and 40 with appropriate E values were then analyzed by sequence alignment using ClustalW available from ExPasy (http://www.ch.embnet.org/software/ClustalW.html) and NPS@ ([https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/ npsa_clustalw.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/)). Secondary structure prediction was carried out using APSSP website (http://imtech.res.in/raghava/apssp/) and assessment was carried out manually by comparing structure predictions with the known structures using the PyMOL Molecular Graphics

System, Schrödinger, LLC. Where necessary, the existence of tandem repeats was analyzed using RADAR, available from EMBL-EBI (http://www.ebi.ac.uk/Tools/pfa/radar/). The raw output of the BLAST search is shown in the supplemental table 8, and significant hits are summarized in table 7.

3. Results & Discussion

3.1 Search for lectins from fungal extracts

3.1.1 Purification with affinity chromatography

The presence of a sialic acid-specific lectin in *A. fumigatus* was previously described and the protein was partially characterized (Tronchin et al. 2002). Our aim was to reproduce the purification of this lectin and to identify its gene for recombinant expression and detailed characterization of its three dimensional structure and carbohydrate specificity. Since hemagglutination inhibition studies suggested an affinity for the glycoprotein fetuin, our first purification attempt was based on affinity chromatography on a fetuin coupled resin.

The initial hemagglutination activity of the crude fungal extract was measured to be 40 HAU ml^{-1}, which is 40 times less than the published activity, even though a 2 fold higher total protein concentration was used. Proteins were present in the flow-through of the fetuin column but no hemagglutination activity was found in the flow-through fractions. Stepwise elution with increasing salt concentration (up to 1.4 M NaCl) and subsequent 0.1 M sialic acid did not result in further protein elution as measured with Bradfort reagent (Bio-Rad) followed by visual analysis.

It was concluded that the putative lectin may have strongly bound to the column and could not be eluted with high salt or sialic acid or that it could be absent from the extract. It may be possible as well that the lectin did not bind to the resin and could not be detected in the flow through due to dilution (data not shown).

3.1.2 Purification by separation

In the following purification attempts, a protocol was setup using conventional separation methods including ammonium sulfate precipitation, ion exchange chromatography and hydrophobic interaction chromatography. An initial gel-filtration step which was proposed in the original protocol (Tronchin et al. 2002) was omitted because the crude extracts may contain pigments, oily components and other complex substances that may damage our column. A total of 200 mg of protein from the crude extract were analyzed by hemagglutination assay and showed 40 HAU ml^{-1}, i.e. 40 times less than the published results, with a 4 fold higher initial protein concentration (data not shown).

The crude fungal extract was directly subjected to ammonium sulfate precipitation and subsequent dialysis for ion exchange chromatography. This concentration step did not increase the hemagglutination activity (data not shown). Elution of ion exchange chromatography was carried out with a linear gradient of increasing NaCl concentration and the proteins eluted in one peak. Peak fractions were analyzed by 12% SDS PAGE and Coomassie based staining. They showed several protein species ranging in molecular weight from 10 kDa to 75 kDa with an apparent concentration of a protein with a molecular weight of ~ 30 kDa (figure 22), which may correspond to the putative lectin with expected molecular weight of 32 kDa. The fractions were pooled and dialyzed for hemagglutination testing. Hemagglutination activity however, did not increase and stayed at 40 HAU ml^{-1}.
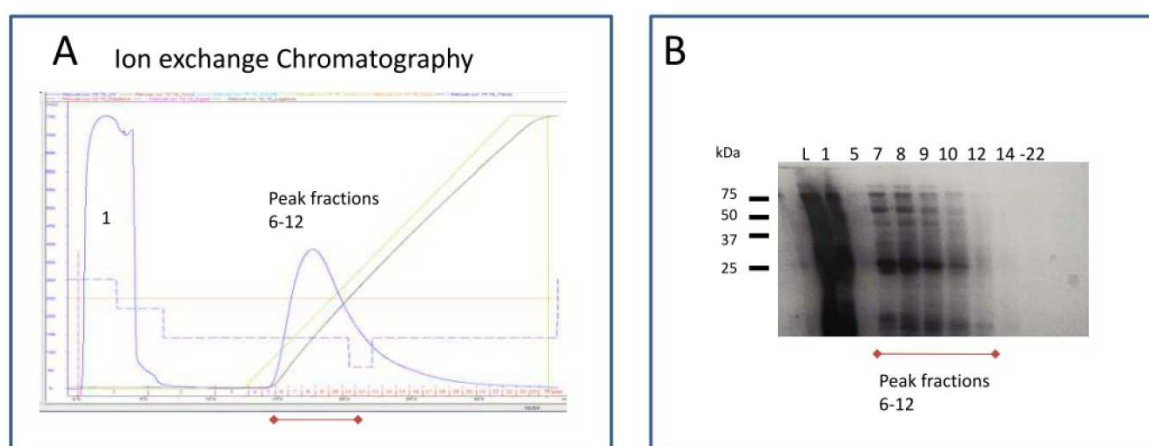
Figure 22: Ion- exchange chromatography of crude fungal extract. Chromatogram (A) and 12% SDS PAGE (B). 1 ml HiTrap QFF (GE Healthcare) column was equilibrated with 20 mM TRIS pH 8.5 before the crude fungal extract was loaded. Elution was carried out with a linear gradient of NaCl with a final concentration of 1M. 20 µl samples of each fraction were analyzed by 12 % SDS-PAGE.

Hydrophobic interaction chromatography was then carried out using a linear gradient of decreasing ammonium sulfate concentration (starting from 1.4 M) in accordance with the initially published purification protocol (Tronchin et al. 2002). Several peaks were eluted and 12 % SDS PAGE revealed several faint bands in most of the peak fractions with an apparent molecular weight higher than 50 kDa (blue circles, figure 23). In fraction 7 (red circle, figure 23) a protein band with an approximate molecular weight of 30 kDa was identified. Its apparent size corresponds to the one of the expected sialic acid lectin. This fraction was transferred to PVDF membrane (figure 24A) to identify the protein using *N*-terminal sequencing. Hemagglutination tests showed partial agglutination in the first well, corresponding to <40 HAU ml^{-1}. The exact hemagglutination titer could not be calculated because of incomplete agglutination. The testing could not be repeated due to small sample size.
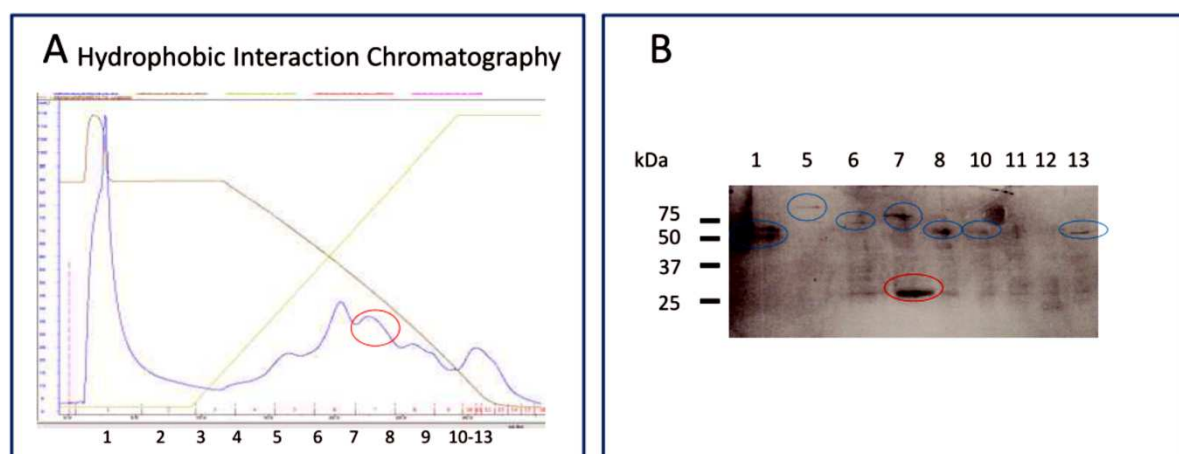
Figure 23: hydrophobic interaction chromatography. Chromatogram (A) and SDS PAGE (B). 1 ml Hi Trap Phenyl HP column (GE healthcare) was equilibrated with 20 mM TRIS buffer pH 7.5, 1.4 M ammonium sulphate. Elution was carried out with a decrease in ammonium sulfate. Fractions were analyzed by SDS PAGE and several proteins (blue circles) with apparent molecular weights higher that 50 kDa could be identified. Fraction 7 contained a protein with an apparent molecular weight of 30 kDa (red circle).

The *N*-terminal sequencing was carried out at the Institute of Structural Biology (IBS) in Grenoble where the first 15 residues could be identified (figure 24B). Search of this sequence in the translated genome of *A. fumigatus* resulted in identification of the gene *Afu8g0050*. The corresponding protein was annotated as putative methyltransferase involved in the hypoxia induced production of the secondary metabolite pseurotinA (figure 24C).
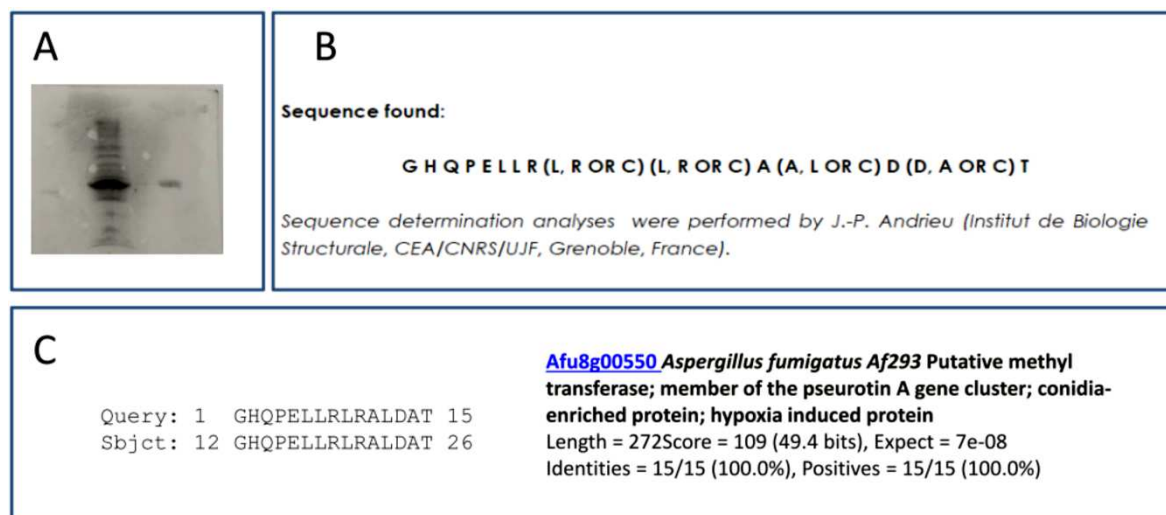
Figure 24: N-terminal sequencing result. Protein transfer to PVDF (A) and N-terminal sequencing (B) were carried out at the IBS, Grenoble. The result was screened by tBLASTn against the translated genome of *A. fumigatus*. The resulting gene was annotated as a putative methyltransferase (C)

.

As hemagglutination tests were not conclusive in the last purification step and the identified protein was annotated as a putative methyltransferase, the results needed confirmation and the purification was repeated using the same protocol. An overview of the purification is shown in figure 25. Interestingly, the hydrophobic interaction chromatography (HIC) shows a different elution profile (figure 25 B). While the hydrophobic interaction chromatography in the first purification showed 5 peaks (figure 23A), the second one only showed 3 distinct peaks in the elution. The peaks were analyzed by SDS PAGE (figure 25C lanes 4, 5 and 6). The first peak in lane 4 shows three protein bands with the apparent molecular weight of 24 kDa, 37 kDa and 39 kDa respectively. Peak II shows several bands with molecular weights ranging from 75 kDa to 25 kDa. The third peak shows a smeared band at ~ 32 kDa and several faint bands below 25 kDa.

Hemagglutination tests confirmed the presence of a lectin in this peak with 40 HAU ml^{-1}, and *N*-terminal sequencing of the smeared band at ~32 kDa revealed the first 8 amino acids (GHQPELLLRALDAT) with a low level of confidence due to low amounts of protein. However, a search for the short sequence in the *A. fumigatus* genome was inconclusive.
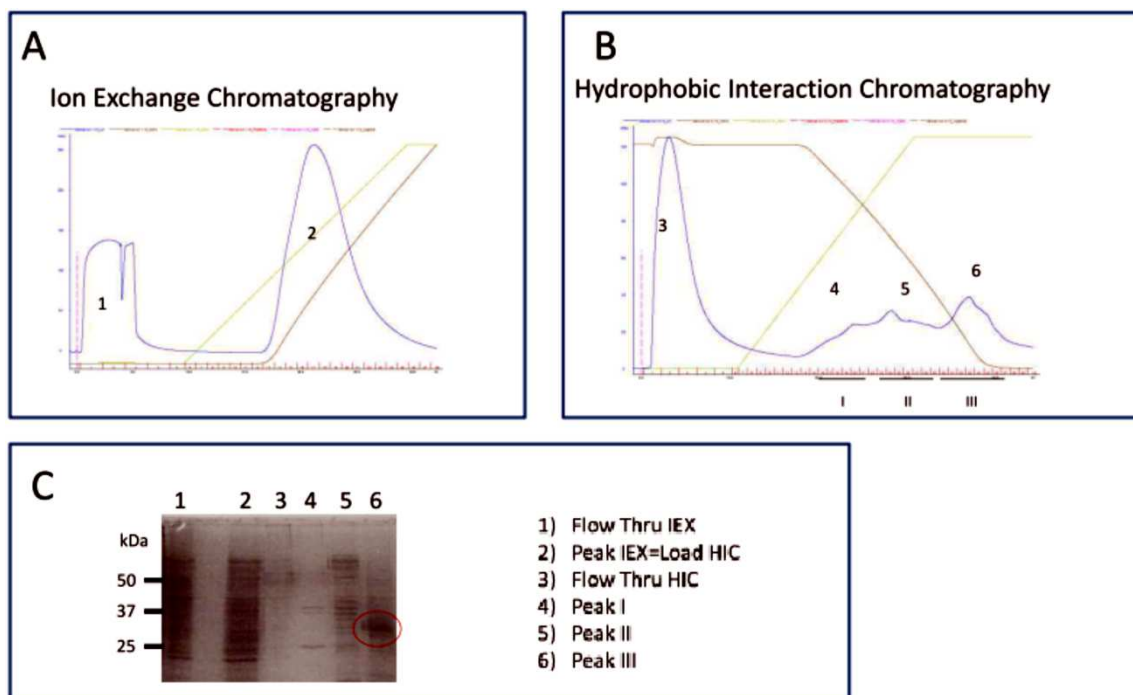
Figure 25: second attempt of purification of the putative lectin AFL5 from crude fungal extracts. Ion exchange chromatogram is shown in panel A and the chromatogram from hydrophobic interaction is shown in B. Fractions from both chromatographies were analyzed by 12% SDS PAGE (C). The putative lectin with an apparent size of ~30 kDa is circled in red.

Taken together these findings suggest the presence of a protein with an approximate size of 30 to 32 kDa and low hemagglutination activity in this fraction.

The low hemagglutination activity is not surprising, given an initial activity in the crude fungal extracts of 40 HAU ml^{-1}. This may be due to the transport of the frozen fungal extracts from Angers to Grenoble. Since limited amounts of fungal extracts were available further purification attempts could not be carried out at this point. Even though the hemagglutination activity was below 40 HAU ml^{-1} after hydrophobic interaction in the first purification, a protein of the approximately same size in the second purification showed hemagglutination activity. The identified gene *Afu8g0050* from the first purification was chosen for recombinant expression and further investigation of this protein towards possible carbohydrate binding activities.

3.2 Preliminary characterization of the putative lectin

3.2.1 Recombinant expression and purification

The gene *Afu8g0050* from *A. fumigatus* was cloned into an expression vector for recombinant expression of the corresponding protein, henceforth called rAFL5. By introducing a cleavable 6-histidines tag at the *N*-terminus of rAFL5 up to 20 mg ml $^{-1}$ pure protein could be purified per 1 liter of bacterial culture (figure 26).The molecular weight was calculated to be 29.5 kDa and the theoretical PI was calculated to be 5.21. The PI was verified to be between 5 and 6 by isoelectric focusing electrophoresis (data not shown). The extinction coefficient was calculated to be 23950 and was used to correctly estimate the protein concentration of rAFL5.
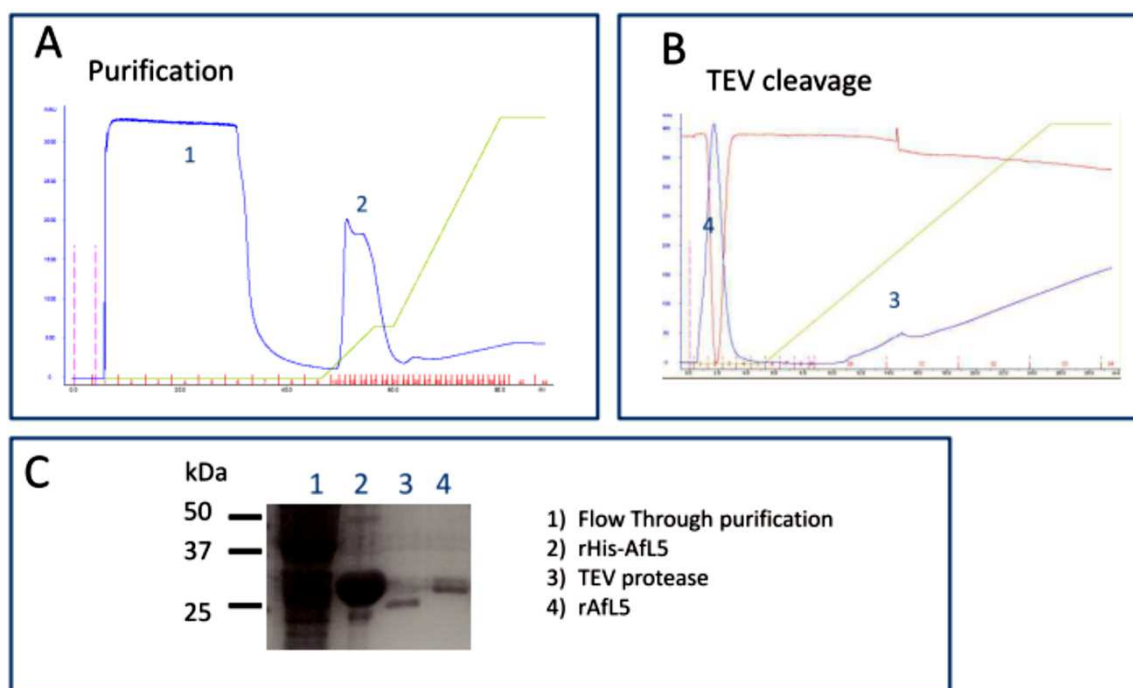
Figure 26: purification of recombinant his-tagged AFL5 via Nickel affinity purification. Chromatogram of crude extract (A) and further purification after TEV cleavage of the His-tag (B) are shown. Fractions were analyzed by 12% SDS PAGE (C).

3.2.2 Hemagglutination

Hemagglutination tests that were carried out with rAFL5 in TRIS Buffer pH 7.5 containing 100 mM NaCl were negative. Since the original purification and hemagglutination assays were carried out in PBS (Tronchin et al. 2002), a comparison was made of rAFL5 in TRIS buffer and PBS over time for several batches. Figure 27 shows the hemagglutination assay of 6 different batches with 1 being freshly prepared and 5 being the oldest, kept at 4˚C for several weeks. rAFL5 was divided into two samples, one staying in TRIS buffer. In the other samples, the buffer was exchanged to PBS (index a, figure 27) by concentration and dilution using membrane ultrafiltration (Vivaspin, Sartorius 10 kDa cutoff). All tests were carried out with a starting concentration of 1 mg ml^{-1} in the first well, serially two-fold diluted in 150 mM NaCl. A bacterial lectin from the *Burkholderia cenocepacia* complex BclB was used as a positive control and negative controls contained only buffer (PBS and TRIS) or 150 mM NaCl and red blood cells.
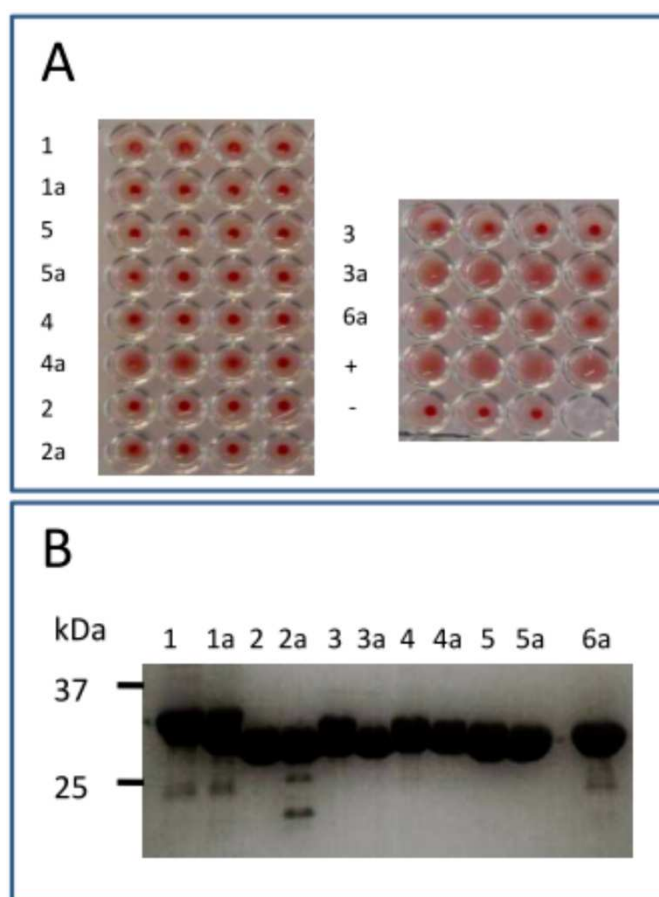
Figure 27: hemagglutination assays (A) of different purification batches of rAFL5. The index **a** indicates that the protein is kept in PBS instead of TRIS buffer. The samples were analyzed by 12% SDSPAGE (B) to evaluate their state of degradation.

In the first samples 1 (TRIS) and 1a (PBS), AFL5 contained the *N*-terminal 6-histidine tag and the sample shows some contamination as visible in the 12% SDS PAGE (figure 27 B). Both samples are unable to agglutinate red blood cells. The same result can be found for samples 2 and 2a where the 6-his-tag was cleaved off. The samples labelled 3 and 3a show that hemagglutination is present when rAFL5 is kept in PBS over time at 4˚C. However, samples 4 and 4a do not exhibit hemagglutination activity, indicating that prolonged incubation of rAFL5 is not a determinating factor. Samples 5a and 6a both agglutinate and show higher activity that sample 3a. The fact that none of the rAFL5 samples that were kept in TRIS buffer show hemagglutination activity may indicate that the amine group present in TRIS may have an inhibitory effect. Indeed, hemagglutination activity could be restored in sample 3 after exchanging the buffer to 20 mM HEPES containing 200 mM NaCl. However, it is noteworthy that hemagglutination activity, if present, remains low with 80 HAU ml^{-1}. Further testing with different salts such as magnesium chloride, calcium chloride and potassium chloride added to the protein solution did not restore hemagglutination activity. To investigate the possibility that a carbohydrate ligand bound to rAFL5 during expression, extensive dialyse against PBS was carried out before testing but this could not restore the activity.

3.2.3 Gel filtration

To further purify rAFL5 for crystallization studies and to investigate its state of oligomerization, gel filtration chromatography was carried out. Since hemagglutination studies suggested that TRIS buffer may have an effect on activity and PBS led to enhanced degradation, gel filtration was carried out in 20 mM HEPES buffer, containing 100 mM NaCl. Figure 28 A shows that rAFL5 elutes in a main peak at 65 ml corresponding to a molecular weight of 30 kDa (1). A shoulder (2) of the main peak was analyzed separately by hemagglutination testing (figure 28C). The chromatogram further shows that another small peak is present in the late eluate at 110 ml (3). All three samples were analyzed for hemagglutination activity and by SDS PAGE. Hemagglutination activity could only be found in that late peak (figure 28C). This active fraction may have bound to the sepharose matrix in the gel filtration column and its elution is therefore delayed. SDS PAGE (figure 28B) showed the same size band of rAFL5 in all peaks, thus excluding the presence of other contaminants.
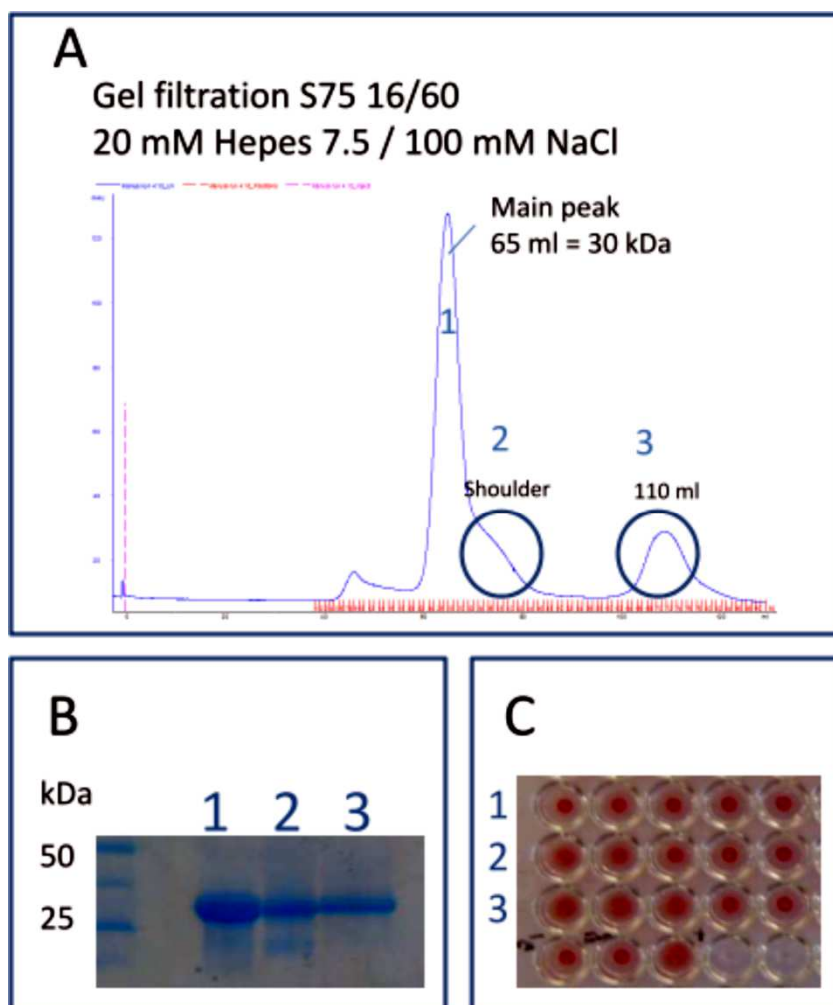
Figure 28 Gelfiltration chromatogram (A). rAFL5 elutes in the main peak at the expected sizes for a monomer. This main peak (1) has a shoulder (2) that was analyzed separately and a third peak (3) in the late eluate. All three peaks were analyzed by 12% SDS PAGE (B) and show a band at 30 kDa, expected for rAFL5. Hemagglutination testing (C) showed that only the late peak (3) is active.

It may be noted that the size of the late peak (3) varied with different purification batches, even though expression conditions and purification methods were not changed. Preliminary hemagglutination inhibition assays were carried out on samples from the late peak, while protein eluted in the main peak (1) was used for crystallization trials.

3.2.4 Hemagglutination inhibition

As mentioned above, inhibition tests were carried out with the active fraction which was separated by gel filtration. A range of monosaccharides were tested for their ability to inhibit the agglutination of rabbit red blood cells. Glucose, mannose, galactose and fucose did not inhibit the hemagglutination activity of rAFL5. GlcNAc and GalNAc inhibited hemagglutination only at the highest concentration of 200 mM which cannot be considered as specific. The glycoprotein fetuin was tested positive for hemagglutination inhibition (data not shown) to a concentration of 625 µg ml^{-1} a concentration similar to the previously published results for the putative lectin from *A. fumigatus* extracts (Tronchin et al. 2002).

3.2.5 Glycan array

rAFL5 was labeled with Alexa Fluor 488 for fluorescent detection on glycan array. The array was performed by the Consortium for Functional Glycomics (CFG), where the labeled protein was sent. Results of the glycan array however, were inconclusive (data not shown).

3.2.5 Stability assessment

Thermal stability of rAFL5 was tested with a thermal shift assay. In a first screen, the influence of the pH with two different salt concentrations was tested (table 4). rAFL5 was found to quickly degrade at low pH with a melting temperature (Tm) of 23°C independently of salt concentration. With increased pH, the protein got more stable with melting temperatures ranging from 43°C at pH 5.6 to 49°C at pH 7.0. The highest Tm of 50°C was found at pH 7.5 with 100 and 200 mM NaCl with a decrease in stability with higher pH. PBS however showed a slightly lower Tm of 47°C even though its pH and salt range were in the optimal range. Further screening was carried out with different additives like detergents, glycerol, different salts as well as sialic acid and fetuin to investigate their influence on thermal stability. Zinc and cobalt chloride were found to greatly destabilize rAFL5 resulting in melting temperatures of 30 and 25°C respectively. All other additives had a minimal impact on thermal stability but all melting temperatures were slightly lower than the control without additives.

Buffer system		addition (final concentration)	Tm (ºC)	addition (final concentration)	Tm (ºC)
Sodium acetate	pH 4.5	+100 mMNaCl	23	+200 mMNaCl	23
Sodium citrate	pH 5.6	+100 mMNaCl	43	+200 mMNaCl	44
Mes	pH 6.0	+100 mMNaCl	43	+200 mMNaCl	43
NaCacodylate	pH 6.5	+100 mMNaCl	46	+200 mMNaCl	46
HEPES	pH 7.0	+100 mMNaCl	49	+200 mMNaCl	48
Phosphate (PBS) pH 7.4			48		
HEPES	pH 7.5	+100 mMNaCl	49	+200 mMNaCl	48
TRIS/HCl	pH 8.0	+100 mMNaCl	48	+200 mMNaCl	48
TRIS/HCl	pH 8.5	+100 mMNaCl	48	+200 mMNaCl	48

Additive	Final concentration	Tm (ºC)
Potassium chloride	50 mM and 25 mM	48
Calcium chloride	10 µM	49
TECEP	25 mM	49
Na thiocyanate	10 mM	49
Taurine	10 mM	49
β-octylglucoside	0.5 %	47
Potassium iodide	100 mM	49
Cobalt (II) chloride	10 mM	25
Magnesium chloride	10 mM	48
Cadmium chloride	10 mM	41
Zinc chloride	10 mM	30
Glycerol	2 %	49
Glycerol	10%	49
Sialic acid (Neu5Ac)	5 µM	48
Fetuin	0.1 mg ml^{-1}	49

Table 4: overview of various conditions and additives that were tested with TSA

3.2.6 Dynamic light scattering

To assess the purity of rAFL5 after gel-filtration, dynamic light scattering experiments were conducted. The results showed a polydispersity index of 0.135 indicating that the solution of rAFL5 is monodisperse and therefore suitable for crystallization trials (data not shown).

3.2.7 Crystallization

The pure and monodisperse protein solution was kept in 20 mM HEPES buffer pH 7.5 with 100 mM NaCl at a concentration of 20 mg ml^{-1}. Large crystals were obtained in TRIS pH 8.5 containing 2.0 M ammonium sulfate that diffracted to 4 Å resolution. Furthermore, thin needles were obtained, also in TRIS buffer pH 8.5 containing 30 % PEG 4000, with either 0.2 M magnesium sulfate or 0.2 M lithium sulfate. Further screening was then carried out varying salt and PEG concentrations. However, these conditions did not result in further improvement of crystal growth but needles grew in TRIS buffer pH 8.5 with 2.0 mM ammonium sulfate. These needles were used for data collection at the European Synchrotron Radiation Facility (ESFR) and diffracted to 2.1 Å resolution at beamline BM30A.

Crystals of rAFL5 are tetragonal and belong to the P4 space group with unit cell dimensions $a= b=$ 46.26 Å and $c=$ 116.1 Å. Statistics of the data collection are summarized in table 2.
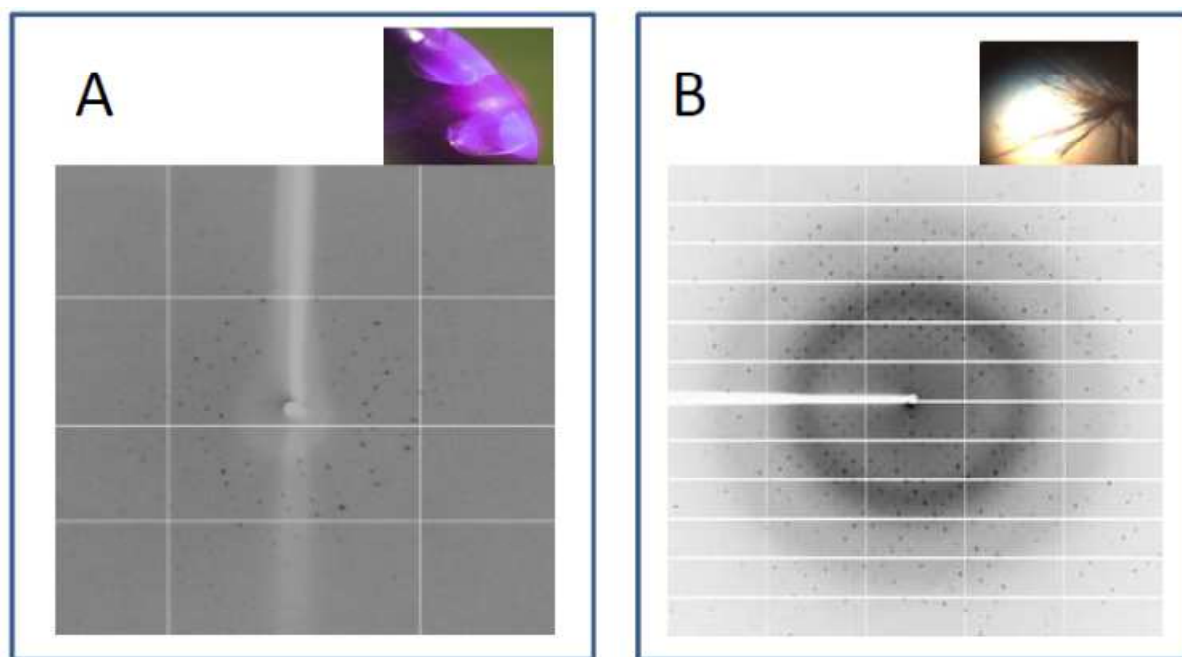
Figure 29: rAFL5 crystals and their diffration patterns. Crystals were obtained with 20 mg ml^{-1} rAFL5 in 100 mM TRIS pH 8.5 containing 2 M ammonium sulfate. Large crystals diffracted to 4 Å (A), while thin needles diffracted to 2.1 Å (B).

Unit Cell: 46.26 46.26 116.1 90 90 90 (Tetragonal)		
Space group: to be determined		
Resolution	36.18- 2.1	2.16-2.10
R merge (within I+/I-)	0.064	0.360
total number of observations	64825	5119
Total number unique	14155	1140
Mean((I)/sd(I))	17.1	4.0
Mn (I) half set correlations CC(1/2)	0.998	0.901
Completeness	99.6	99.7
Multiplicity	4.6	4.5

Table 5: crystallographic statistics

Single-wavelength anomalous diffraction (SAD) phasing attempts with sulfur in the native rAFL5 were unsuccessful, because the anomalous signal of sulfur was not strong enough. Phasing was then aimed to be done with multi-wavelength anomalous diffraction (MAD) with the use of selenomethionyl-rAFL5. The selenomethionyl-rAFL5 was successfully expressed and purified. Thin needles grew in hanging drops containing 1µl 20 mM TRIS pH 8.5, 2 M ammonium sulphate and 1µl selenomethionine-rAFL5 at 20 mg ml^{-1}. However, the thin needles did not diffract due to their small size. Optimization trials were carried out by seeding mircocrystals into new drops with the above mentioned crystallization conditions but sufficiently large crystals could not be obtained.

3.2.8 Surface Plasmon Resonance

The carbohydrate binding properties of rAFL5 remained questionable after hemagglutination testing. Therefore the more sensitive Surface Plasmon Resonance (SPR) was used for further investigation. Since inhibition studies showed an apparent specificity of rAFL5 for the glycoprotein fetuin a chip was created where fetuin was immobilized as ligand on the chip surface. The chip was tested with a known lectin from the mushroom *Psathyrella velutina* (PVL), which binds to the glycan structures present on the glycoprotein fetuin. Figure 30 shows the sensorgram where the binding of PVL is shown in green, indicating that the fetuin was successfully immobilized on the chip surface. The injection of rAFL5 did not result in a change in response suggesting that rAFL5 does not bind to immobilized fetuin.
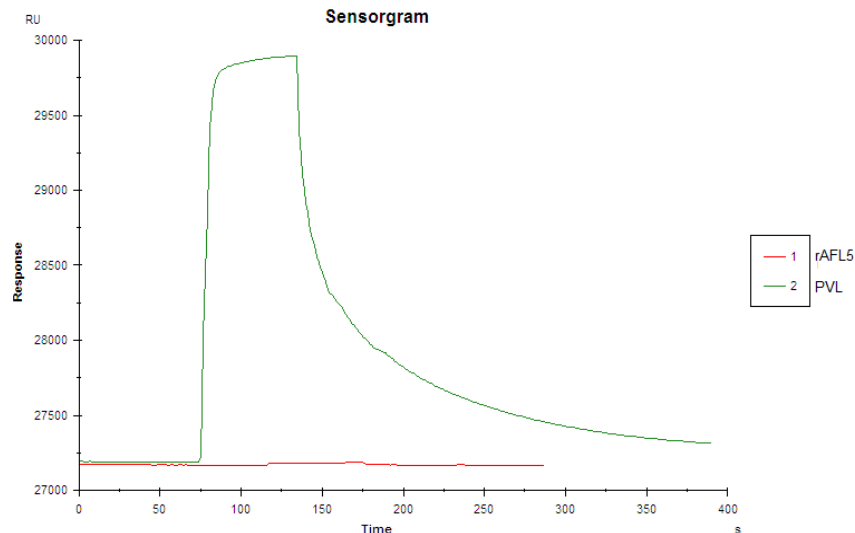
Figure 30: Surface Plasmon Resonance with rAFL5 (red) and PVL (green) using a fetiun coupled CM5 chip.

Given the fact that rAFL5 showed inconclusive hemagglutination results and that no interaction could be detected by SPR we concluded that the identified protein corresponding to the gene *Afu8g0050* from *A. fumigatus* was not identical to the sialic-acid specific lectin described earlier (Tronchin et al. 2002). We therefore decided to start another purification attempt with new fungal extracts. Since the lectin was described to bind to fetuin in hemagglutination inhibition we decided to monitor the purification steps with SPR using the fetuin chip, thus providing a more sensitive and unbiased detection method.

We received two new batches of fungal extracts that were both analyzed by SPR. No binding was found in the crude extracts. The extracts were then concentrated by ammonium sulfate precipitation and dialyzed for SPR experiments but again, no binding could be detected. Taken together these findings suggest that no active fetuin binding lectin was present in any provided extracts. The freezing and thawing during the travel from Angers to Grenoble may have been a contributing factor why it was not possible to purify the sialic acid specific lectin from crude fungal extracts.

3.2.9 Sequence alignment studies

Sequence alignment studies with AFL5 using protein BLAST (Altschul & Madden 1997) showed that the protein has high sequence similarity with methyltransferases from other filamentous ascomycetes such as *Scedosporium apiospermum*, *Metharhizium anisopliae* (Figure 31), *Fusarium oxysporum*, *Talaromyces stipitatus* and others. However, it may be worth mentioning that most of them are putative or hypothetical proteins identified by sequence similarity and none of these proteins has been studies in further detail. In the case of Afu8g00550 (rAFL5) its property is deduced from sequence similarity and gene clustering (Maiya et al. 2007).
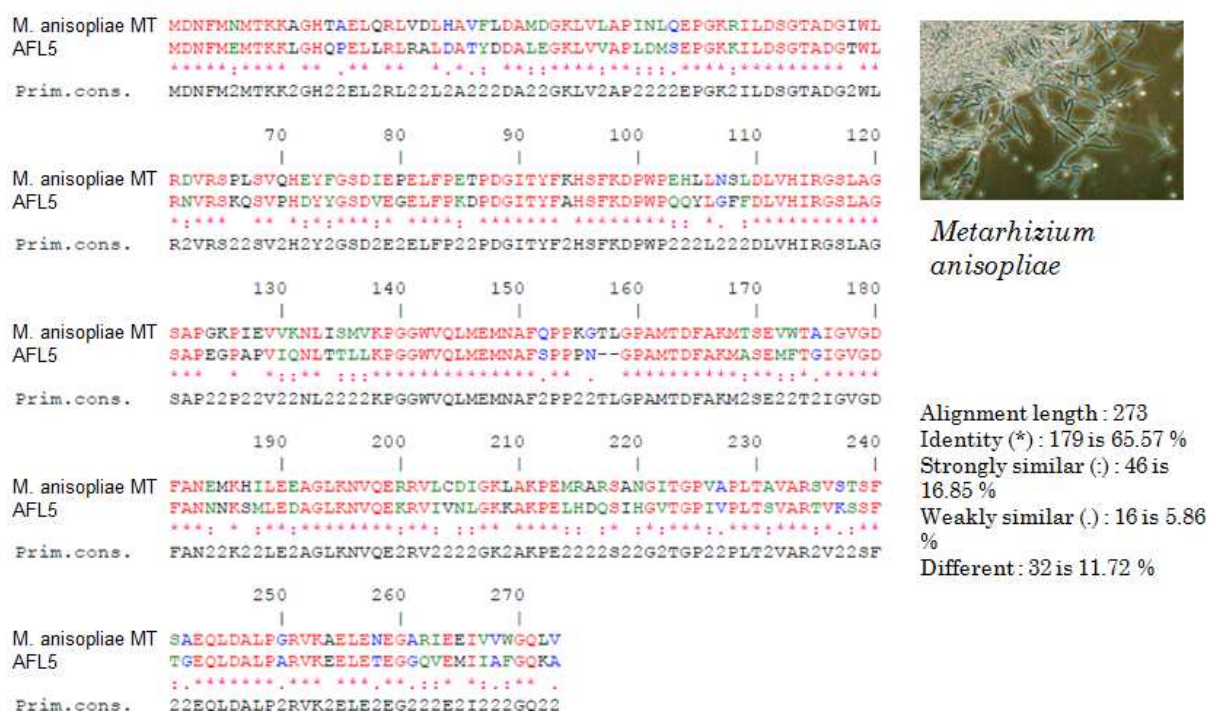
Figure 31: sequence alignment of AFL5 with a SirN- like Methyltransferase (MT) from the Ascomycota *M. anisopliae*. Alignment was carried out with ClustalW. Identical residues are shown in red, similar residues in green and blue.

3.3 Search for novel lectins by sequence similarity

Sequence similarity occurs between homologous proteins and can therefore be exploited to search for proteins of interest, in this case small soluble lectins in translated genomes.

A bioinformatical search using conserved domain databases of the *A. fumigatus* genome was carried out by collaboration partners in the Czech Republic beforehand and resulted in the identification of 9 potential lectins, summarized in table 6.

Name/ ORF location	Similarity with known lectins
AFL 1/ Afu5g14740	Fucose specific fungal lectin 6-blades β-propeller
AFL2 / Afu7g00560	Fungal lectin Actinoporin-like
AFL3/ Afu5g00910	C-type lectin
Afu2g12590	C-type lectin
AFL4/ Afu2g01405	Cyanovirin homolog
pAFL11/ Afu2g16800	L-type lectin
pAFL10/ Afu2g12180	L-type lectin
Afu4g128050	Calnexin
Afu5g13830	R-type (and galactosidase domain)

Table 6: previously identified lectins in *A fumigatus*.

In this thesis the search was expanded by searching the translated genome for putative proteins that share sequence similarity with other known fungal lectins summarized in the CERMAV Lectin 3D database. A total of 27 lectin sequences from 20 fungal species were used for screening the *A. fumigatus* 1136 and *A. fumigatus* af293 translated genome. The query lectins were categorized in 9 groups based on their structural fold as presented on the Lectin 3D database. Using the BLAST tool on the comprehensive AspGD website allowed a highly sensitive bioinformatical search by restricting the search space to only two genome sequences. The results are therefore vast and insignificant results had to be sorted out manually by evaluating the bit score and E- value as well as the alignment data. Most of the identified hits resulted in patchy alignments shorter than 10 amino acids and had a bit score lower than 30 and high Expect (E) values ranging from 9 to 1 indicating insignificant sequence similarities due to the high sensitivity of the search. The raw data of the BLAST Hits are shown in the supplemental table 8 in the annex at the end of this chapter. Furthermore, the results show redundancy because of similarity in the query sequences. For a better overview of the identified ORFs encoding a putative lectin, names were assigned to the ORF. pAFL stands for putative *A. fumigatus* lectin followed by consecutive numbering. The results are shown in table 7.

Name/ accession number	MW (kDa)	Similarity type	Similarity species (Uniprot ID)	bit Score/ E-value	Orthologs
AFL 1/ Afu5g14740 AFUB_062410	34.7	6-blades β-propeller	*Aleuria. aurantia (P18891)*	125/ 2 E-29	*A. flavus/ A. oryzae A. aculeatus*
AFL2 / AFUB_087150 Afu7g00560	14.0	Actinoporin,	*Agaricus. bisporus (Q00022) Boletus. edulis (F2Z266) Sclerotium rolfsii (D2D056) Xerocomellus. chrysenteron (Q8WZC9)*	67/ 2 E-6	*A. fumigatus*
AFL4/ AFUB_018490 Afu2g01405	12.1	Cyanovirin homologs	*Tuber borchii (Q5MK11) Neurospora crassa (Q7S6U4) Gibberella zeae (F2Z242) Magnaporthe. oryzae (G4N991)*	84/ 7 E-18	*A. clavatus/ N. fischeri*
AFL6/ Afu1g02290 AFUB_002680	11.8	Cyanovirin homologs	*Tuber borchii (Q5MK11) Neurospora crassa (Q7S6U4) Gibberella zeae (F2Z242) Magnaporthe. oryzae (G4N991)*	109/ 2 E -25	*A. nidulans A. versicolour A. clavatus N. fischeri A. sydowii/A. flavus A. oryzae/A. terreus A. niger A. acidus A. aculeatus A. carbonarius A. brasiliensis*
pAFL7/ AFUB_095010	12,8	Cyanovirin homologs	*Tuber borchii (Q5MK11) Neurospora crassa (Q7S6U4) Gibberella zeae (F2Z242)*	56/ 2 E-10	Only in *A.fumigatus* 1163
pAFL8/ Afu7g04870/ AFUB_090420	36.6	Cyanovirin homologs	*Tuber borchii (Q5MK11) Neurospora crassa (Q7S6U4) Gibberella zeae (F2Z242) Magnaporthe. oryzae (G4N991)*	58/ 8 E-10	*A. nidulans/ A. oryzae/ A. niger*
pAFL10/ AFUB_027870/ Afu2g12180	36.2	Legume like, Membrane protein with signal peptide	*Coprinopsis cinerea (A8N4Z2)*	243/ 1 E-64	*A. nidulans/ A. versicolor/ A. sydowii/ A. flavus/ A. oryzae/ A. terreus/ A. niger/ A. acidus/ A. aculeatus/ A. carbonarius/ A. brasiliensis/ A. clavatus/ N. fischeri*
pAFL11/ Afu2g16800 (AFUB_032470)	45.2	Legume like, Membrane protein with signal peptide	*Saccharomyces cerevisiae (E7KFG8)*	51/ 3 E-7	*A. oryzae/ A. nidulans/ A. niger*
pAFL12 Afu5g01120/ AFUB_049600	36.6	Ricin B like	*Coprinopsis. cinerea (A8NCX0)*	495/ 1 E-140	*A. nidulans/ A. sydowii/ A. flavus/ oryzae/ A. terreus/ N. fischeri/*

Table 7:Summary of putative lectins in *A. fumigatus,* identified by sequence similarity with known fungal lectins which have been structurally characterized.

3.3.1 β-propeller (AFL1)

The representative of a 6-blade β-propeller was AAL the lectin from *Aleuria aurantia*. Expectedly AFL was found with a high bit score of 125, validated by a low E- value of 2 E-29. In addition, a second protein (AFUB_075950/ pAFL9) showed similarity with a bit score of 31 (E-value: 0.47), just above the cutoff and no conserved domains were annotated in this protein. The alignment with AAL (36 kDa) and AFL (34.7 kDa) was inconsistent, which may be due to the smaller size of AFUB_075950/ pAFL9 (11.5 kDa). In contrast, alignment with the bacterial lectin BamBL (9.4 kDa) of the same structural family from *Burkholderia ambifaria* showed 48% similarity. Further cluster alignment with AAL, AFL and BamBL showed that conserved residues involved in fucose binding were not conserved in AFUB_075950/ pAFL9. In addition, secondary structure prediction using Phyre and search for tandem repeats using RADAR could not bring evidence that the putative protein AFUB_075950 belonged to the lectin family of β- propeller (data not shown). It was therefore concluded, that AFUB_075950/ pAFL9 cannot be considered a lectin. Moreover this investigation demonstrates that the chosen cutoff of a bit score of 30 ensures a high sensitivity while securely cutting out insignificant hits. In fact, insignificant hits were found with bit scores around 30 using *Psathyrella velutina* lectin with 7 blade β-propeller fold. The 3 hits were annotated as esterase, demethylase and oxidoreductase respectively and therefore immediately identified as insignificant.

3.3.2 Actinoporin like (AFL2)

The significant hit (Afu7g00560) with the lowest bit score was a member of the actinoporin-like lectin family confirming the previous search.It was identified in all query lectins from *Agaricus bisporus, Boletus edulis, Sclerotium rolfsii* and *Xerocomellus chrysenteron* with bit scores ranging from 47 to 68.

In addition, the lectin from *Agaricus bisporus* showed similarity with a putative glycosyl hydrolase and a putative β- galactosidase in *A. fumigatus*. These hits were designated with a low significance (table 8) as they are not lectins but involved in carbohydrate metabolism.

3.3.3 Cyanovirin- homolog (AFL4, AFL6, pAFL7, and pAFL8)

Among the putative lectins, three new cyanovirin-N homologs (CVNH) were identified from this genomic analysis. Cyanovirin-N (CV-N) was first identified from the cyanobacterium *Nostoc ellipsosporum* for its high antiviral activity (Boyd et al. 1997). Since then several homologs have been identified in fungi and seedless plants (Koharudin et al. 2008). They have been categorized in three types according to their domain structure. AFL4 and AFL6 belong to the type I CVNH, which contain either a single or multiple CVNH domains but belong to different subgroups. AFL4 had already been identified previously and its characterization is underway in our laboratory. AFL6 was previously identified (Koharudin et al. 2008) and its characterization was carried out in this thesis (see Chapter IV). pAFL7 with the accession number AFUB_095010 is only present in the 1163 strain of *A. fumigatus* and belongs to the type I CVNH as well. pAFL8 (AFUB_090420)is a type II CVNH and thus contains an *N*-terminal domain of different function and a *C*-terminal CVNH domain. Type II CVNHs have not yet been characterized. The N- terminal domain of pALF8 shows similarities with an ATPase that is part of the endosomal sorting complex required for transport (ESCRT) (Henne et al. 2013).

3.3.4 L-type lectins lectins (pAFL10 pAFL11)

The legume- like lectins pAFL10 (AFUB_027870/ Afu2g12180) and pAFL11 (Afu2g16800/ AFUB_032470) had been already identified in the previous bioinformatical search, carried out by our collaboration partners. Both lectins are part of highly conserved eukaryotic intracellular cargo transport protein family. Their carbohydrate recognition domain which has the same fold as L-type lectins binds high-mannose type oligosaccharides on N-glycosylated proteins for transport from the endoplasmatic reticulum to the Golgi. Their carbohydrate specificity may differ greatly while amino acid and structure are highly conserved.

3.3.3 β- trefoil/ R-type (pAFL12)

pAFL12 (AFUB_049600) was found using a putative ricin-B lectin (R-type) from the inky cap fungus *Coprinopsis. cinerea*. It is further associated with the pectin lyase fold superfamily and contains three parallel beta-helix repeats indicative of glycosidase function. It contains a signal peptide and is a secreted protein.

Examples for low-significance hits in this family are a member of the pectin methyltransferase family (AFUB_080500), a glycosidase involved in the cellulose catabolic

process (AFUB_077630) and a putative exo-beta-1,3-glucanase (AFUB_001500), that show low similarity with β- trefoil lectins.

4. Concluding remarks

In an attempt to identify novel lectins from the opportunistic pathogen *A. fumigatus,* we used two different approaches, namely purification from fungal extracts and identification by sequence similarity. In the first approach, we attempted to purify a presumed sialic acid specific lectin from crude fungal extracts. The extracts were prepared in Angers and sent to our facilities in Grenoble. During this transport the accompanying freezing/ thawing process, lectin activity may have decreased as we observed in the first sample a 40 times reduction in hemagglutination compared to what was described in the literature (Tronchin et al. 2002). In the second sample batch, a complete lack of activity was observed. With this low activity in starting material the successful purification of the lectin was greatly hindered. Nevertheless, we were able to identify and partially characterize a protein that showed hemagglutination activity with a low titer. Its carbohydrate binding ability however could not be conclusively proven and the project was abandoned when no interaction was found using the highly sensitive SPR- technique. Further research should be initiated with fresh fungal extracts where freezing should be avoided. To investigate whether AFL5 (Afu8g00550) has indeed methyltransferase activity, assays could be carried out with its natural ligands pseurotinA and its derivatives, which are commercially available.

The second approach was made possible thanks to the availability of the sequenced genome of *A. fumigatus* which allowed searching for sequence similarities with other known fungal lectins. Although lectins had been already identified with this method we were able to further present four novel putative lectins. While more detailed bioinformatical research is needed for most of the identified lectins, one cyanovirin-N homolog was already identified in a previous study (Koharudin et al. 2008). This lectin is henceforth called AFL6 and its characterization can be found in the following chapter.

5 Supplemental table

Raw data of the bioinformatical search for putative lectins in the *A .fumigatus* genome

Structural fold	Query Species	UniProtKB/ Swiss-Prot Query	Similarity in *A. fumigatus* Af293 (A11630)	Bit score	E-value	Name/ Annotation	Significance
6-blades β-propeller	*Aleuria aurantia*	P18891	Afu5g14740 (AFUB_0602410)	125	2 E-29	AFL1 (FleA)/ L-fucose specific lectin	high
			Afu6g09890 (AFUB_075950)	31	0.47	hypothetical protein/ uncharacterized ORF	low
7-blades β-propeller	*Psathyrella velutina*	Q309D1	Afu1g03170 (AFUB_003580)	33	0.16	Putative esterase/ uncharacterized ORF	none
			Afu5g03430 (AFUB_051950)	31	0.83	Histon demethylase activity/ uncharacterized ORF	none
			Afu7g06420 (AFUB_091980)	28	5.1	Putative NADH- dependent flavin oxidoreductase/ verified ORF	none
Actinoporin-like	*Agaricus bisporus*	Q00022	Afu7g00560 (AFUB_087150)	47	2 E-6	AFL2/ Characterization underway	high
			Afu3g01660 (AFUB_046730)	27	2.5	Putative glycosyl hydrolase/ uncharacterized ORF	low
			Afu5g14550 (AFUB_062230)	26	5.2	Putative β- galactosidase/ uncharacterized ORF	low
			Afu6g08570 (AFUB_074530)	26	5.9	Putative DNA damage repair protein/ uncharacterized ORF	none

		Afu1g02820 (AFUB_003200)	26	6.1	Pst2/ Putative NADH-quinone oxidoreductase/ verified ORF	none
Boletus edulis	F2Z266	Afu7g00560 (AFUB_087150)	57	3.9 E-9	AFL2/ Characterization underway	high
		Afu2g09960 (AFUB_025800)	30	2.1 E-1	Ssc70/Putative mitochondrial Hsp70 chaperone/verified ORF	none
		Afu3g14500 (AFUB_034730)	30	2.2 E-1	Predicted catalytic activity and role in metabolic process (Isochorismatase family)/ uncharacterized ORF	none
		Afu8g07110 (AFUB_080690)	28	8.8 E-1	Putative alanyl-tRNA synthetase/ uncharacterized ORF	none
		Afu2g12700 (AFUB_028340)	26	4.9	Predicted zinc ion binding (HET domain)/ uncharacterized ORF	none
		Afu2g08650 (AFUB_024570)	26	5.9	Predicted zinc ion binding (RING finger membrane protein)/ uncharacterized ORF	none
Sclerotium rolfsii	D2D056	Afu7g00560 (AFUB_087150)	67	2.0 E-12	AFL2/ Characterization underway	high
		Afu2g08650 (AFUB_024570)	29	7.4 E-1	Predicted zinc ion binding (RING finger membrane protein)/ uncharacterized ORF	none
		Afu6g04390 (AFUB_093900)	28	1.6	Putative ubiquitinationc protein (Bre1)/ uncharacterized ORF	none
		Afu5g04170 (AFUB_052690)	25	8.2	Hsp90 molecular chaperone Mod-E/ verified ORF	none
		Afu6g03260 (AFUB_095040)	25	7.9	Ap1(pepAb) putative aspartic endopeptidase/ uncharacterized ORF	none
Xerocomellus chrysenteron	Q8WZC9	Afu7g00560 (AFUB_087150)	58	1.0 E-9	AFL2/ Characterization underway	high
		Afu2g08650	31	1.7 E-1	Predicted zinc ion binding (RING finger	none

			(AFUB_024570)			membrane protein)/ uncharacterized ORF
			Afu8g07110 (AFUB_080690)	28	1.6	Putative Alanyl-tRNA synthetase/ uncharacterized ORF	none
β-trefoil domain	*Clitocybe nebularis*	B2ZRS9	Afu3g10060 (AFUB_039120)	27	2.2	Putative protein/ predicted Golgi transport complex/ Uncharacterized ORF	none
	Coprinopsis cinerea (CCL2)	B3GA02	Afu5g08190 (AFUB_055720)	27	1.7	Putative protein of unknown function/ Uncharacterized ORF	none
			Afu8g05740 (AFUB_081810)	26	3.1	Putative RTA1 protein/ predicted hydrolase activity/ uncharacterized ORF	none
			Afu8g07250 (AFUB_080500)	26	5.7	Pectin methylesterase family protein/ uncharacterized ORF	low
			Afu3g09500 (AFUB_039670)	26	6.1	Putative aryl-alcohol oxidase; vanillyl-alcohol oxidase/ uncharacterized ORF	none
			Afu2g06150 (AFUB_023230)	25	7.7	pdiA/ protein disulfide isomerase/ verified ORF	none
	Coprinopsis cinerea (ricin B)	A8NCX0	Afu5g01120 (AFUB_049600)	499	1 E-141	Putative protein of unknown function/ Uncharacterized ORF	high
	Coprinopsis cinerea (CCL1)	E3VW05	Afu6g11610 (AFUB_077630)	31	1.5 E-1	cbhB/ Probable 1,4-beta-D-glucan cellobiohydrolase B/ uncharacterized ORF	low
			Afu6g00370 (AFUB_097150)	28	1.5	Putative aminotransferase/ uncharacterized ORF	none
			AFUB_044190	26	4.7	Putative protein of unknown function/ Uncharacterized ORF	none
	Laetiporus sulpherus	Q7Z8V1	Afu8g03900 (AFUB_083680)	30	1.2	Putative RING finger domain/ uncharacterized ORF	none

		Afu6g09620 (AFUB_075670)	29	2.8	Predicted acyl- transferase activity/ uncharacterized ORF	none
		Afu6g13270 (AFUB_001500)	27	9.9	Putative Exo-beta-1,3-glucanase/ uncharacterized ORF	low
		Afu6g07300 (AFUB_073270)	28	5.3	Putative mRNA splicing factor (Prp17)/ uncharacterized ORF	none
Marasmius oreades (agglutinin)	Q8X123	n.a	n.a	n.a	n.a	n.a
Marasmius oreades (mannose lectin)	I7H471	Afu5g09190 (AFUB_056730)	29	3.4	Putative ABC bile acid transporter/ verified ORF	none
		Afu2g13260 (AFUB_028890)	26	2.5	medA transcriptional regulator/ verified ORF	none
		Afu5g00230	26	3.0	Unknown protein of unknown function/ uncharacterized ORF	none
		Afu5g12870 (AFUB_060580)	25	4.9	Putative NAD+ kinase/ uncharacterized ORF	none
		Afu1g08860 (AFUB_008280)	25	9.1	Putative Meiotic recombination protein Ski8/Rec14/ uncharacterized ORF	none
Polyporus. squamosus	Q75WT9	Afu6g02000 (AFUB_096340)	29	2.5	NCS1 allantoate transporter/\uncharacterized ORF	none
Sclerotinia sclerotiorum	A7XUK7	Afu5g09190 (AFUB_056730)	29	3.4	Putative ABC bile acid transporter/ verified ORF	none
		Afu2g13260 (AFUB_028890)	26	2.5	medA/transcriptional regulator/ verified ORF	none
		Afu5g00230	26	3.0	Unknown protein of unknown function/ uncharacterized ORF	none
		Afu5g12870 (AFUB_060580)	25	4.7	Putative NAD+ kinase/ uncharacterized ORF	none
		Afu1g08860	25	9.1	Putative Meiotic recombination protein	none

			(AFUB_008280)			Ski8/Rec14/ uncharacterized ORF	
Cyanovirin-N homolog	*Tuber borchii*	Q5MK11	Afu1g02290 (AFUB_002680)	66	3.0 E-12	AFL6/ characterization underway	high
			AFUB_095010	47	2.0 E-06	Putative protein of unknown function/ uncharacterized ORF	high
			Afu2g01405 (AFUB_018490)	43	2.0 E-05	AFL4/ characterization underway	high
			Afu7g04870 (AFUB_090420)	42	6.0 E-05	Putative glutamine-serine-proline rich protein/ uncharacterized ORF	high
			Afu6g01970 (AFUB_096370)	27	1.4	GATA transcriptional activator AreA/ verified ORF	none
			Afu3g10550 (AFUB_038610)	25	4.4	Putative protein of unknown function/ uncharacterized ORF	none
			Afu3g10820 (AFUB_038310)	25	4.3	Putative Ferric-chelate reductase/ uncharacterized ORF	none
	Neurospora crassa	Q7S6U4	Afu1g02290 (AFUB_002680)	109	2.5 E-25	AFL6/ characterization underway	high
			Afu2g01405 (AFUB_018490)	69	4.0 E-13	AFL4/ characterization underway	high
			AFUB_095010	56	2.0 E-09	Putative protein of unknown function/ uncharacterized ORF	high
			Afu7g04870 (AFUB_090420)	41	8.0 E-05	Putative glutamine-serine-proline rich protein/ uncharacterized ORF	high
			Afu4g03090 (AFUB_100040)	28	4.9 E-01	Putative protein of unknown function/ uncharacterized ORF	none
			Afu4g04660 AFUB_098340	27	1.1	Putative proteasome regulatory particle subunit Rpt6/ uncharacterized ORF	none
			Afu8g06360	26	2.9	Putative alpha-1,3-glucanase	low
	Gibberella zeae	F2Z242	Afu1g02290	92	3.0 E-20	AFL6/ characterization underway	high

			(AFUB_002680)				
			Afu2g01405 (AFUB_018490)	84	7.0 E-18	AFL4/ characterization underway	high
			Afu7g04870 (AFUB_090420)	58	8.0 E-10	Putative glutamine-serine-proline rich protein/ uncharacterized ORF	high
			AFUB_095010	56	2.0 E-09	Putative protein of unknown function/ uncharacterized ORF	high
			Afu1g1181 (AFUB_011260)	26	2.2	Putative protein of unknown function/ uncharacterized ORF	none
	Magnaporthe oryzae	G4N991	Afu7g04870 (AFUB_090420)	49	6.0 E-07	Putative glutamine-serine-proline rich protein/ uncharacterized ORF	high
			Afu1g02290 (AFUB_002680)	43	4.0 E-20	AFL6/ characterization underway	high
			Afu2g01405 (AFUB_018490)	39	8.0 E-04	AFL4/ characterization underway	high
			Afu4g11690 (AFUB_068700)	27	3.3	RNA polymerase II transcription factor related protein/ uncharacterized ORF	none
			Afu2g11760 (AFUB_027510)	26	4.9	Putative Signal transduction protein BroA/ uncharacterized ORF	none
Galectin	*Agrocybe aegerita*	Q6WY08	Afu5g07390 (AFUB_054940)	27	2.3	Putative 60S ribosome biogenesis protein Sqt1/ uncharacterized ORF	none
			Afu5g14380 (AFUB_062100)	27	3.4	Alpha-glucuronidase aguA/ verified ORF	low
			Afu7g04670 (AFUB_090210)	26	4.1	F1F0 ATP synthase assembly protein Atp11	none
			Afu5g13190 (AFUB_060900)	25	8.4	Kinesin-like protein/ Uncharacterized ORF	none

	Coprinopsis cinerea (CGL1)	Q06100	Afu7g01160 (AFUB_087740)	25	9.6	Putative Cytochrome P450 alkane hydroxylase/ uncharacterized ORF	none
	Coprinopsis cinerea (CGL2)	Q9P4R8	Afu1g15130 (AFUB_014680)	25	7.5	Putative GPI anchored transmidase/ uncharacterized ORF	none
	Coprinopsis. cinerea (CGL3)	Q206Z5	Afu2g15610 (AFUB_031260)	26	4.8	DNA-directed RNA polymerase/ uncharacterized ORF	none
Ig like	*Flammulina velutipes* (Immunomodulatory protein FIP-Fve)	P80412	Afu2g13520 (AFUB_029160)	28	7.9 E-01	Putative fermentation associated protein (Csf1)/ uncharacterized ORF	none
			Afu2g12860 (AFUB_028480)	27	1.1	DUF726 domain protein/ uncharacterized ORF	none
			Afu1g04260 (AFUB_004590)	27	1.5	Endo-1,3-beta-glucanase Engl1/ uncharacterized ORF	none
			Afu3g02900 (AFUB_045350)	25	3.7	Oxidoreductase, short chain dehydrogenase/reductase family/ uncharacterized ORF	none
			Afu4g00290 (AFUB_100820)	25	4.2	PutativeSuccinyl-CoA synthetase beta subunit/ uncharacterized ORF	none
L-type	*Saccharomyces cerevisiae* (Emp46p)	E7KFG8	Afu2g16800 (AFUB_032470)	51	3 E-07	Putative Lectin family integral membrane protein/ uncharacterized ORF	high
			Afu5g06870 (AFUB_054430)	27	6.5	Putative protein of unknown function/ uncharacterized ORF	none
			Afu5g03120 (AFUB_051630)	26	9	5'-3' exoribonuclease 1/ uncharacterized ORF	none
	Coprinopsis cinerea	A8N4Z2	Afu2g12180 (AFUB_02787)	243	1 E -64	Putative Lectin family integral membrane protein/ uncharacterized ORF	high
			Afu1g14680 (AFUB_014220)	40	1 E-03	DNA directed RNA polymerase/ uncharacterized ORF	none
			Afu2g16800 (AFUB_032470)	36	2.2 E-02	Putative Lectin family integral membrane protein/ uncharacterized ORF	high

			Afu2g13450 (AFUB_029090)	36	2.8 E-02	Putative dynactin/ uncharacterized ORF	none
			Afu1g11450 (AFUB_010880)	33	2.5 E-01	Putative myosin class II heavy chain (MHC)/ uncharacterized ORF	none
			Afu3g08500 (AFUB_040590)	32	6.2 E-01	Putative morphogenesis protein (Msb1)/ uncharacterized ORF	none
			Afu1g02730 (AFUB_003120)	31	7.8 E-01	Putative mitochondrial phosphate carrier protein (Ptp)/ uncharacterized ORF	none
			Afu3g05630 (AFUB_043370)	30	2.0 E-01	Putative phospholipase D1 (PLD1)/ uncharacterized ORF	none
Yeast adhesin	*Candida glabrata*	Q6VBJ0	Afu2g10820 (AFUB_026600)	31	5.2 E-01	Putative DEAD/DEAH box helicase/ uncharacterized ORF	none
			Afu1g10950 (AFUB_010370)	27	8.2	Putative Vesicle-mediated transport protein Vid24/ uncharacterized ORF	none
	Saccharomyces. cerevisiae	P38894	n.a	n.a	n.a	n.a	n.a.

Table8: supplemental table containing the raw data from the bioinformatical search. Aminoacid sequences from selected known fungal lectins were BLASTed against the genome of two *Aspergillus fumigatus* strains. Annotations are marked if available and the significance of the hits is indicated.

Chapter IV

Characterization of the cyanovirin-N homolog AFL6

Summary

In this Chapter, we present the biochemical and structural characterization of *Aspergillus fumigatus* lectin 6. The lectin had been identified from genomic data analysis as a member of the recently discovered cyanovirin–N homolog (CVNH) family (Koharudin et al. 2008). After recombinant expression in *E. coli*, affinity and size exclusion chromatography was carried out to obtain pure and homogenous protein. Commercially available crystal screening kits were used for crystallization trials and crystals of AFL6 were obtained. Diffraction data was collected at the European Synchrotron Radiation Facility (ESRF) in Grenoble. The structure was solved to high resolution in absence of the ligand. Attempts to determine the specificity of the lectin were performed using isothermal titration calorimetry and glycan array.

Résumé

Dans ce chapitre, nous présentons la caractérisation biochimique et structurale de la lectine 6 d'*Aspergillus fumigatus*. Cette lectine a été identifiée lors de l'analyse de données génomiques comme étant un membre de la famille des homologues de la Cyanovirin–N (CVNH) qui a été récemment découverte. Après la production de la protéine sous forme recombinante chez *E. coli,* une purification par chromatographie d'affinité suivie d'une chromatographie d'exclusion par taille a été réalisée afin d'obtenir une protéine pure et homogène. Des essais de cristallisation ont été réalisés, et ont permis d'obtenir des cristaux d'AFL6. Des données de diffraction ont été collectées au synchrotron à Grenoble et ont permis de résoudre à haute résolution par la méthode du remplacement moléculaire, la structure d´AFL6 en absence de ligand. Des essais pour déterminer la spécificité de la lectine ont été conduits par titration calorimétrique isotherme et par du criblage à haut débit avec des puces à sucre.

1. Introduction

1.1 Cyanovirin-N: a potent HIV inhibitor

Cyanovirin–N (CV-N) is a small soluble 11 kDa lectin from the cyanobacterium (blue- green algae) *Nostoc ellipsosporum.* It was first identified in 1997 by its potent antiviral activity (Boyd et al. 1997). CV-N was found to be non toxic for the host cells even at high concentration and its biological activity was highly resistant to physicochemical denaturation, making it an excellent candidate for a microbicide. It was shown that CV-N inactivated several laboratory strains and primary isolates of the human immunodeficiency virus HIV-1 and 2 as well as simian (SIV) and feline (FIV) strains and other enveloped viruses, such as Ebola (Barrientos et al. 2003). Furthermore, *in vitro* fusion assays of infected and non infected cells showed that CV-N prevented intercellular HIV transmission (Boyd et al. 1997). The antiviral property of CV-N is mediated by its interaction with the viral envelope glycoprotein gp120.The infection of HIV is initiated by adhesion of the virus to the host cell. Lectins are at the interface of these interactions. For example, the C-type lectin DC-SIGN recognizes high-mannose glycans on the viral envelope glycoproteins gp120. The virus hijacks the dendritic cells presenting DC-SIGN resulting in increased efficiency of infection of secondary lymphoid organs and their target CD4$^+$ cells (da Silva et al. 2011). Figure 32 shows an overview of HIV infection in the dendritic cell.
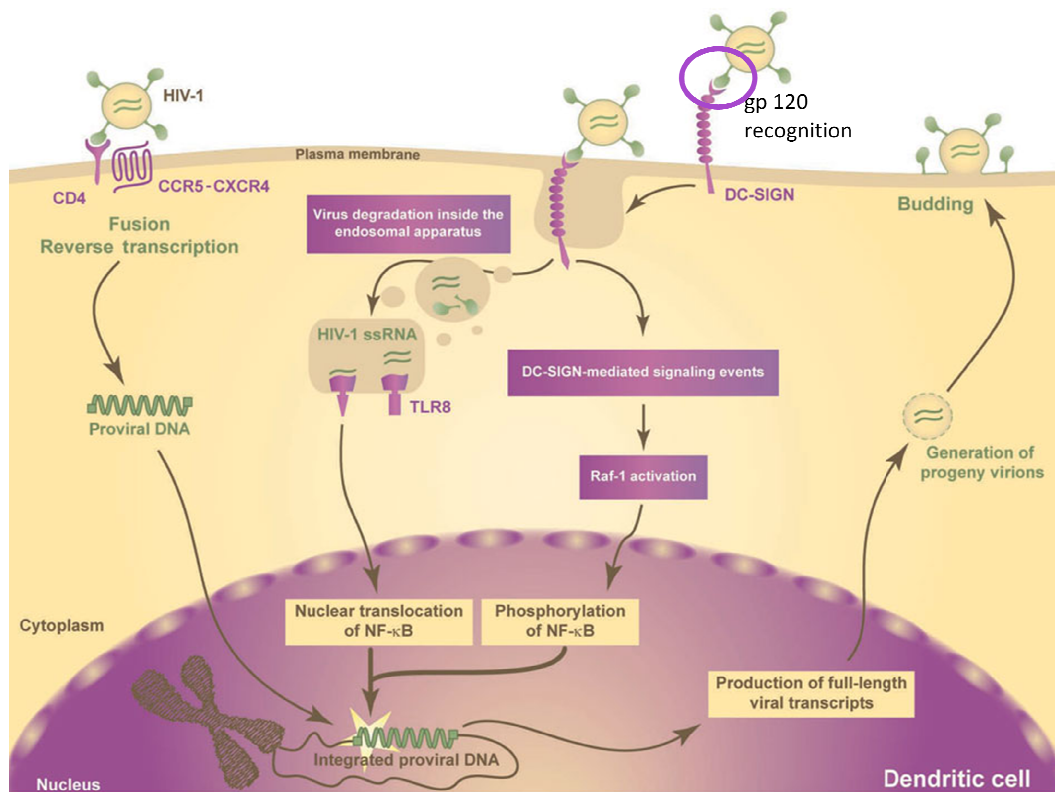
Figure 32: Overview of HIV infection pathways in dendritic cells. The infection is mediated by binding of gp120 to DC-SIGN and CD4. Picture adapted from (Tremblay 2010)

The finding that CV-N interacts with the virus envelope glycoproteins gp120 and gp41 and blocks their interaction with the host receptors, spurred great interest in the scientific community towards the development of a topical HIV prevention (Esser et al. 1999; Dey et al. 2000; Bolmstedt et al. 2001). The viral surface of HIV is covered with heterodimers of the heavily glycosylated glycoproteins gp120 and gp41. Gp120 contains more than 20 N-glycosylation sites and N-glycans, that include high mannose type and complex type oligosaccharide, account for about half its molecular weight (Zhu et al. 2000).
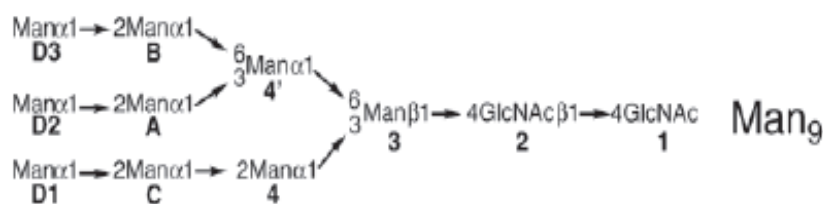
Figure 33: schematic representation of Man9, a high mannose type N-glycan present on gp120.Picture reproduced from (Bewley & Otero-Quintero 2001)

The anti HIV activity of CV-N is due to its high affinity binding to α1-2 mannosides presented at the non- reducing end of the D1 and D3 arms of high mannose type glycans (figure 33). The structural base of the carbohydrate recognition and thermodynamical parameters of CV-N has been studied intensely (Bewley & Otero-Quintero 2001; Bolmstedt et al. 2001; Shenoy et al. 2001; Fujimoto & Green 2012; Botos et al. 2002; Dey et al. 2000). CV-N is a small protein, comprising 101 amino acids and is made up of two pseudo-symmetric domains A and B that share 32% sequence identity. Each domain contains one carbohydrate binding site. Isothermal titration calorimetry experiments with gp120 and gp41 showed high affinity binding with a K_D of 37 nM and a stoichiometry of 5:1 that is 5 molecules of CV-N bind to 1 glycoprotein. While both binding sites A and B are specific for Manα1-2Man motifs, they differ in their affinity. The binding site on domain A shows a lower affinity with a K_D around 3 µM compared to the binding site on domain B which shows affinity in the nanomolar range (Bewley et al. 2002). NMR structures of CV-N with Manα1-2Man confirmed that the morphology of the two binding sites is different; with binding site A being a semicircular cleft that partially surrounds the disaccharide. The high affinity binding site B on the other hand was shown to be a deep pocket that nearly envelopes the disaccharide thus providing structural evidence for the observed difference in affinity (Bewley & Otero-Quintero 2001).

The overall solution structure of monomeric CV-N has the shape of an elongated ellipsoid, about 55 Å in length with a maximum width of about 25 Å. Secondary structure elements consist of 10 β- strands and four short helical turns (Bewley et al. 1998). Interestingly, in crystal structures, CV-N can be observed as a domain swapped dimer, where domain A of one monomer interacts with domain B` from the other monomer (figure 34).
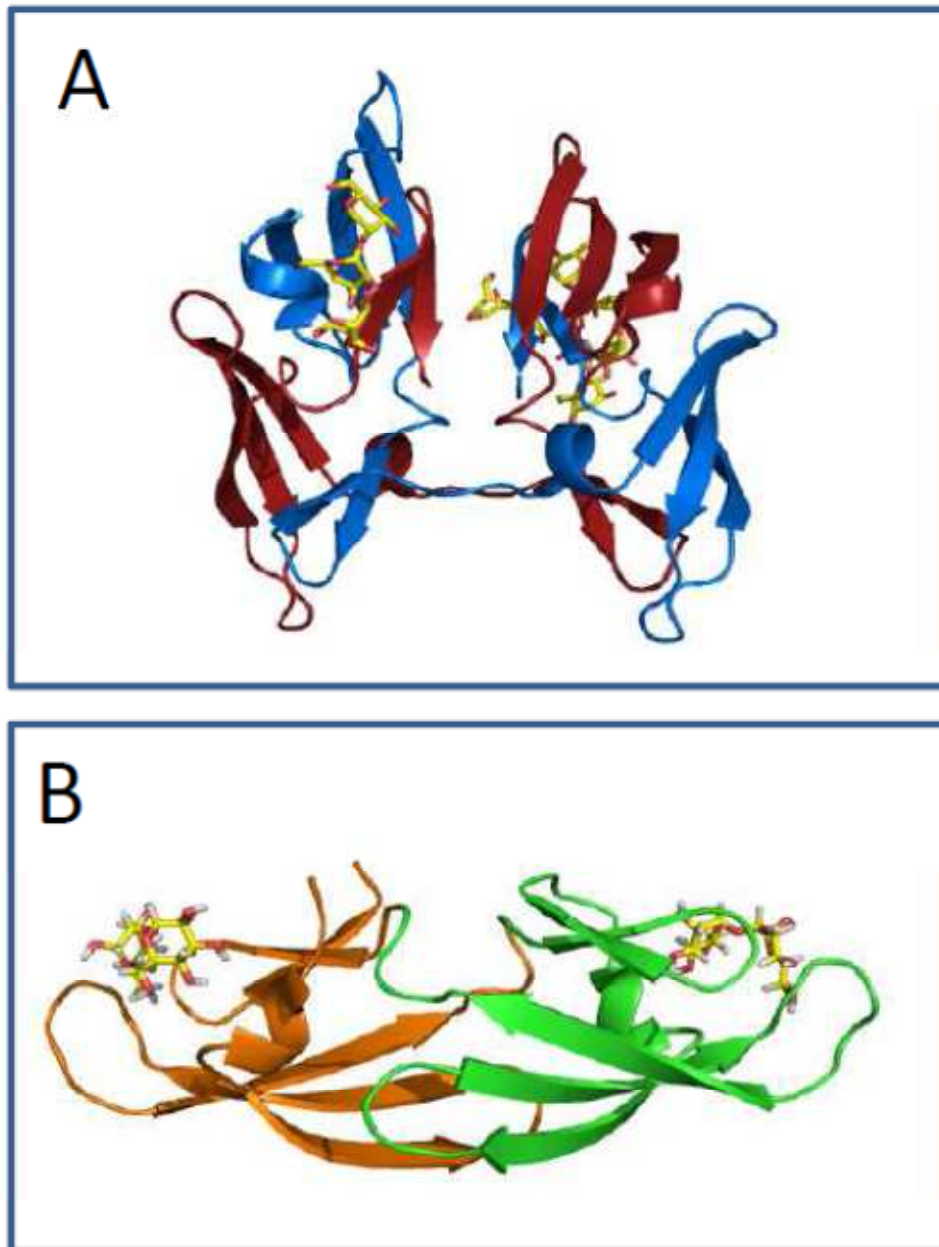
Figure 34: 3 dimensional structures of CV-N complexed with oligosaccharides. Domain swapped dimer (A) complexed with the D3 arm of oligomannose -9 (PDB: 3GXZ). The two monomers are colored in red and blue respectively.The CV-N monomer in complex with α1-2dimannoside is shown in panel B (PDB: 1IIY) where domains A and B are coloured in orange and green respectively,

This domain swapped state however, does not result in additional intramolecular interaction and does not affect HIV inhibition. It is much more a metastable, kinetically trapped structure resulting from crystallisation conditions and CV-N is thought to be monomeric in its naturally occurring state (Yang et al. 1999; Barrientos et al. 2002). However, it has recently been shown, that CV-N oligomers designed to exhibit the domain swapped form in solution, not only displayed enhanced HIV neutralization but were active against a broader spectrum of different HIV strains (Keeffe et al. 2011). Taken together, the intense studies of the interactions between CV-N and its carbohydrate ligands enabled researchers to develop HIV entry inhibitors that can be used as topical microbicides for vaginal and possible rectal application. For pharmacological evaluation of such virucides, cost effective large scale production of CV-N is of utmost importance. Recombinant expression in *E. coli* has been successful in the periplasm but yields are low (Mori et al. 1998). CV-N is often misfolded and purification from inclusion bodies and subsequent refolding is cost- and time consuming (Colleluori et al. 2005). The most promising attempt of expression in *E. coli* included using a SUMO (Small-Ubiquitin-like Modifier) tag that is known to facilitate correct protein folding. A yield of 150 mg L^{-1} culture of recombinant CV-N in a fed-batch fermentor is reported for this system (Gao et al. 2009). Furthermore, other bacterial expression systems, including *Steptococus gorondii* and *Lactobacillus jensenii,* show promising applications (Giomarelli et al. 2002; Brichacek et al. 2013). Eukaryotic systems have been exploited as well, including the yeast *Pichia pastoris* but glycosylation inactivated the protein and yields were low (Mori et al. 2002; Xiong et al. 2010). Even though large scale production is still a challenge for this lectin, it has been proven to be an excellent candidate for a future HIV prevention that is desperately needed.

1.2 Cyanovirin-N homologs in fungi and plant

Besides the investigation towards a microbicide based on CV-N, bioinformatical research revealed that the CV-N lectin fold is an evolutionary conserved family of proteins that are mainly found in fungi and some seedless plants. AFL6 belongs to this recently discovered family of cyanovirin-N homologs. First indications that the cyanovirin-N is part of an evolutionary conserved motif were found in 2005. Perucudani and co-workers conducted a gene expression profiling study in the ascomycetous truffle *Tuber borchii* and found a nutrient regulated mRNA (M5G11) that showed significant similarities with other hypothetical proteins in filamentous fungi and seedless plants. Further bioinformatical

analyses revealed a significant similarity with CV-N. All of the identified proteins had an internal tandem repeat with an average internal similarity of 30% and the conserved amino acids matched those in CV-N that were mainly located at the hydrophobic core region, which is important for the structural integrity of the protein. Taken together, these findings led to the creation of the term cyanovirin-N homolog (CVNH). Among the identified family members of CVNH were several hypothetical proteins from *Aspergillus* spp including *A. flavus*, *A. nidulans*, *A. niger* and later, when the whole genome was available, *A. fumigatus* (Percudani et al. 2005; Koharudin et al. 2008). Based on their domain organization the CVNH family has been grouped into three types (figure 35).
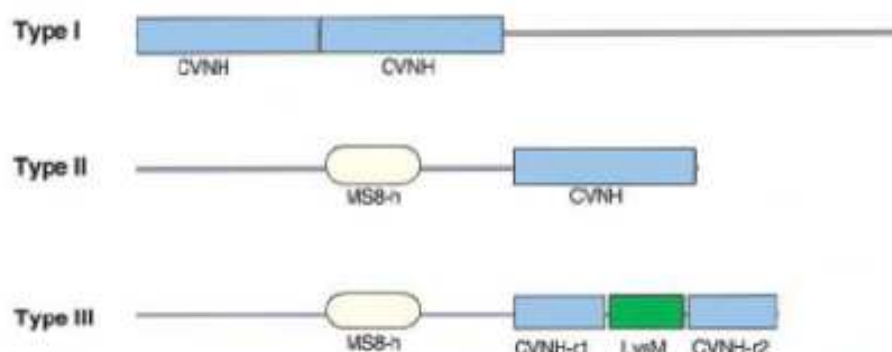
Figure 35: Schematic overview of the domain structure in the three types of CVNH
Picture reproduced from (Percudani et al. 2005)

In type I CVNH, one or multiple CVNH domains are found at the N-terminus of the protein sequence. Type I CVNH are the most studied and structures are available from the truffle *Tuber borchii* (TbCVNH), the mold *Neurospora crassa* (NcCVNH), the underwater fern *Ceratopteris richardii* (CrCVNH) (Koharudin et al. 2008) and the wheat head blight fungus *Gibberealla zeae* (GzCVNH) (Matei et al. 2011). Type II CVNH encoding genes have been found in *Aspergillus* species, where the CVNH domain is preceded by an MS8 coding region of unknown function. To our knowledge no type II CVNH has been structurally characterized yet. In type III the CVNH domain is interrupted by a lysin motif (LysM) in between the repeats. NMR solution structures and carbohydrate interaction studies of a type II LysM CVNH are available for the rice blast fungus *Magnaporthe oryzae* (Koharudin et al. 2011).

Type I CVNH remains the most intensely studied type of CVNHs, and AFL6 is a member of this type. Thus the focus will be directed towards the description of type I CVNH lectins. Based on multiple sequence alignments type I CVNH can be subcategorized into three groups. Group I contains CV-N itself and microvirin (MVN), from the cyanobacterium

Microcystis aeruginosa (Kehr et al. 2006) as well as CVNH from fern and fungi. The other groups are entirely composed of CVNH polypeptides from filamentous Ascomycetes (Koharudin et al. 2008). The cladogram shows the relations among the different organisms and is depicted in figure 36. Among them, a lectin from *A. fumigatus,* henceforth called AFL6 was identified, belonging to the group III CVNH.
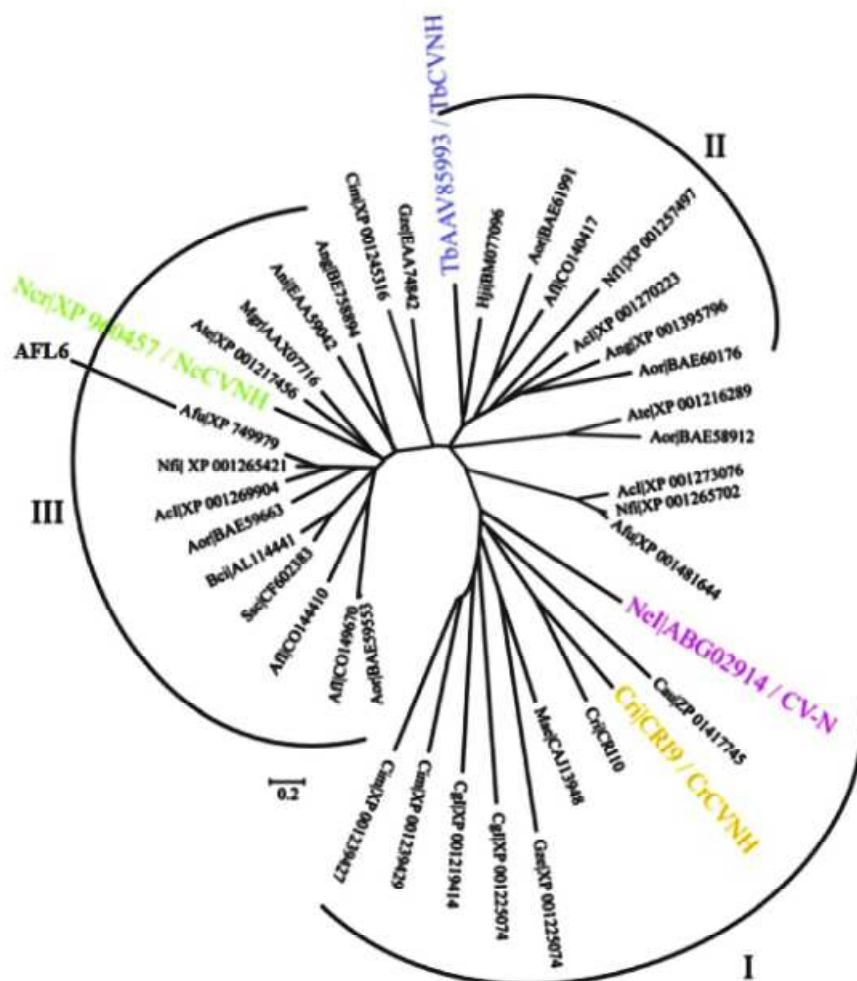
Figure 36: Cladogram of 39 predicted polypeptide sequences that are CVNH family members. Picture adapted from (Koharudin et al. 2008)

3D structures for one member of each type I CVNH group are available (depicted in green, blue and yellow in figure 37). Apart from their conserved CV-N fold, they share carbohydrate specificity for α1-2 linked mannosides in high mannose type glycans. However, only one binding site appeared to be active in CVNH from *Tuber borchii* (TbCVNH, group II) and *Neurospora crassa* (NcCVNH, group III) as discovered by ^{1}H-^{15}N HSQC spectroscopy. On the other hand, the representative of group I from the fern *Ceratopteris richardii* (CrCVNH), the closest relative of CV-N, showed two active binding sites. Further investigations toward

111

carbohydrate specificity using glycan array showed that TbCVNH interacted with linear glucose sugars, and a combination of glucose, galactose and N-acetylglucosamine containing glycans as well as sucrose (Glcα1,2Frc) (Koharudin et al. 2008). These findings led to the design of a chimeric CVNH, consisting of the active domain A from TbCVNH and the active domain B of NcCVNH. This chimera binds Manα1-2Man with very low affinity with K_d values in the millimolar range. It also binds sucrose even though in the parent NcCVNH sucrose binding could not be detected (Koharudin et al. 2009). In this chapter, we present the biochemical and structural characterization of AFL6, a CVNH from *A. fumigatus*.

2. Experimental Details

2.1. Cloning of AFL6 in expression Vector pET TEV

For recombinant expression in *E.coli*, the gene comprising 321 bp was ordered at GenScript (USA) with codon optimization and appropriate restriction sites. It was provided in a pUC57 vector and sub-cloned into the expression vectors pET-TEV (Houben et al. 2007) and pET-TEVbiot (figure 37) using *Nde I* and *Xho I* FastDigest™ restriction enzymes from Fermentas. The resulting recombinant AFL6 protein contained an N-terminal 6-Histidin tag for affinity purification which can be cleaved off by Tobacco Etch Virus (TEV) protease produced in the laboratory. After cloning into pETTEVbiot the resulting AFL6 contained an additional N-terminal avi-tag, which encodes the sequence GLNDIFEAQKIEWHE for *in vitro* biotinylation.
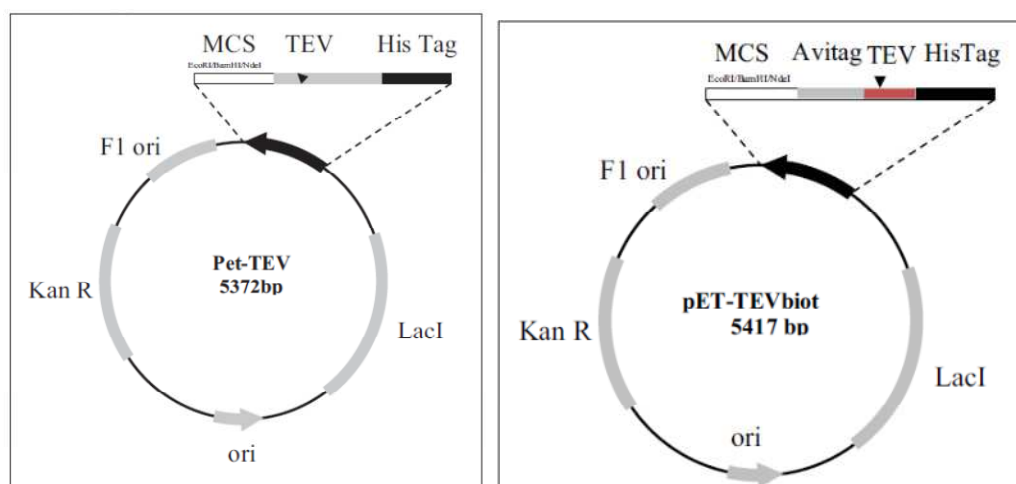
Figure 37: Plasmid maps for pET-TEV (left) and pET-TEVbiot (right). The N-terminal HisTag can be cleaved of by TEV-protease. The plasmids contain resistance to kanamycin (KanR) for selective growth. MCS: multiple cloning site.

The Ligation was carried out using DNA Ligation Kit Mighty Mix (TaKaRa) and the ligated plasmid was amplified in *E.coli* XL-1 cells. Colony-PCR was carried out to identify positive clones using Taq`Ozyme Purple Mix (OZYME) and T7 primers. The plasmid DNA of positive clones was purified using Wizard® SV Miniprep DNA Purification System (Promega). The appearance of a band at 321 bp confirmed the correct insertion of the insert after control restriction with the same enzymes. The plasmid was then send to GATC were sequencing was carried out using T7 primers and the correct sequence of the optimized *Afl6* gene was confirmed.

2.2 Recombinant expression of AFL6 in *E. coli*

1 µl of *Afl6* pET-TEV 28a was transformed into the *E. coli* expression strain BL21 (DE3) from Novagen. The day before expression a pre-culture of 15 ml LB liquid medium containing 30 µg ml^{-1} kanamycin was inoculated with several colonies of *E. coli* BL21 (DE3) containing the above mentioned AFL6 plasmid. The pre-culture was used to inoculate a 1 L of liquid LB medium containing 30 µg ml^{-1} kanamycin. Bacteria were grown until an OD$_{600}$ of 0.6- 0.8 and protein expression was induced with 0.1 mM IPTG and performed 16 ˚C overnight. Cells were then harvested by centrifugation at 6000 g for 20 min. The pellet was frozen and stored at -20˚C until purification.

2.3 Purification of recombinant AFL6

The frozen pellet was resuspended in 15-20 ml Buffer A containing 20 mM Tris/ HCl pH 7.5, 500 mM NaCl and 10 mM imidazole. The cells were disrupted at 1.9 kbar with the One Shot Cell Disrupter (Constant Systems Ltd.) and centrifugation at high speed separated the cell debris, so that the supernatant contained the recombinant His-tagged rAFL6. Affinity chromatography was carried out using Fast Protein Liquid Chromatography (FPLC, GE Healthcare) where the clear cell lysate was loaded onto a 1 ml Nickel-NTA His-Trap™ column (GE Healthcare) for immobilized metal affinity chromatography (IMAG). His-rAFL6 was eluted with a linear gradient of 10 – 500 mM Imidazole. Peak fractions were pooled and concentrated in a centrifugation- ultrafiltration unit (Vivaspin, Sartorius) to a concentration higher than 1 mg ml^{-1} for subsequent cleavage. The His-tag was cleaved off using TEV protease in Buffer A supplemented with 0.5 mM EDTA, 0.25 mM TCEP and a ratio 1:50 TEV protease: His-rAFL6. After incubation at 10˚C overnight the TEV protease and uncleaved His-rAFL6 were separated by IMAG and cleaved rAFL6 was collected in the flow through. For crystallization studies AFL6 was further purified by gel filtration using a

Superdex S75 16/60 column (GE Healthcare) in 20 mM MES or Citrate Buffer at pH 6.0 with 150 mM NaCl.

To verify the expression and purification of rAFL6 and to monitor the purity, each step was analyzed on 18% SDS-PAGE. Samples of 16 µl were mixed with 4 µl 5x loading buffer and denatured at 95˚C for 5 min. before loading. The Precision Plus Protein™ Standards Unstained (BioRad) and InstantBlue (Expedeon) were used as molecular weight marker and gel staining solution respectively.

2.4 Hemagglutination

Hemagglutination tests were carried out in U-shaped 96-well plates using rabbit red bold cells (RBC) from Biomérieux, Lyon. 50 µl were serially two-fold diluted in PBS and the same volume of 2% RBC in 150 mM NaCl was added. The hemagglutination was incubated for 1h at room temperature and the results were visually analyzed.

2.5 Thermal Shift Assay

To test the thermal stability of AFL6, a range of different buffers with a pH ranging from 4.5 to 8.5 were tested. Buffer stocks were prepared at 100 mM. To screen these 24 conditions a mix was prepared containing 70 µl AFL6 at 2.7 mg ml^{-1}, 7 µl 500x Sypro Orange and 63 µl H$_2$O. In each well 5 µl of this mix were combined with 12.5 µl 100 mM Buffer and 7.5 µl H$_2$O. This leads to a final concentration of 50 mM Buffer and 0.27 mg ml^{-1} AFL6. The thermal shift assay was then performed with Bio-Rad Real-time PCR system MiniOpticon using the Bio-Rad CFX Manager software. The temperature was gradually raised in 1ºC steps from 20ºC to 100ºC.

2.6 Labelling of AFL6 with biotin and glycan array

In vitro biotinylation was carried out in according to the manufacturer's protocol using rAFL6avi and NHS-linked biotin (Sigma). Brief, 25 µl of 100 mM biotin in DMF was slowly added to a solution of 10 mg ml^{-1} of AFL6avi in PBS pH 7.2 and incubated for 30 min at room temperature with gentle agitation. Biotinylated protein was then purified with a 5 ml polyacrylamide desalting column (Thermo Scientific) in 20 mM MES buffer pH 6.0 with 150 mM NaCl and fractions of 500 µl were collected. To identify biotin-labeled fractions 2.5 µl of each fraction were spotted on a nitrocellulose membrane which was blocked with 1% Bovine serum albumin (BSA) in TRIS buffered saline containing 0.1 % Tween-20 (TBST) and subsequently incubated with peroxidase-labelled streptavidin (Sigma). Colorimetric detection of the peroxidase reporter was performed using 1-Step UltraTMB Blotting Solution

(Thermo Scientific) and fractions containing biotinylated rAFL6avi appeared as blue spots on the membrane.

The biotinylated rAFL6avi was then sent to the Consortium for Functional Glycomics (CFG) in the USAwhere it was analyzed on their printed glycan array version 5.2 containing 609 glycans in replicates of 6. For fluorescent detection, Cyan-5-streptavidin was provided by the CFG.

2.7 Isothermal titration calorimetry (ITC)

Calorimetric measurements were carried out using the Microcal iTC 200® calorimeter from GE Healthcare. In a typical experiment, 0.4 µl ligand (30 mM) were injected in 120 second intervals from a 40 µl syringe into the rapidly stirred (1000 rpm) sample cell containing 200 µl of 60 µM AFL6. Control experiments involved injection of ligand into buffer without AFL6. Titrations were carried out at 25˚C in 20 mM sodium citrate buffer pH 6.0 containing 150 mM NaCl. The isotherms were fit using the Origin ITC Analysis software according to the manufacturer`s protocols. The stoichiometry was fixed to extract enthalpy and binding affinities from the hyperbole binding curve. Thermodynamic parameters were calculated using $\Delta G = -RT \ln K_a$ and $\Delta G = \Delta H - T\Delta S$.

2.8 Dynamic light scattering

Experiments of dynamic light scattering were carried out using the Zetasizer Nano ZS from Malvern Instruments Ltd and a 40 µl quartz cuvette. Prior to measurement, the purified protein solution was centrifuged at high speed for 20 min. to remove bigger particles like dust contamination or aggregated proteins.

2.9 Structural characterization

2.9.1 Crystallization

Crystallization trials were carried out by hanging drop vapor diffusion on 24 well plates at 20˚C. To screen initial crystallization conditions commercially available screening kits Structure Screen2 and Morpheus from Molecular Dimensions Ltd were used. Drops comprised 1 µl 10 mg ml^{-1} AFL6 and 1 µl mother liquor. Single needles were collected and cryo-protected in 80% saturated lithium sulfate solution and frozen in liquid nitrogen.

2.9.2 X-ray data collection and processing

X-ray diffraction data were collected at the European Synchrotron Radiation Facility (Grenoble) at beamline BM30A using an ADSC Q4 CCD detector. Integration of the collective data set was done using XDS (Kabsch 2010). All further computation was made using programs within the CCP4 suite (Winn et al. 2011).The space group as determined by POINTLESS (Evans 2006) and the data was scaled and merged with aimless (Evans & Murshudov 2013). The number of molecules in the asymmetric unit and the solvent content was determined by Matthews (Kantardjieff & Rupp 2003).

2.9.3 Molecular replacement

The phases were obtained by used molecular replacement using the coordinates of the chimeric CVNH (PDB: H3P8 (Koharudin et al. 2009)). After modification using chainsaw (Stein 2008) the phases was used to solve the structure (McCoy et al. 2007). After generation of the initial model, the chain visualization and model building was done using Coot (Emsley & Cowtan 2004). Iterative refinement was carried out by alternating between manual rebuilding in Coot and automated refinement in Refmac5 (Murshudov et al. 2011). Solvent water molecules were added and 5% of the reflections were kept aside for cross validation. The last round of refinement was done using anisotropic B-factors.

3. Results and Discussion

3.1 Cloning, expression and purification of recombinant AFL6

Optimization of expression conditions resulted in the production of rAFL6 exclusively in the soluble fraction, indicating optimal conditions for expression and protein folding (data not shown). Large scale expressions were henceforth carried out at these conditions and led to a protein production of 8-10 mg per litre culture that was purified via Nickel-affinity chromatography. Figure 38 shows the purification chromatogram and SDS PAGE analysis.
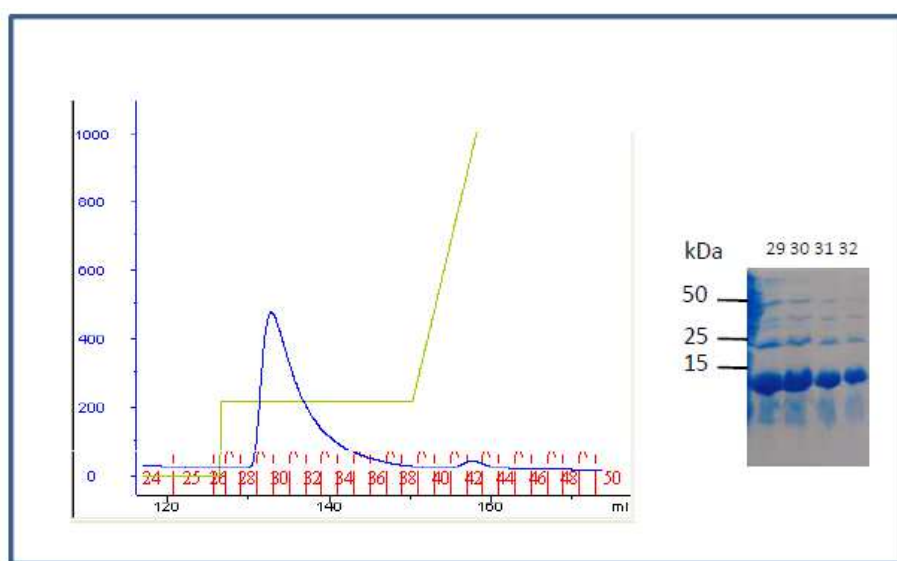
Figure 38: Elution profile and 18% SDS PAGE of rAFL6 affinity chromatography. The protein eluted with 150 mM imidazole. UV absorption is coloured in blue and the increase of imidazole is shown in green.

The Nickel affinity chromatogram shows the elution of His-rAFL6 150 mM imidazole. The height of the elution peak is dependent on the UV absorbance of the protein at 280 nm. Since AFL6 only contains one tryptophan residue and no tyrosine residue the absorbance is low, resulting in an extinction coefficient of 5500 M^{-1} cm^{-1}. The his-tag was cleaved using TEV as described above. Cleaved rAFL6 was found in the flowthrough of subsequent Nickel affinity chromatography (data not shown).

Gel filtration chromatography was carried out to determine the oligomerization state of AFL6 and to further purify the protein for crystallisation studies. AFL6 eluted at 83 ml corresponding to an estimated size of 12 kDa which corresponds to the theoretical size of a monomer with 11.7 kDa. This is in good accordance with previously published characterizations of different CVNH which were described as monomeric in solution

(Koharudin et al. 2008). The homogeneity of the protein solution was investigated using dynamic light scattering where no aggregation was observed (data not shown) and monodisperse protein solution was used for crystallization trials.

3.2 Hemagglutination assay

rAFL6 showed no hemagglutination activity at 1 mg ml^{-1} on rabbit erythrocytes (data not shown). While CV-N shows hemagglutination activity, its fungal homologs NcCVNH and TbCVNH are not able to agglutinate erythrocytes (Koharudin et al. 2008). The lack of hemagglutination activity may be due either to the absence of the putative carbohydrate ligand on erythrocytes, or to the existence of only one active binding site on AFL6 as it is the case for NcCVNH and TbCVNH. Besides this, it may be possible that the binding site was occupied by a tightly bound ligand during expression.

3.3 Thermal shift assay: stability assessment

Investigating the stability of rAFL6 in different buffer conditions using varying pH and salt concentrations is an important step on the way to determine optimal crystallization conditions. AFL6 shows a melting temperature Tm of 47˚C at a pH of 4.5 in acetate buffer, which is independent of salt concentration. Augmentation of the pH to 5.6 and 6.0 shows an increase of Tm to 60˚C with 100 and 200 mM NaCl.

Buffer system		Tm ºC	+100 mM NaCl Tm ºC	+200 mM NaCl
Sodium acetate	pH 4.5	47	47	47
Sodium citrate	pH 5.6	54	60	60
Mes	pH 6.0	55	60	60
Sodium Cacodylate	pH 6.5	55	56	53
Hepes	pH 7.0	53	54	51
Hepes	pH 7.5	53	53	55
Tris/HCl	pH 8.0	53	54	54
Tris/HCl	pH 8.5	51	54	53

Table 9: overview of the conditions tested on TSA

For crystallization studies, AFL6 was kept in 20 mM MES buffer at pH 6.0 containing 150 mM NaCl. For ITC studies however, results were better when rAFL6 was kept in 20 mM Citrate buffer at the same pH and salt concentration.

3.4 Structural characterization

3.4.1 Crystal structure of rAFL6

Crystals of rAFL6 were obtained in 2.0 M ammonium sulphate with 5% v/v 2-propanol at a concentration of 10 mg ml^{-1} and resulted in a star of thick needles (figure 39). Diffraction data was obtained on single crystals broken from such stars and indicated an orthorhombic unit cell belonging to space group C222$_1$. The unit-cell dimensions were a=42.74Å, b=43.85Å, c=86.7Å with one molecule in the asymmetric unit. Data collection statistics are provided in Table 10. The crystal diffracted to 1.35 Å and the final model showed clear electron density for all amino acids after refinement. Refinement statistics are given in table 11.

Figure 39: Crystal of AFL6 obtained in 2.0 M ammonium sulfate and 5% v/v 2-propanol with a concentration of 10 mg ml^{-1} AFL6 and a protein to mother liquor ratio of 1:1.Crystals trials were carried out with hanging drop vapor diffusion using Structure Screen 2 from Molecular Dimensions. UV light was used to ensure the crystal is made of protein and not salt molecules.

Unit cell: 42.74 43.85 86.7 90 90 90 (orthorhombic)
Space group: C222$_1$

Resolution	overall 43.38Å-1.35Å	outer shell 1.385Å-1.35Å
R merge (within I+/I-)	0.049	0.449
total number of observations	103368	5132
Total number unique	18003	863
Mean((I)/sd(I))	15.8	3.6
Mn (I) half set correlations CC(1/2)	0.999	0.908
Completeness	98.5	96.7
Multiplicity	5.7	5.9

Table 10: Crystallographic statistics for rAFL6

R$_{crys}$/R$_{free}$	**15.3/ 19.7**
Rmsd bonds (Å)	0.014
Rmsd angles (^{o})	1.68
Protein atoms	841
B-factor (Å^2)	15.03
water molecules	107
B-factor (Å^2)	26.78

Table 11 : refinement statistics

The structure of rAFL6 can be described as an elongated ellipsoid with two domains. The two pseudo-symmetric halves consist of domain A and B respectively. The two domains are formed by strand exchange between the two tandem repeats. The stand exchange is shown in figure 40.
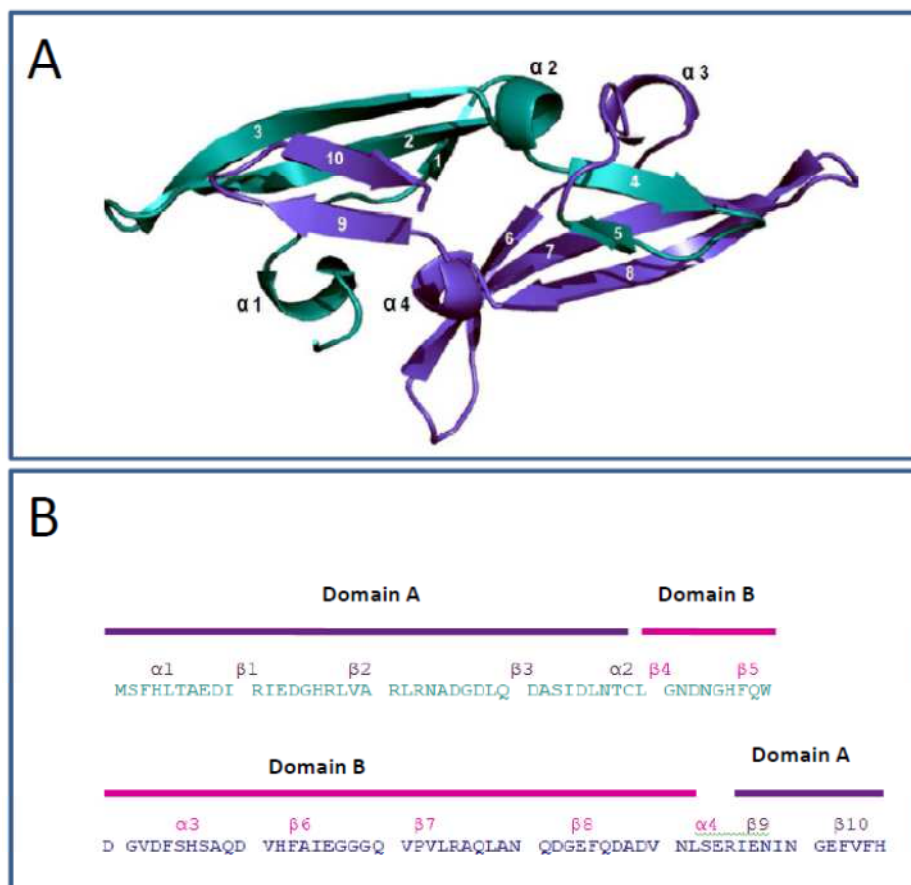
Figure 40: Cartoon representation and sequence of AFL6. The tandem repeats are shown in cyan and blue respectively. The secondary structure features are written in purple for domain A and magenta for domain B. The β-hairpin of the first tandem repeat is part of domain B and vice versa.

Domain A comprises residues 1-40 and 96-106 (50 residues) and domain B extends from residue 41-95 (54 residues). Thus the β-hairpin of the first tandem repeat is part of domain B and vice versa. Domain A contains two helical turns (α1 and α2), a three stranded β-sheet (β1-β3) and a β-hairpin (β9 and β10) wherein the β10 consists only of 3 amino acids. Domain B is composed likewise with two helical turns (α3 and α4), a triple stranded β-sheet (β6-β8) and a small β-hairpin (β4 and β5) (Figure 41).
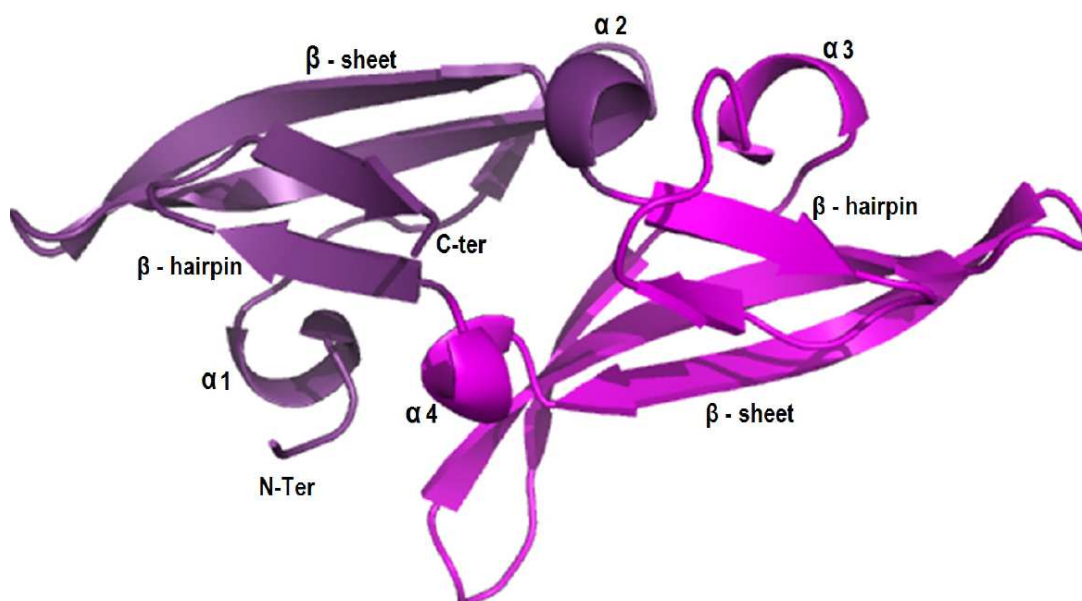
Figure 41: Cartoon representation with secondary structre features of AFL6. Domain A and B are shown in purple and magenta respectively.

3.4.2 Comparison of AFL6 with other CVNH

In agreement with other known CVNHs AFL6 is monomeric in crystal form as well as in solution, as determined by gel filtration and DLS. CV-N can exist both in monomeric form and as a domain swapped dimer, where one of the domains is swapped between the two dimer halves as observed in non-physiological conditions such as the crystal state and at high concentration in solution (Barrientos et al. 2002; Yang et al. 1999). In CV-N, the domain swap is enabled by an unusual trans configuration of proline 51 in the hinge region (W_{49} Q_{50} P_{51} S_{52} N_{53}) (Bewley et al. 1998; Barrientos et al. 2002). Some sequence variability is found at this position in the CVNHs since the proline 51 is replaced by a methionine in TbCVNH (W_{50} G_{51} M_{52} Q_{53} N_{54}) and a glycine in both NcCVNH (W_{52} G_{53} G_{54} Q_{55} N_{56}) and AFL6 (W_{49} D_{50} G_{51} V_{52} D_{53}). The sequence variation in AFL6 results in a structural difference in this area and domain swapping or dimerization has not been observed in the CVNHs characterized so far.

AFL6 belongs to the group III CVNH as identified by sequence similarity. Interestingly, molecular replacement worked best when the coordinates of the chimera LKAMG (PDB:3HP8) was used, which was designed from domain A of TbCVNH (group II) and domain B of NcCVNH (group III) (Koharudin et al. 2009). AFL6 shares only 65% sequence similarity (34% identical) with TbCVNH compared to 77% similarity with NcCVNH (52%

identity) and the chimera LKAMG (43% identity). Figure 42 shows the multiple alignment of the two tandem repeats of AFL6 with the founding member CV-N (group I) and two members of each group namely TbCVNH and NcCVNH. Detailed structural analysis will focus on these two members of group II and III respectively in comparison with AFL6.

First repeat:
```
                              α1        β1          β2           β3        α2      β4        β5
CVN        --LGKFSQTCYNSAIQGS----VLTSTCERTNGGYNTSSIDLNSVIENVDGSLKW
NcCVNH     GSHMSFHVTAEDARIEVRDNRTILFARLRREDGEWNDASYELDQIIGNNDGHFQW
TbCVNH     GSHMSYADSSRNAVLTNGGR--TLRAECRNADGNWVTSELDLDTCIGNPNGFLGW
AFL6       ---MSFHLTAEDIRIEDGHR---LVARLRNADGDLQDASIDLNTCLGNDNGHFQW
              .:   :. :  :        * :   .. :*     :. :*:   : * :* : *
```

Second repeat
```
                              α3        β6          β7           β8        α4      β9        β10
CVN        QPSNFIETCRNTQLAG---SSE---LAAECKTRAQQFVSTKINLDDHIANIDGTLKYE-
NcCVNH     GGQNFTETAEDIRFHPKEGAAEQPILRARLRDCNGEFHDRDVNLTEIVENVNGEFQAKF
TbCVNH     GMQNFSHSSEDIKLEE---GGRK--LTCRPKTVDGGFRERQGIDLNRIQNVNGRLVFQ-
AFL6       DGVDFSHSAQDVHFAIE-GGGQVPVLRAQLANQDGEFQDADVNLSERIENINGEFVFH-
              :*  .:..: ::          ...   * ..        * . .    : : *::* :  .
```

Figure 42: multiple sequence alignment of the two tandem repeats of CV-N with homologs from *N. crassa* (NcCVNH), *T. borchii* (TbCVNH) and *A. fumigatus* (AFL6). Identical residues are colored in purple, similar residues are colored in blue. Cysteine residues are colored yellow.

The overall structure of the elongated ellipsoid is conserved throughout the CVNH- family. Figure 43 shows the structural alignment with NcCVNH (figure 43A) and TbCVNH (figure 43 B). A great variability can be found in the N-terminus, were the first α-helix is not structurally conserved in NcCVNH. The following β1 stand is much shorter than the β1 in TbCVNH or NcCVNH where all β-strands of the three stranded β-sheet are about the same size. Furthermore, differences in the structure are found in the loops connecting the β-strands β1 and β2 in domain A. In NcCVNH, these loops are significantly longer than in AFL6 and in TbCVNH. Variations are observed in both the number of residues and the orientation of the loops. In NcCVNH, the β1 strand is also elongated compared to AFL6, accounting for significant structural differences in this area. The same is true for the loop connecting β-strands 6 and 7. While the β7 stand is not elongated, its position is significantly different from β6 in AFL6. Indeed considerable variation of this loop can be found in CV-N, TbCVNH, NcCVNH as well as AFL6. In the model, used for molecular replacement to solve the structure of AFL6 this loop had to be removed, because of its flexibility throughout the

CVNH structures. In TbCVNH the β-strands 7 and 8 are significantly smaller, allowing more flexibility in the loop that connects these β-strands.
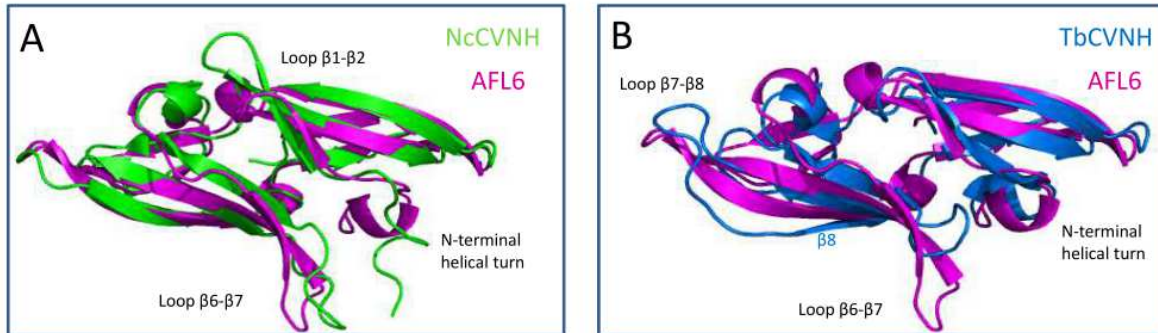
Figure 43: structural alignment of AFL6 (magenta) with (A) NcCVNH (green PDB: 2JZL) and (B)TbCVNH (blue, PDB: 2JZK). Structural differences can be found in the N-terminal helical turn and the indicated loops connecting β-strands.

Highly conserved amino acids are predominantly found at the domain interfaces, resulting in the overall structural similarity. (Koharudin et al. 2008). In AFL6 these residues are F3, I12, L18, L36 and L40 for domain A which correspond to residues F4, I13, L18, L27 and I41 in CV-N; Y3 L12, L19, L37, and I41 in TbCVNH and F3,I12, L21, L39 and I43 in NcCVNH. For domain B the conserved amino acids comprise F54, F63, L74, L92 and F103 in AFL6 which relate to residues F54, L63, L69, L87 and I91 in CV-N; F55, L64, L70, L88 and V92 in TbCVNH and F57, F66, L78, L96 and V100 in NcCVNH. These residues are shown as stick representations in figure 44.
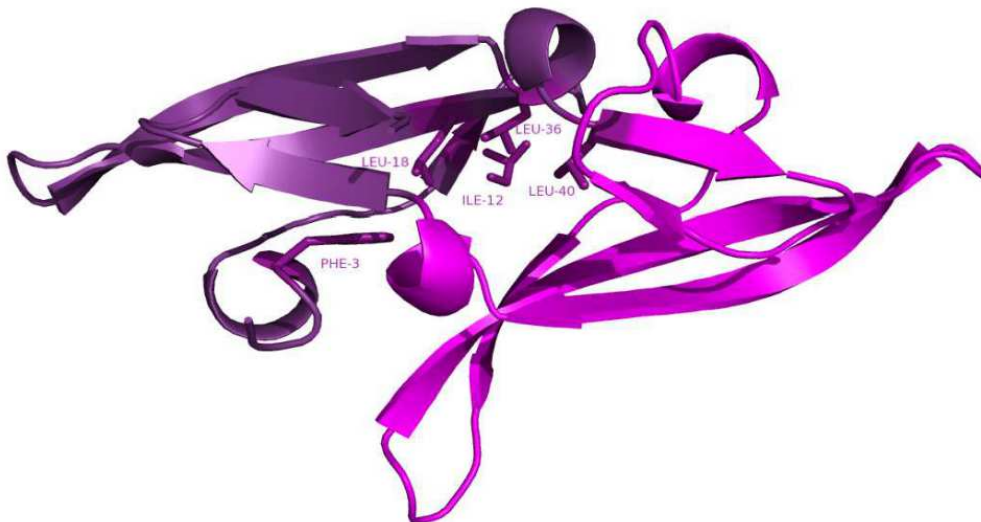
Figure 44: Cartoon representation of AFL6, where conserved amino acids of the hydrophobic core are shown as stick representations. These residues are F3, I12, L18, L36 and L40for domain A (purple) and F54, F63, L74, L92 and F103 for domain B (magenta).

One of the major differences between the bacterial CV-N and its homologs in fungi and plants are the disulfide bridges, which are formed by C8-C22 and C58-C73 in CV-N. While these disulfide bridges are conserved in CVNH from the underwater fern *Ceratopteris richardii* (CrCVNH), belonging to the same group than CV-N (group I), no disulfide bridges are found in CVNHs from groups II (TbCVNH) and III (NcCVNH). In NcCVNH, cysteines are replaced by hydrophobic residues (A7/61 and L25/82) (Koharudin et al. 2008). In AFL6, these residues are conserved at positions A7/58 and L22/78 and in both proteins, hydrophobic interactions compensate for the covalent disulfide bridges in CV-N and are thus important to maintain the structural integrity. It may be noted that AFL6 contains one cysteine residue (C39) equivalent to C40 in TbCVNH at the helical turn α2 which is replaced by isoleucine in NcCVNH.

3.4.3 Structural evaluation of putative carbohydrate binding sites

As mentioned above CV-N and its homologs possess two carbohydrate binding sites, one at each domain at opposite ends of the ellipsoid. NMR titration studies conducted on NcCVNH and TbCVNH with Manα1-2Man showed that the positions of the binding sites are equivalent to those in CV-N. In the Cyanovirin-N homologs however, only one of the binding sites from *T. borchii* (domain A) and *N. crassa* (domain B) is active and affinities are much lower (Koharudin et al. 2008). Interestingly, the chimera LKAMG binds sucrose with both domains, even though sucrose binding was only reported for TbCVNH (domain A in LKAMG) and not for NcCVNH (domain B in LKAMG) (Koharudin et al. 2009).

Structural evaluation of the putative carbohydrate binding site on domain A

In NcCVNH and TbCVNH, the binding site of domain A involves residues located at the helical turn α1, the loop that connects the β strands β2 and β3 and on the hairpin of strands β9 and β10 as shown by $^1H^{15}NHSQC$ spectroscopy (Koharudin et al. 2008). Comparison of these areas in AFL6 shows that for this protein, α1 is slightly shifted and the following β1 is much shorter allowing for more flexibility in this area. The loop between β2 and β3 is structurally well conserved in all three lectins (figure 45). The hairpin of domain A composed of β9 and β10 appears to be structurally conserved in AFL6 with a shorter β9 strand. The lack of activity in NcCVNH of this binding site may be explained by the flexibility of the first N-terminal residues which are not forming a stable α helix.
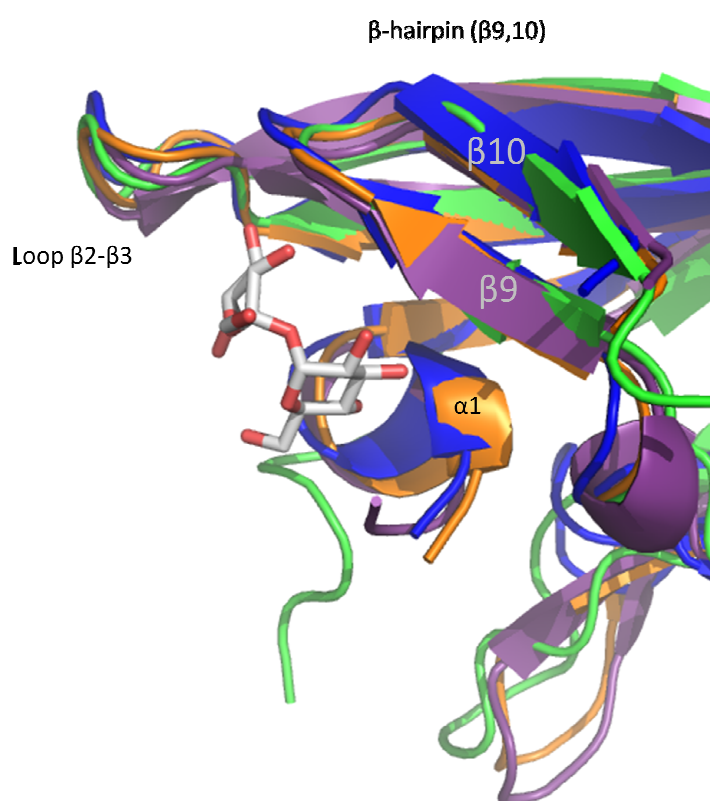
Figure 45: structural comparison of the putative binding site on domain A. Residues involved in binding are located on α1, the loop connecting β2 andβ3 and on the β hairpin comprising β9 andβ10. Cartoon representations are colored in purple, green and blue for AFL6, NcCVNH (PDB:2JZL) and TbCVNH (PDB: 2JZK) respectively. LKAMG (PDB: 3HP8) is shown in orange and its sucrose ligand in grey

Further evaluation of the amino acid that are involved in sucrose binding in the chimera LKAMG show that hydrogen bonds are predominantly formed with backbone amide and carbonyl groups. Those residues involve R24, N101, N99, and S2, which are all conserved in AFL6. However, the position of S2 is shifted in AFL6. The same shift can be found at residue T6 in AFL6, where the side chain of the equivalent S6 in LKAMG forms a hydrogen bond with sucrose. Hydrogen bonds are also formed with E 97, which corresponds to E 96 in AFL6 (Figure 46).
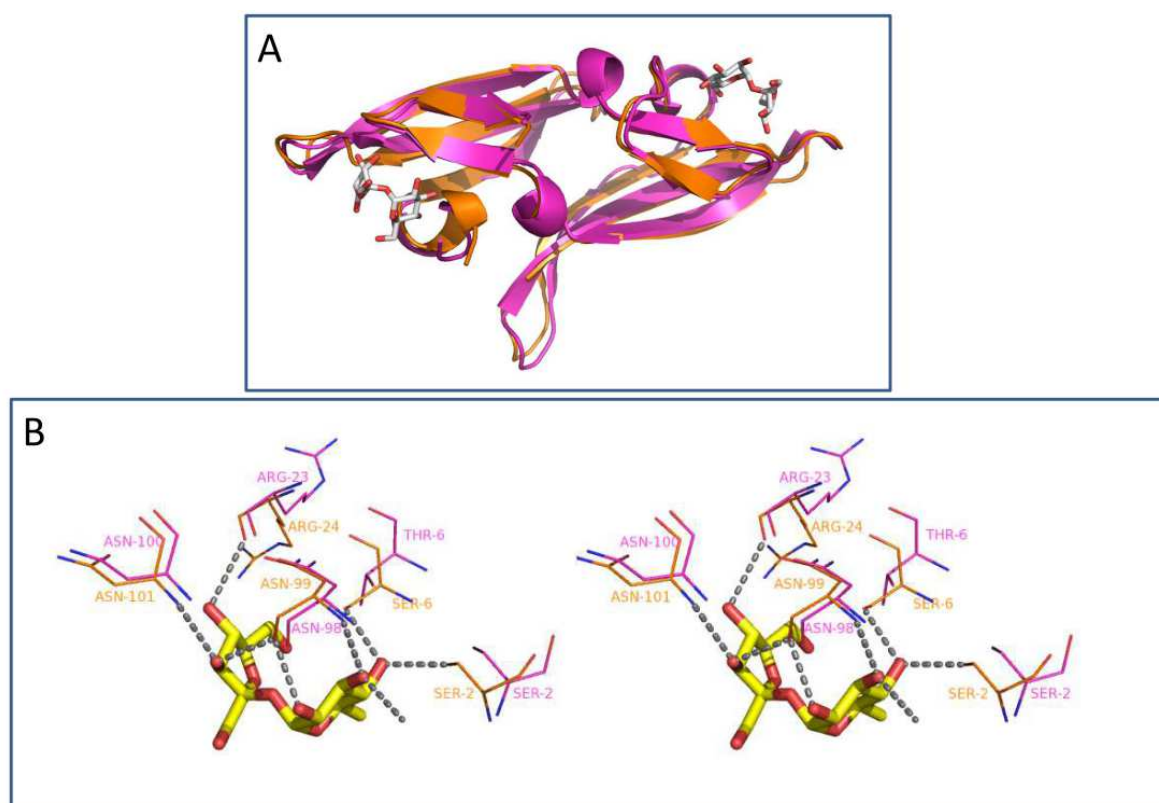
Figure 46: Structural alignment of LKAMG (PDB: 3HP8) with AFL6 (A). Stereo view of sucrose binding site on domain A of LKAMG . Hydrogen bonds (dark grey) are formed with R24, N101, N99 and S2 backbone atoms, all of which are conserved in AFL6. Side chain interaction involves residues S6 and E97. Sucrose is colored with yellow carbon atoms and LKAMG and AFL6 are colored in orange and magenta respectively.

Binding of manα1,2 man in CV-N shows that hydrogen bonds are formed with the backbone amide group of D95, which corresponds to N100 and N99 in AFL6 and LKAMG respectively. Furthermore the hydrogen bond formed by R24 in sucrose binding is conserved in CV-N, where E23 engages in hydrogen bonding with the mannoside ligand figure 47.
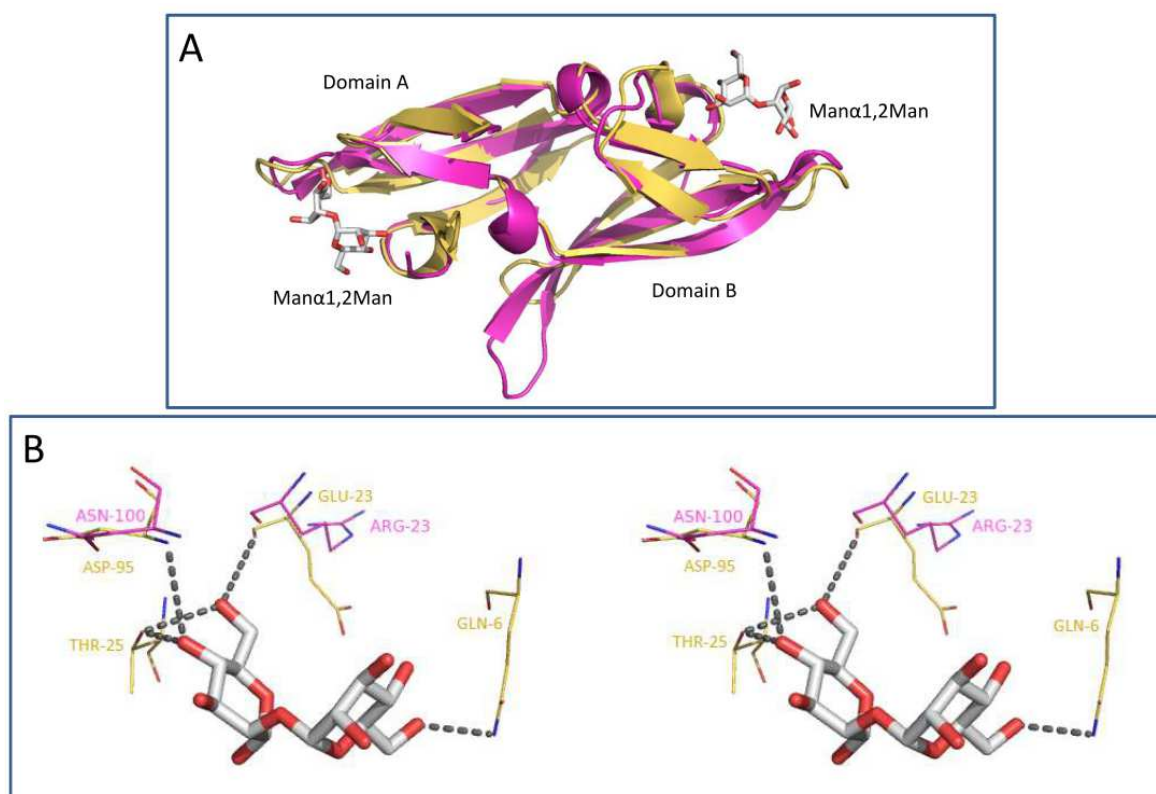
Figure 47: Structural alignment of AFL6 and CV-N (PDB: 1IIY) with manα1-2man ligand (Panel A). Stereo view of manα1-2man in binding site of domain A (Panel B). Hydrogen bonds (yellow) are formed with E 23 and D 95, both of which are conserved in AFL6. Sucrose is colored with grey carbon atoms and CV-N and AFL6 with yellow and magenta respectively.

The binding site in domain B is located on the opposite site of the pseudo symmetrical ellipsoide. Consequently, residues involved in mannoside binding in NcCVNH are located on the helical turn α3, the loop connecting β strands β7 and 8 of the three stranded sheet and on the hairpin of strands β9 and 10. Again in AFL6 the helical turn α3 is conserved but its position is shifted. The most significant structural difference in AFL6 is found at the loop connecting β7 and 8 (figure 48). While its overall orientation is in good alignment with NcCVNH, residue Q81, located at the turn of the loop is not only larger than C 85 in NcCVNH and V 77 in TbCVNH, but it is also oriented towards the inside of the binding pocked. The putative binding cleft is therefore much smaller. The hairpin comprising β4 and β5 shows high structural variability among the three CVNHs. The absence of carbohydrate binding in domain B of TbCVNH may be due to the overall disorder of the loop between β7 and 8 (Koharudin et al. 2008).
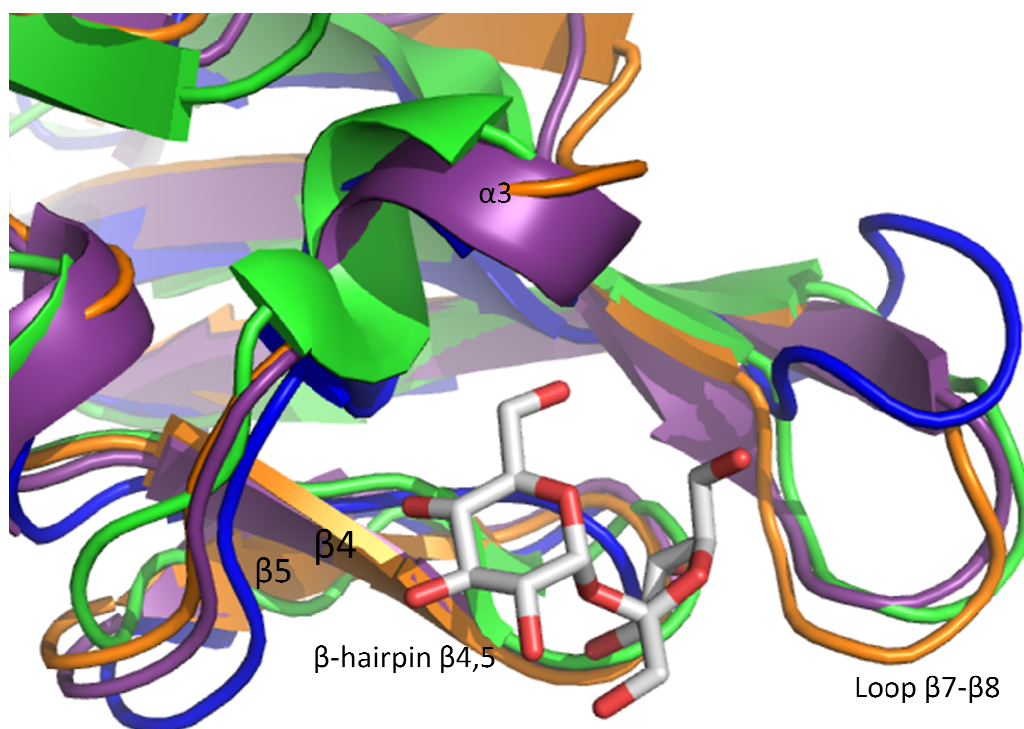
Figure 48: structural comparison of the putative binding site on domain B. Residues involved in binding are located on α3, the loop connecting β7 and β8 and on the β hairpin comprising β4 and β5. Cartoon representations are colored in purple, green and blue for AFL6, NcCVNH (PDB:2JZL) and TbCVNH (PDB: 2JZK) respectively. LKAMG (PDB: 3HP8) is shown in orange and its sucrose ligand in grey.

Asparagine, at position 41 in LKAMG is involved in H-bonding with the sucrose ligand and is well conserved in AFL6 (N41) as well as in CV-N (N42), where it involved in mannoside binding as well. Arginine 81 is substituted by glutamine in CV-N (Q78) and both are important in binding sucrose and mannoside respectively. Structural alignment shows that AFL6 contains an alanine at this position (A79). Interestingly, AFL6 contains a glutamine in position 81, which is directed towards the binding pocket resulting in a distortion of the cleft (figure 49).
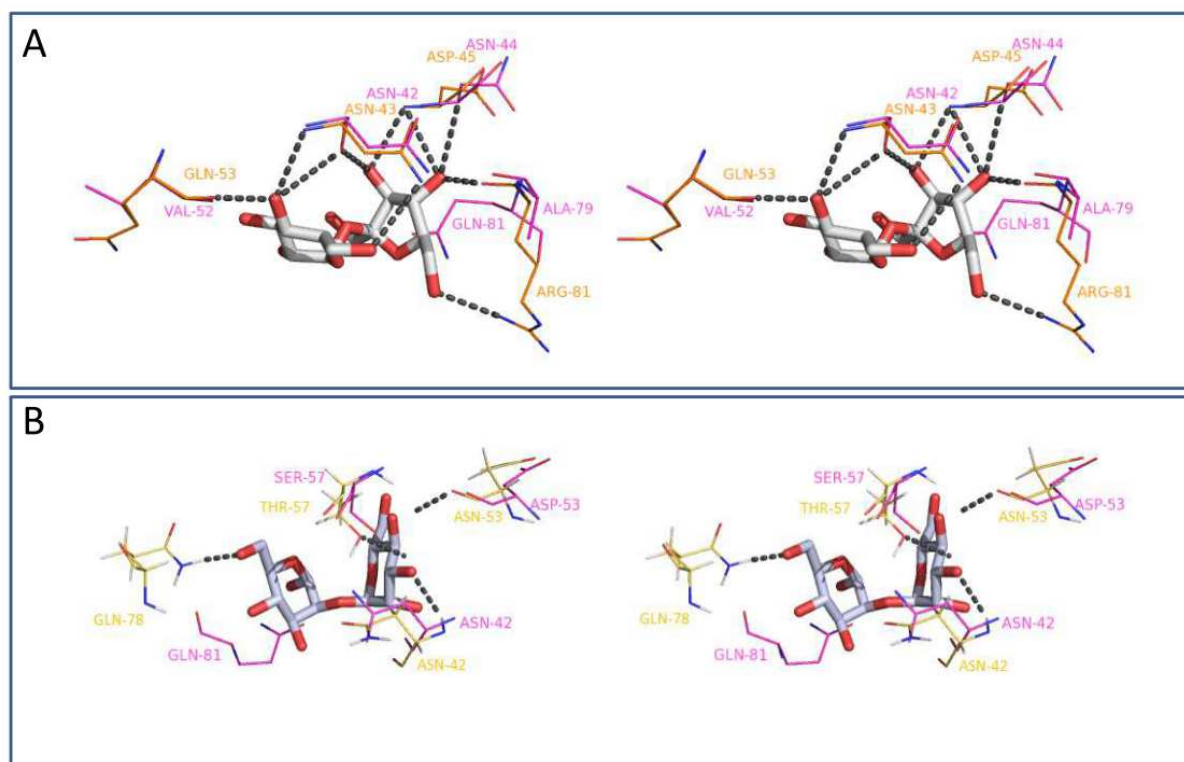
Figure 49: stereo views of binding sites of domain B with (A) sucrose in LKAMG (orange PDB: 3HP8) and (B) manα1,2man binding of CV-N (yellow, PDB: 1IIY) in alignment with AFL6 (magenta). The ligands are shown sticks colored with grey carbon atoms.

From this comparison, AFL6 may indeed present two binding sites corresponding to those observed in domain A and domain B for the other proteins in this family. The site in domain A is however more conserved than the site in domain B that presents a glutamine residue which could be involved in a steric conflict with putative carbohydrate ligand (figure 50).
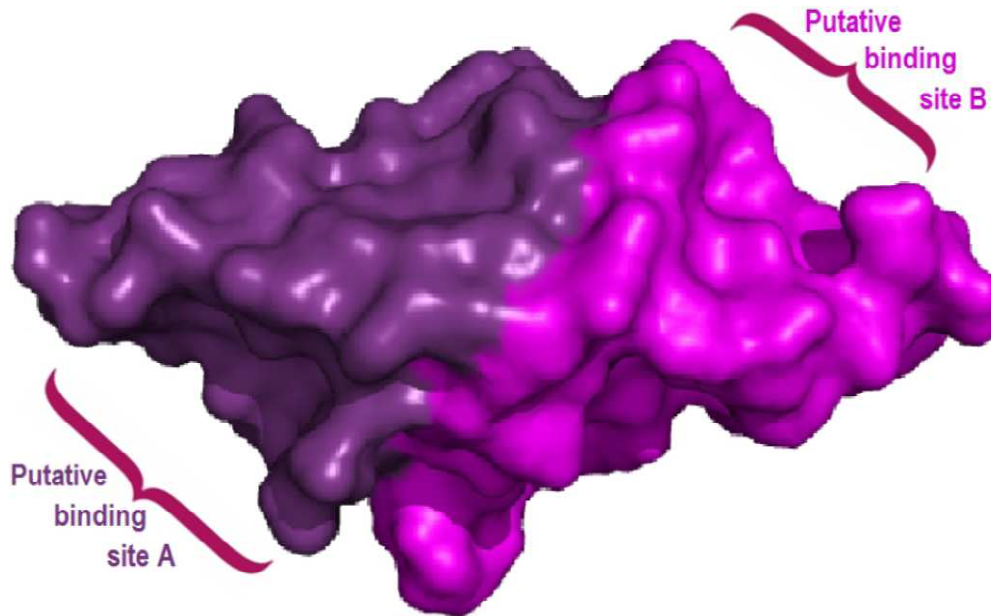
Figure 50: surface representation of AFL6. The putative carbohydrate binding sites are located at opposite ends of the ellipsoid, domains A and B are colored in purple and magenta respectively.

3.4 Carbohydrate Interactions

3.4.1 Labelling and Glycan Array

Since AFL6 is a novel lectin, its natural ligands are not known. A simple and efficient way to investigate putative ligands is the glycan array. This service is offered by the Consortium for Functional Glycomics (CFG) in the USA. To use this service the user needs to provide the lectin with a label for detection. The standard method is labeling with Alexa Fluor 488, but monomeric lectins are known to present low affinity for the glycan on the chips, and therefore low detection. It is therefore recommended to assemble such lectin in a multimeric form before testing them on the array. For example monovalent lectins can be biotinylated to form a tetramer with fluorescent streptavidin to enhance signal strength or the biotin can be detected with a fluorescent biotin antibody.

Standard *in vitro* biotinylation and Alexa Flour 488 labeling of proteins is often carried out using *N*-hydroxysuccinimid (NHS)-linked biotin which allows covalent binding to the primary amine of lysine-residues. However, in the case of AFL6, there are no lysine-residues present in the primary amino acid sequence. Therefore a lysine containing avi-tag was introduced to the N-terminus. This was achieved by simple sub-cloning of AFL6 into a pETTEVbiot vector that was available in our facilities. Cloning, expression and purification was carried out as described earlier, and the resulting AFL6avi was biotinylated and the sample was sent to the CFG for glycan array. The detection was performed with Cyan-5 labeled streptavidin.

The mammalian glycan chip version 5.2 that was used contains 609 glycans. The results show weak binding signal under 2000 rfu with high standards deviations resulting in a high noise background. Only hits with a mean fluorescence response of 700 rfu with a standard deviation smaller than 20% were considered positive hits. This results in a total number of 8 glycans that are recognized by AFL6 under these conditions. All these glycans contain a common disaccharide motif of lactosamine (Galβ1-4GlcNAc or LacNAc) in repetitive disaccharides known as polylactosamine. Interestingly, glycans are recognized with different terminal glycosylation such as Neu5Ac, Fuc, GlcNAc or Gal.
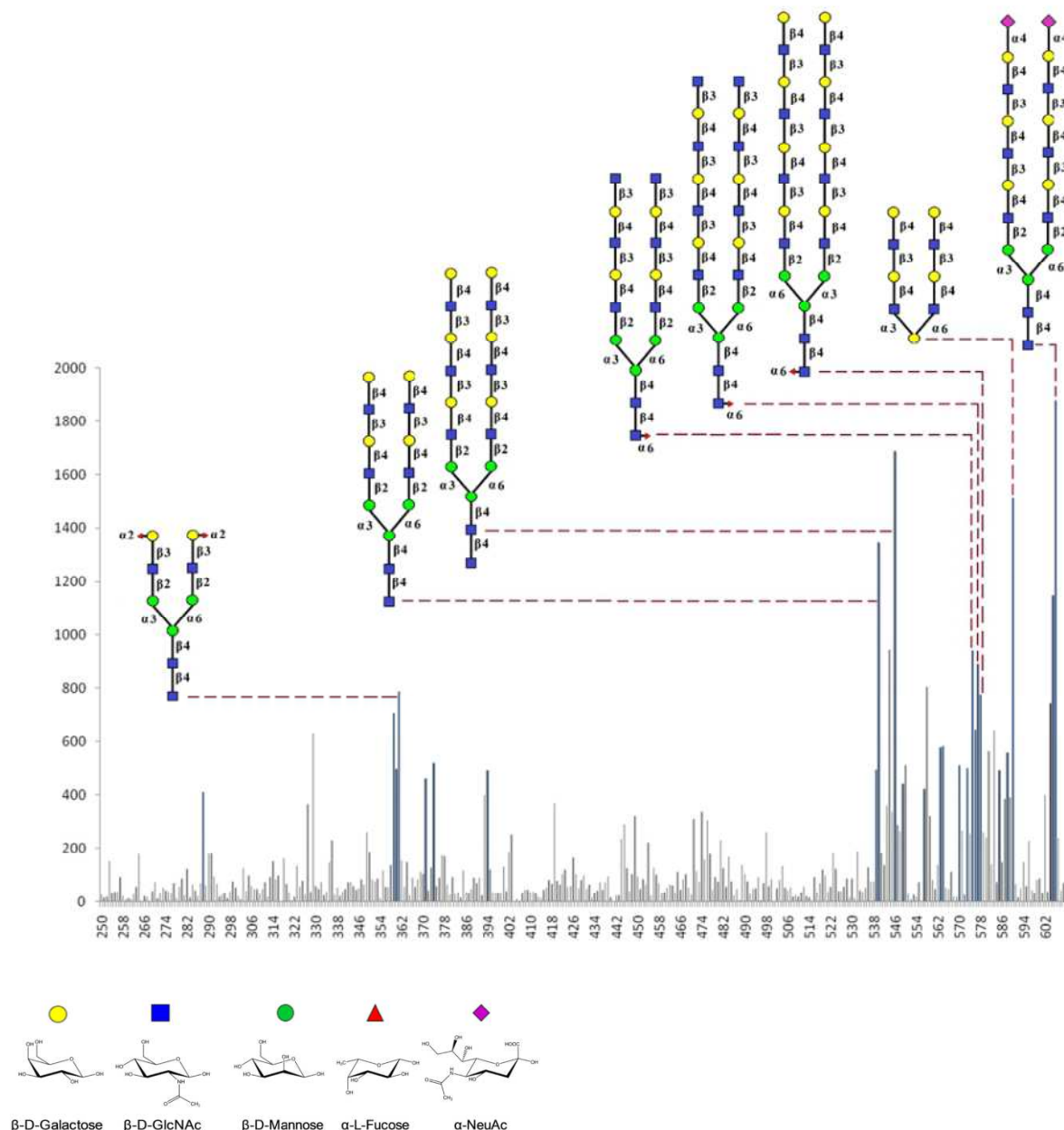
Figure 51: glycan array of biotinylated AFL6, detected with Cyan-5-labeled streptavidin. Potential hits are represented as blue columns and labeled with the corresponding glycan. For a better overview of the hits, glycan number 1- 280, where no hits were found are omitted from the graph

Conversely, binding of high mannose type glycans were not detected in the glycan array even though CV-N and the characterized CVNH show affinity for α2 mannosides. Further investigations towards the carbohydrate binding of AFL6 were carried out using isothermal titration calorimetry (ITC).

3.4.2 AFL6 binds Manα1,2Man with low affinity (ITC studies)

Given the result of the glycan array, ITC studies were conducted with the disaccharide lactosamine. However, even in 500 molar excess of the ligand, no interaction could be observed (data not shown). This may be due to the too small size of the ligand, indicating that maybe tetrasaccharide or a longer polylactosamine fragments may be necessary for binding. However, such ligands were not readily available and could not be carried out due to lack of time. The same goes for investigations using other techniques for carbohydrate binding evaluation such as ELLA or membrane based techniques. Alternatively, the binding to sucrose was investigated since the chimeric LKAMG, which shows close structural relation with AFL6 binds sucrose with affinity in millimolar range. However, sucrose binding could not be observed for AFL6 in ITC studies using 500 molar excess of the ligand.

Manα1,2Man binding

The anti HIV activity of CV-N is based on the binding of high mannose glycans on the viral envelope, more specifically the D1 and D3 arms of Man-8 and Man-9, comprising Manα1,2Man residues shown in figure 52 (Shenoy et al. 2002), (Barrientos & Gronenborn 2005).
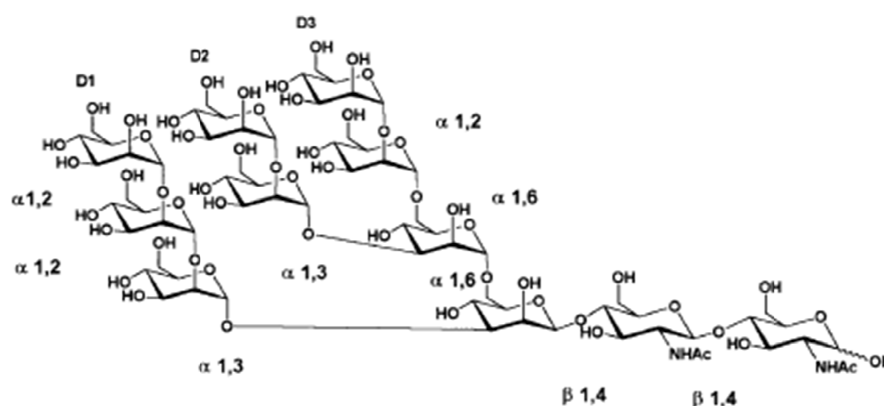
Figure 52. Carbohydrate structure of Man-9 adapted form (Shenoy et al. 2002)

Since some of the cyanovirin family members bind to α 2-mannoside with high affinity, this ligand was also tested by ITC for AFL6. ITC studies were carried out with AFL6 and a 500 fold molar excess of Manα1-2Man. Figure 53 shows the raw ITC data (A) and the integrated curve with experimental points. Because of the low affinity, an excess of ligand has been used and the stoichiometry of binding interaction cannot be measured, as it can only be calculated from a sigmoide curve, which can only be produced with higher affinities.

Therefore it is not possible to determine whether both binding sites are active in AFL6. However, agglutination assays with rabbit red blood cells were negative indicating that only one bonding site may be active.

To extract the enthalpy and calculate the affinity and entropic contributions from the hyperbolic curve, it is therefore necessary to fix the stoichiometry. The experimental data was integrated using a one side binding model (stoichiometry fixed at N=1), resulting in a calculated association constant K_a of 123 M^{-1} (dissociation constant K_d of 8 mM). The negative ΔH that dominates the binding (ΔH = -6 kcal/mol) with unfavorable entropic contributions ($T\Delta S$ = -3.2 kcal/mol) suggests that the binding is dominated by polar/ electrostatic interactions, van der Waals interactions and hydrogen bonds.
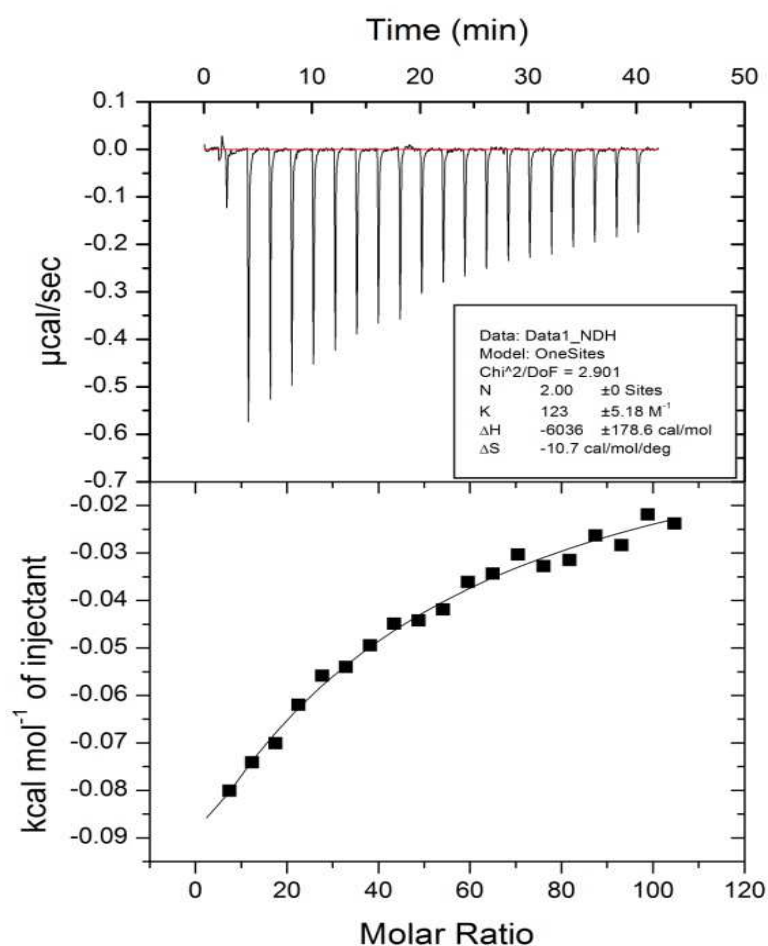
Figure 53: Calorimetric titration of AFL6 with Manα1-2Man. Top: instrument feedback necessary to maintain a constant temperature as 30 mM mannoside are periodically titrated into rapidly mixing 60 µM AFL6. These results are then converted to the binding isotherm (bottom).

ITC experiments were also conducted using α1-3 and α1-6 linked mannosides, both present in Man-9 oligosaccharide. The absence of detectable interactions indicates that the α1-2 linkage is necessary for mannoside binding.

4. Conclusion

AFL6 was identified by homology search in the genome of *A. fumigatus* and confirmed the previous report (Koharudin et al. 2008). After recombinant expression and purification, crystals diffracting to high resolution were obtained. Structural analysis showed that AFL6 shares the common fold of CV-N and other CVNH form fungi comprising two pseudo-symmetrical domains A and B. Evaluation of the putative binding sites demonstrated that amino acids that could create polar interactions between ligand and protein backbone are mostly conserved. However, hydrogen bonds to amino acid side chains were only partly conserved. Furthermore the position of glutamine 81 in AFL6 could result in disruption of the binding cleft on domain B. It is likely that the binding site on domain B may not be active for mannoside binding. This is in good accordance with other fungal CVNH in which only one binding site is active. The lack of hemagglutination activity of AFL6 supports this theory. Mannoside binding was confirmed using ITC, but the affinity was very low with a K_D of 8 mM. Interestingly, AFL6 was found to bind bi- antennary polylactosamine containing glycans in glycan array. However, lactosamine binding was not confirmed by ITC experiments. Further investigations using polylactosamine in ITC or SPR experiments are necessary to explore the carbohydrate specificity of AFL6.

Conclusion and Perspectives

The first part of my thesis comprised the purification of a sialic acid specific lectin from crude *A. fumigatus* extracts that had been reported earlier (Tronchin et al. 2002). Since our laboratory in Grenoble is not equipped to handle live *A. fumigatus*, the extracts were prepared by collaboration partners in Angers and sent on dry ice to Grenoble. The extracts showed low hemagglutination activity after thawing them on ice, which hampered the search for the lectin. The lectin was described to possess affinity for the glycoprotein fetuin (Tronchin et al. 2002). Therefore, a first purification attempt was carried out using affinity chromatography with a fetuin- coupled resin. However, no protein was found with subsequent elution with NaCl and NeuAc. This may indicate either that the sialic acid specific lectin was not present in the extracts or that it had bound too strongly to the resin and could not be eluted. We then decided to reproduce the published protocol, omitting the first step of gel filtration. Ammonium sulfate precipitation, ion exchange- and hydrophobic interaction chromatography steps were followed by hemagglutination assays to locate the lectin containing fractions. With this protocol, we were able to purify a putative lectin corresponding to the gene *Afu8g00550* and the corresponding protein was named AFL5. After recombinant expression and purification hemagglutination assays were carried out to investigate whether the identified protein was indeed a lectin. The assays showed contradictory results. Firstly, hemagglutination activity was absent when the protein was kept in TRIS buffer, but could be restored when the buffer was exchanged to PBS. Secondly, we noticed that some purification batches completely lacked hemagglutination activity even though expression and purification was carried out using the same protocol. The activity could not be restored by addition of salts like potassium chloride and calcium chloride. We hypothesized that a carbohydrate ligand may have bound to rAFL5 during expression and carried out dialysis against PBS over several days without restoring activity. Separation of a hemagglutinating fraction of rAFL5 was possible using gel filtration chromatography. rAFL5 eluted in a main peak corresponding to the monomer and a smaller peak that was found in the late eluate. Interestingly, hemagglutination activity was found only in the late peak and SDS PAGE revealed that the peak contained rAFL5. The late fraction of AFL5 may have bound to the gel filtration resin comprising dextran and agarose. Hemagglutination inhibition tests carried out with this fraction, showed a specificity for fetuin, with a minimal concentration of 625 µg ml^{-1} that

completely inhibited agglutination of the RBCs. Even though this finding was in agreement with the published results, we cannot conclude that our identified protein corresponds to the sialic acid specific lectin. The majority of recombinant AFL5 was not able to agglutinate RBC and only a fraction was found to bind to the column and show activity. Binding to the gel- filtration matrix and fetuin may therefore be considered an artifact. One could hypothesize that binding may be unspecific due to hydrogen bonding or ionic interactions on the protein surface. Structural characterization could help identify such areas. While crystals were obtained for this protein that diffracted well, the structure could not be solved due to the phase problem. SAD phasing with sulfur was unsuccessful because the anomalous signal was not strong enough. The crystals obtained with the Selenomethionin derivative need further optimization as they were too small and didn`t diffract sufficiently for MAD phasing.

AFL5 was annotated as a methyltransferase, and its gene *Afu8g00550* is part of a putative gene cluster that is involved in the biosynthesis of the secondary metabolite pseurotin A (Maiya et al. 2007). Indeed, sequence alignment studies showed that AFL5 shares homology with other putative proteins from filamentous fungi such as *Scedosproium apiospermum*, *Metarhizium aniosoplia* and *Fusarium* spp. all of which contain a Sir-N like methyltransferase domain. Sir-N like methyltransferases are known to catalyze *O*-methylation, using S-adenosyl methionine (SAM) as a donor. However, none of these SirN-like methyltrasferases from filamentous fungi have been structurally characterized and some are putative proteins. Therefore, it would be interesting to investigate the methyltransferase activity of AFL5. In *Aspergillus fumigatus* the putative methyltransferase that we named AFL5 is thought to catalyze the methylation of 8-*O*- demethylpseurotin A. Methyltransferase assays could be carried out using SAM as a donor and 8-*O*- demethylpseurotin A or azaspirene, a related compound, as an acceptor.

A second attempt of purifying the sialic acid lectin was carried with fresh extracts that were sent to us. This time, the purification steps were monitored not only with hemagglutination assays but also with immobilized fetuin in SPR. In these extracts However, neither hemagglutination activity nor binding to immobilized fetuin could be observed, even after ammonium sulphate precipitation and ion exchange chromatography. Since activity in the sent extracts was either very low or absent, one may conclude that the involved freezing and thawing process may have led to inactivation or denaturation of the lectin. Further attempts to reproduce the purification protocol should therefore be carried out on fresh samples in a laboratory that is equipped to handle *A. fumigatus*, so that exact reproduction of the

publication is possible. Characterization of this sialic acid specific lectin would be highly interesting as there are so far only two fungal lectins charccterized with specificity towards sialic acid. Both are from mushrooms (*Agrocybe cylindracea* and *Polyporus velutina*) and information about sialic acid specific lectins from opportunistic filamentous fungi are scarce.

Besides purification form natural extracts lectins can be also identified by sequence similarity. We therefore screened the genome of *A. fumigatus* for ORFs that show sequence similarity with other known fungal lectins. With simple BLAST search we were able to identify four novel lectins in *A. fumigatus*. The fact that we could not identify the sialic acid binding lectin is not surprising since the search was very limited and search templates comprised only 27 fungal lectins whose 3-dimensional structure was known. Among them, only *P. velutina* lectin and *A. cylindracea* galectin binds sialic acid. It would be interesting to extend the genome search to *A. nidulans*, *A. niger*, *A. oryzae* and other available *Aspergillus* genomes and to include other than just fungal lectins. For a sequence based search, a query database should contain a list of well known fungal, bacterial, animal, and plant lectins that could be retrieved from the CERMAV lectin database (http://lectin3d.cermav.cnrs.fr). The combined database should then be made non-redundant by clustering sequences with high sequence similarity (90% identity). Databases containing the translated ORF protein sequences of *Aspergillus* spp genomes should then be used to search for sequence similarity in the query database. The output should be further analyzed by secondary structure prediction and compared to the structural features of the query lectins. Such a search would greatly contribute to the knowledge about lectins in filamentous fungi and comparison of lectin occurrence between pathogenic and non- pathogenic *Aspergillus* species would shed light into the role of lectins in virulence and provide information about possible drug targets.

With the simple BLAST search we carried out, we identified an R-type and two L- type lectins, both of which are unlikely to be involved in adhesion. We therefore decided to concentrate our studies on the cyanovirin-N homolog, that we named AFL6. Given the fact that the biological function of CVNHs is still unknown it would be interesting to create an *A. fumigatus* AFL6 knockout strain and investigate its pathology in an animal model. Transkriptome sequencing, carried out at the Pasteur Institute in Paris (personal communication with Annabelle Varrot) has shown that AFL6 is expressed in resting, swollen and germinating conidida and expression increases over time. Furthermore, its expression is

elevated in biofilms, indicating an important biological role in germination and possibly biofilm formation.

AFL6 was recombinantly expressed and purified and crystals were obtained, that diffracted to 1.3 Å. The structure was solved using molecular replacement. Our structural characterization demonstrated that AFL6 shares the common fold of CVNHs that can be described as elongated ellipsoid comprising two pseudosymmetric domains A and B. Domains are formed by strand exchange between two tandem repeats. Each domain contains one putative carbohydrate binding pockets at opposite ends of the ellipsoid. The putative binding cleft on domain B of AFL6 contains a glutamine residue that protrudes into the binding pocket in the crystal structure and may pose a steric hindrance for carbohydrate ligands. We conclude therefore that only the binding site on domain A may be active. This theory is further supported by the lack of hemagglutination activity. Indeed, other fungal CVNH demonstate carbohydrate binding at only one of their domains and lack hemagglutination activity. CV-N on the other hand, is able to agglutinate RBCs and binds α2-mannosides on both domains. Co- crystallization trials with manα1-2man where carried out to allow detailed structural characterization of the putative carbohydrate binding of AFL6. However, only thin needles were obtained which needed further optimization. These co- crystallization trials would identify the active binding site and amino acid residues involved in binding. Furthermore, it would be highly interesting to see the possible flexibility of the glutamine side chain, which protrudes into the binding site in our crystallographic model. Questions as to whether this residue is responsible for the low affinity or completely abolishes binding at domain B could thus be answered.

The specificity for α2- mannosides is conserved in the CVNHs, characterized so far, but their affinities are lower than that of CV-N. However, TbCVNH showed specificity for sucrose as well. An artificial chimera, comprising the active domains from TbCVNH and NcCVNH binds sucrose with both domains. Interestingly, this chimera worked best to solve the structure of AFL6 by molecular replacement. In spite of the structural similarity AFL6 was not found to bind sucrose in ITC studies using 500 molar excess of ligand. Although AFL6 binds manα1-2man like other CVNHs and CV-N itself, it shows much lower affinity. Glycan array revealed that AFL6 bind to bi- antennary polylactosamine ((Galβ1-4GlcNAc)$_{2-4}$ or (LacNAc)$_{2-4}$). However, binding of LacNAc could not be observed in ITC studies with 500 molar excess of the ligand.

Binding to α2-mannosides was not found in glycan array which may be due to the high noise to signal ratio and the overall low fluorescent response that was observed. It may be helpful to create an artificial tetramer by using biotinylated AFL6 and incubate it with fluorescently labelled streptavidin before the glycan array. In addition, molecular docking studies of glycan databases could shed light into the carbohydrate specificity of AFL6. The simulations should be flexible-receptor docking studies, to evaluate our hypothesis concerning the presence of a binding-site in only domain A and not domain B; notably, different conformations of the glutamine side-chain in the binding-site of domain B should be generated and used for different receptor conformations.

Furthermore other screening methods could be exploited to find carbohydrate ligands with higher affinity using ELLA tests, or SPR. AFL6 could be coupled to a SPR chip and differend glycans could be tested for binding.

Besides the here characterized AFL6 it would be interesting to study the type II CVNH named pAFL8. So far, none of the type two CVNH has been structurally characterized and the N –terminal domain, shows similarity with proteins involved in the ESCRT pathway. It was shown that members of the ESCRT family may contribute to pathogenesis in *Candida albicans* (Wolf et al. 2010). pAFL8 with a dual function may therefore be worth investigating and it may play a role in virulence of *A. fumigatus*, by enabling adaptation to the host environment and may even turn out to be a valuable drug target.

Bibliography

Abhinav, K., 2013. Identification of mycobacterial lectins from genomic data. *... : Structure, Function, and ...*, 81(4), pp.644–657.

Aimanianda, V. et al., 2009. Surface hydrophobin prevents immune recognition of airborne fungal spores. *Nature*, 460(7259), pp.1117–21.

Alcazar-Fuoli, L. & Mellado, E., 2014. Current status of antifungal resistance and its impact on clinical practice. *British journal of haematology*, (April), pp.471–484.

Altschul, S. & Madden, T., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids ...*, 25(17), pp.3389–3402.

Armstead, J., Morris, J. & Denning, D.W., 2014. Multi-country estimate of different manifestations of aspergillosis in cystic fibrosis. *PloS one*, 9(6), p.e98502.

Arnaud, M.B. et al., 2012. The Aspergillus Genome Database (AspGD): recent developments in comprehensive multispecies curation, comparative genomics and community resources. *Nucleic acids research*, 40(Database issue), pp.D653–9..

Balloy, V. & Chignard, M., 2009. The innate immune response to Aspergillus fumigatus. *Microbes and infection / Institut Pasteur*, 11(12), pp.919–27..

Banerjee, S.S. & Chen, D.-H., 2007. Glucose-Grafted Gum Arabic Modified Magnetic Nanoparticles: Preparation and Specific Interaction with Concanavalin A. *Chemistry of Materials*, 19(15), pp.3667–3672.

Barasch, J. & Al-awqati, Q., 1993. Defective acidification of the biosynthetic pathway in cystic fibrosis Departments of Medicine and Physiology , College of Physicians and Surgeons of Columbia University , 630 W 168th St , New. *Journal of Cell Science*, 233(Suppl 17), pp.229–233.

Barrientos, L.G. et al., 2003. Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral research*, 58(1), pp.47–56.

Barrientos, L.G. et al., 2002. The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure (London, England : 1993)*, 10(5), pp.673–86.

Barrientos, L.G. & Gronenborn, A.M., 2005. The highly specific carbohydrate-binding protein cyanovirin-N: structure, anti-HIV/Ebola activity and possibilities for therapy. *Mini reviews in medicinal chemistry*, 5(1), pp.21–31.

Beauvais, A. et al., 2007. An extracellular matrix glues together the aerial-grown hyphae of Aspergillus fumigatus. *Cellular microbiology*, 9(6), pp.1588–600.

Beauvais, A. et al., 2014. Aspergillus Cell Wall and Biofilm. *Mycopathologia*, pp.0–6.

Beauvais, A. et al., 2013. Deletion of the α-(1,3)-glucan synthase genes induces a restructuring of the conidial cell wall responsible for the avirulence of Aspergillus fumigatus. *PLoS pathogens*, 9(11), p.e1003716.

Bennett, J., 1842. XVII. On the Parasitic Vegetable Structures found growing in Living Animals. *Transactions of the Royal Society of Edinburgh*, x(August 1841), pp.277–294..

Bernstein, F.C. et al., 1977. The protein data bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112(3), pp.535–542..

Bewley, C. a et al., 1998. Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nature structural biology*, 5(7), pp.571–8.

Bewley, C. a & Otero-Quintero, S., 2001. The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120. *Journal of the American Chemical Society*, 123(17), pp.3892–902.

Bewley, C. a., Kiyonaka, S. & Hamachi, I., 2002. Site-specific Discrimination by Cyanovirin-N for α-Linked Trisaccharides Comprising the Three Arms of Man8 and Man9. *Journal of Molecular Biology*, 322(4), pp.881–889.

Bhowal, J., Guha, A.K. & Chatterjee, B.P., 2005. Purification and molecular characterization of a sialic acid specific lectin from the phytopathogenic fungus Macrophomina phaseolina. *Carbohydrate research*, 340(12), pp.1973–82.

Boi, C. et al., 2006. Adsorption of lectins on affinity membranes. *Journal of Membrane Science*, 273(1-2), pp.12–19.

Bolmstedt, a J. et al., 2001. Cyanovirin-N defines a new class of antiviral agent targeting N-linked, high-mannose glycans in an oligosaccharide-specific manner. *Molecular pharmacology*, 59(5), pp.949–54..

Bonfield, T.L. et al., 2012. Absence of the cystic fibrosis transmembrane regulator (Cftr) from myeloid-derived cells slows resolution of inflammation and infection. *Journal of leukocyte biology*, 92(5), pp.1111–22..

Botos, I. et al., 2002. Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *The Journal of biological chemistry*, 277(37), pp.34336–42.

Bouchara, J.P. et al., 1997. Sialic acid-dependent recognition of laminin and fibrinogen by Aspergillus fumigatus conidia. *Infection and immunity*, 65(7), pp.2717–24..

Bourgeois, C. & Kuchler, K., 2012. Fungal pathogens-a sweet and sour treat for toll-like receptors. *Frontiers in cellular and infection microbiology*, 2(November), p.142.

Boyd, M.R. et al., 1997. Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrobial agents and chemotherapy*, 41(7), pp.1521–30.

Boyd, W.C. & Shapleigh, E., 1954. Specific Precipitating Activity of Plant Agglutinins (Lectins). *Science (New York, N.Y.)*, 119(3091), p.419.

Bozza, S. et al., 2009. Immune sensing of Aspergillus fumigatus proteins, glycolipids, and polysaccharides and the impact on Th immunity and vaccination. *Journal of immunology (Baltimore, Md. : 1950)*, 183(4), pp.2407–14.

Van Breedam, W. et al., 2014. Bitter-sweet symphony: glycan-lectin interactions in virus biology. *FEMS microbiology reviews*, 38(4), pp.598–632.

Brennan, S., 2008. Innate immune activation and cystic fibrosis. *Paediatric respiratory reviews*, 9(4), pp.271–9; quiz 279–80.

Brichacek, B. et al., 2013. In vivo evaluation of safety and toxicity of a Lactobacillus jensenii producing modified cyanovirin-N in a rhesus macaque vaginal challenge model. *PloS one*, 8(11), p.e78817..

Brummer, E. & Stevens, D. a, 2010. Collectins and fungal pathogens: roles of surfactant proteins and mannose binding lectin in host resistance. *Medical mycology*, 48(1), pp.16–28.

Chabre, Y.M. et al., 2011. Combining glycomimetic and multivalent strategies toward designing potent bacterial lectin inhibitors. *Chemistry (Weinheim an der Bergstrasse, Germany)*, 17(23), pp.6545–62.

Chotirmall, S.H., Greene, C.M. & McElvaney, N.G., 2010. Candida species in cystic fibrosis : A road less travelled. *Medical mycology*, 48(Suppl1), pp.114–124.

Chotirmall, S.H. & McElvaney, N.G., 2014. Fungi in the cystic fibrosis lung: bystanders or pathogens? *The international journal of biochemistry & cell biology*, 52, pp.161–73..

Colleluori, D.M. et al., 2005. Expression, purification, and characterization of recombinant cyanovirin-N for vaginal anti-HIV microbicide development. *Protein expression and purification*, 39(2), pp.229–36..

Cooper, M., 2002. Optical biosensors in drug discovery. *Nature Reviews Drug Discovery*, 528(July), p.2002.

Coughlan, C. a et al., 2012. The effect of Aspergillus fumigatus infection on vitamin D receptor expression in cystic fibrosis. *American journal of respiratory and critical care medicine*, 186(10), pp.999–1007.

Coulot, P. et al., 1994. Specific interaction of Aspergillus fumigatus with fibrinogen and its role in cell adhesion. *Infection and immunity*, 62(6), pp.2169–77.

Coutinho, H.D.M., Falcão-Silva, V.S. & Gonçalves, G.F., 2008. Pulmonary bacterial pathogens in cystic fibrosis patients and antibiotic therapy: a tool for the health workers. *International archives of medicine*, 1(1), p.24.

Dam, T. & Brewer, C., 2002. Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chemical reviews*, 102(718), pp.387–429.

Davis, P.B., 2006. Cystic fibrosis since 1938. *American journal of respiratory and critical care medicine*, 173(5), pp.475–82.

Davril, M. et al., 1999. The sialylation of bronchial mucins secreted by patients suffering from cystic fibrosis or from chronic bronchitis is related to the severity of airway infection. *Glycobiology*, 9(3), pp.311–321..

Delhaes, L. et al., 2012. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community--implications for therapeutic management. *PloS one*, 7(4), p.e36313.

Desh Deepak Singh, B.S.P. et al., 2007. Scanning the Genome of Mycobacterium tuberculosis to Identify Potential Lectins. *Protein & Peptide Letters*, 14(7), pp.683–691.

Dey, B. et al., 2000. Multiple antiviral activities of cyanovirin-N: blocking of human immunodeficiency virus type 1 gp120 interaction with CD4 and coreceptor and inhibition of diverse enveloped viruses. *Journal of virology*, 74(10), pp.4562–9.

Díaz, C., Melchers, L. & Hooykaas, P., 1989. Root lectin as a determinant of host–plant specificity in the Rhizobium–legume symbiosis. *Nature*, 338(April), p.1989.

Dupeux, F. et al., 2011. A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta crystallographica. Section D, Biological crystallography*, 67(Pt 11), pp.915–9.

Emsley, P. & Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 12 Pt 1), pp.2126–32..

Esser, M.T. et al., 1999. Cyanovirin-N binds to gp120 to interfere with CD4-dependent human immunodeficiency virus type 1 virion binding, fusion, and infectivity but does

not affect the CD4 binding site on gp120 or soluble CD4-induced conformational changes in gp120. *Journal of virology*, 73(5), pp.4360–71.

Evans, P., 2006. Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 1), pp.72–82..

Evans, P.R. & Murshudov, G.N., 2013. How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography*, 69(Pt 7), pp.1204–14..

Ferec, C. & Cutting, G.R., 2012. Assessing the Disease-Liability of Mutations in CFTR. *Cold Spring Harbor perspectives in medicine*, 2(12), p.a009480..

Free, S., 2013. Fungal cell wall organization and biosynthesis. *Advances in genetics*, 81, pp.33–82..

Frizzell, R. a & Pilewski, J.M., 2004. Finally, mice with CF lung disease. *Nature medicine*, 10(5), pp.452–4.

Fujimoto, Y.K. & Green, D.F., 2012. Carbohydrate recognition by the antiviral lectin cyanovirin-N. *Journal of the American Chemical Society*, 134(48), pp.19639–51..

Fujimoto, Z., Tateno, H. & Hirabayashi, J., 2014. Lectin Structures : Classification Based on 3D structures. In *Lectins: Methods and Protocols*. pp. 579–605.

Gabius, H.-J. et al., 2011. From lectin structure to functional glycomics: principles of the sugar code. *Trends in biochemical sciences*, 36(6), pp.298–313.

Gallegos, B. et al., 2014. Lectins in human pathogenic fungi. *Revista iberoamericana …*, 31(1), pp.72–75.

Gao, X. et al., 2009. Soluble cytoplasmic expression, rapid purification, and characterization of cyanovirin-N as a His-SUMO fusion. *Applied Microbiology and Biotechnology*, 85(4), pp.1051–1060.

Gil, M.L. et al., 1996. Binding of extracellular matrix proteins to Aspergillus fumigatus conidia. *Infection and immunity*, 64(12), pp.5239–47.

Giomarelli, B. et al., 2002. The microbicide cyanovirin-N expressed on the surface of commensal bacterium Streptococcus gordonii captures HIV-1. *AIDS*, 16, pp.1351–1356.

Glick, M.C. et al., 2001. Activity of fucosyltransferases and altered glycosylation in cystic fibrosis airway epithelial cells. *Biochimie*, 83(8), pp.743–7.

Goldstein, I.J. & Agrawal, B.B.L., 1965. Specific Binding of Concanavalin A to CrossßLinked Dextran Gels. , 96(1965), pp.23–25.

Grahn, E. et al., 2007. Crystal structure of the Marasmius oreades mushroom lectin in complex with a xenotransplantation epitope. *Journal of molecular biology*, 369(3), pp.710–21.

Guss, A.M. et al., 2011. Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *The ISME journal*, 5(1), pp.20–9.

Gustafsson, J.K. et al., 2012. Bicarbonate and functional CFTR channel are required for proper mucin secretion and link cystic fibrosis with its mucus phenotype. *The Journal of experimental medicine*, 209(7), pp.1263–72..

Hamilton, A. & Jeavons, L., 1998. Sialic Acid-Dependent Recognition of Laminin byPenicillium marneffei Conidia. *Infection and ...*, 66(12).

Henne, W.M., Stenmark, H. & Emr, S.D., 2013. Molecular mechanisms of the membrane sculpting ESCRT pathway. *Cold Spring Harbor perspectives in biology*, 5(9).

Henry, C., Latgé, J.-P. & Beauvais, A., 2012. α1,3 glucans are dispensable in Aspergillus fumigatus. *Eukaryotic cell*, 11(1), pp.26–9.

Hoffman, L. et al., 2006. Selection for Staphylococcus aureus small-colony variants due to growth in the presence of Pseudomonas aeruginosa. *Proceedings of the ...*, 103(52), pp.19890–19895..

Hoo, A.-F. et al., 2012. Lung function is abnormal in 3-month-old infants with cystic fibrosis diagnosed by newborn screening. *Thorax*, 67(10), pp.874–81.

Houben, K. et al., 2007. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase- nucleocapsid interaction with paramyxovirus familiy. *J virol*, 81(13), pp.6807–16.

Houser, J. et al., 2013. A soluble fucose-specific lectin from Aspergillus fumigatus conidia-- structure, specificity and possible role in fungal pathogenicity. *PloS one*, 8(12), p.e83077.

Houser, J. et al., 2015. Structural insights into Aspergillus fumigatus lectin specificity: AFL binding sites are functionally non-equivalent. *Acta crystallographica. Section D, Biological crystallography*, 71(Pt 3), pp.442–53.

Howard, S.J. & Arendrup, M.C., 2011. Acquired antifungal drug resistance in Aspergillus fumigatus: epidemiology and detection. *Medical mycology*, 49 Suppl 1(April), pp.S90–5.

Hubbuch, J. et al., 2001. High gradient magnetic separation versus expanded bed adsorption: a first principle comparison. *Bioseparation*, 10(1-3), pp.99–112.

Ielasi, F., Decanniere, K. & Willaert, R., 2012. The epithelial adhesin 1 (Epa1p) from the human-pathogenic yeast Candida glabrata: structural and functional study of the carbohydrate-binding domain. *Acta Crystallographica Section …*, D68, pp.210–217.

Imberty, A., Chabre, Y.M. & Roy, R., 2008. Glycomimetics and glycodendrimers as high affinity microbial anti-adhesins. *Chemistry (Weinheim an der Bergstrasse, Germany)*, 14(25), pp.7490–9.

Imberty, A., Mitchell, E.P. & Wimmerová, M., 2005. Structural basis of high-affinity glycan recognition by bacterial and fungal lectins. *Current opinion in structural biology*, 15(5), pp.525–34.

Imberty, A. & Varrot, A., 2008. Microbial recognition of human cell surface glycoconjugates. *Current opinion in structural biology*, 18(5), pp.567–76.

Ishida, H.I. et al., 2002. Molecular Cloning and Overexpression of fleA Gene Encoding a Fucose-specific Lectin of Aspergillus oryzae. *BiosciBiotechnolBiochem*, 66(5), pp.1002–1008.

Kabsch, W., 2010. Xds. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2), pp.125–32.

Kadirvelraj, R. et al., 2011. Structure and binding analysis of Polyporus squamosus lectin in complex with the Neu5Ac{alpha}2-6Gal{beta}1-4GlcNAc human-type influenza receptor. *Glycobiology*, 21(7), pp.973–84.

Kahn, F., Kahn, M. & Islam, 2011. Khan and Khan 2011 fungal lectins.pdf. *Internation journal of biological chemistry*, 5(1), pp.1–20.

Kantardjieff, K.A. & Rupp, B., 2003. Matthews coefficient probabilities : Improved estimates for unit cell contents of proteins , DNA , and protein – nucleic acid complex crystals. *Protein Science*, 12, pp.1865–1871.

Kawagishi, H. et al., 1994. A sialic acid-binding lectin from the mushroom Hericium erinaceum. *FEBS letters*, 340(1-2), pp.56–8.

Keeffe, J.R. et al., 2011. Designed oligomers of cyanovirin-N show enhanced HIV neutralization. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34), pp.14079–84.

Kehr, J.-C. et al., 2006. A mannan binding lectin is involved in cell-cell attachment in a toxic strain of Microcystis aeruginosa. *Molecular microbiology*, 59(3), pp.893–906.

Knutsen, A.P. & Slavin, R.G., 2011. Allergic bronchopulmonary aspergillosis in asthma and cystic fibrosis. *Clinical & developmental immunology*, 2011, p.843763.

Koharudin, L.M.I. et al., 2011. Structure-function analysis of a CVNH-LysM lectin expressed during plant infection by the rice blast fungus Magnaporthe oryzae. *Structure (London, England : 1993)*, 19(5), pp.662–74.

Koharudin, L.M.I. et al., 2008. The evolutionarily conserved family of cyanovirin-N homologs: structures and carbohydrate specificity. *Structure (London, England : 1993)*, 16(4), pp.570–84..

Koharudin, L.M.I., Furey, W. & Gronenborn, A.M., 2009. A designed chimeric cyanovirin-N homolog lectin: structure and molecular basis of sucrose binding. *Proteins*, 77(4), pp.904–15..

Kondori, N. et al., 2014. Development of IgG antibodies to Exophiala dermatitidis is associated with inflammatory responses in patients with cystic fibrosis. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*, 13(4), pp.391–9.

Kopito, R.R., 1999. Biosynthesis and degradation of CFTR. *Physiological reviews*, 79(1 Suppl), pp.S167–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9922380.

Kreda, S., Davis, C. & Rose, M., 2012. CFTR, mucins, and mucus obstruction in cystic fibrosis. *Cold Spring …*, 2, pp.1–32.

Kuboi, S. et al., 2013. Molecular characterization of AfuFleA, an L-fucose-specific lectin from Aspergillus fumigatus. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*, 19(6), pp.1021–8.

Lamoth, F., Rubino, I. & Bochud, P.-Y., 2011. Immunogenetics of invasive aspergillosis. *Medical mycology*, 49 Suppl 1(April), pp.S125–36.

Lannoo, N. & Van Damme, E.J.M., 2014. Lectin domains at the frontiers of plant defense. *Frontiers in plant science*, 5(August), p.397.

Larroque, M. et al., 2012. The unique architecture and function of cellulose-interacting proteins in oomycetes revealed by genomic and structural analyses. *BMC genomics*, 13, p.605.

Latgé, J., 1999. Aspergillus fumigatus and aspergillosis. *Clinical microbiology reviews*, 12(2), pp.310–350.

Lazarescu, R.E. & Vinelli, M., 2014. Dissemination of invasive aspergillosis: diagnostic and management dilemmas. *BMJ case reports*, 2014, p.204642.

Leslie, A. & Powell, H., 2007. Processing diffraction data with Mosflm. *Evolving methods for macromolecular ...*, 245, pp.41–51.

Levdansky, E. et al., 2010. The Aspergillus fumigatus cspA gene encoding a repeat-rich cell wall protein is important for normal conidial cell wall architecture and interaction with host cells. *Eukaryotic cell*, 9(9), pp.1403–15.

Liu, J.C., Modha, D.E. & Gaillard, E. a, 2013. What is the clinical significance of filamentous fungi positive sputum cultures in patients with cystic fibrosis? *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*, 12(3), pp.187–93.

Loussert, C. et al., 2010. In vivo biofilm composition of Aspergillus fumigatus. *Cellular microbiology*, 12(3), pp.405–10.

Maiya, S. et al., 2007. Identification of a hybrid PKS/NRPS required for pseurotin A biosynthesis in the human pathogen Aspergillus fumigatus. *Chembiochem : a European journal of chemical biology*, 8(14), pp.1736–43..

Mancheño, J.M. et al., 2005. Structural analysis of the Laetiporus sulphureus hemolytic pore-forming lectin in complex with sugars. *The Journal of biological chemistry*, 280(17), pp.17251–9.

Marchler-Bauer, A. et al., 2005. CDD: a Conserved Domain Database for protein classification. *Nucleic acids research*, 33(Database issue), pp.D192–6.

Matei, E. et al., 2011. NMR solution Structure of a Cyanovirin Homolog from Wheat Head Blight Fungus. *Proteins*, 79(5), pp.1538–1549.

McCoy, A.J. et al., 2007. Phaser crystallographic software. *Journal of applied crystallography*, 40(Pt 4), pp.658–674.

Mori, T. et al., 2002. Functional homologs of cyanovirin-N amenable to mass production in prokaryotic and eukaryotic hosts. , 26, pp.42–49.

Mori, T. et al., 1998. Recombinant production of cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. *Protein expression and purification*, 12(2), pp.151–8.

Murshudov, G.N. et al., 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography*, 67(Pt 4), pp.355–67.

Muszkieta, L. et al., 2013. Investigation of Aspergillus fumigatus biofilm formation by various "omics" approaches. *Frontiers in microbiology*, 4(February), p.13.

Nascimento, K.S. et al., 2012. An overview of lectins purification strategies. *Journal of molecular recognition : JMR*, 25(11), pp.527–41.

Nierman, W., Pain, A. & Anderson, M., 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. *Nature*, 438(7071), pp.1151–6.

Parize, P. et al., 2014. Impact of Scedosporium apiospermum complex seroprevalence in patients with cystic fibrosis. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*.

Park, S.J. & Mehrad, B., 2009. Innate immunity to Aspergillus species. *Clinical microbiology reviews*, 22(4), pp.535–51.

Peñalver, M.C. et al., 1996. Binding of human fibronectin to Aspergillus fumigatus conidia. *Infection and immunity*, 64(4), pp.1146–53.

Percudani, R., Montanini, B. & Ottonello, S., 2005. The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins*, 60(4), pp.670–8.

Philippe, B. & Ibrahim-Granet, O., 2003. Killing of Aspergillus fumigatus by alveolar macrophages is mediated by reactive oxidant intermediates. *Infection and …*, 71(6), pp.3034–3042.

Pihet, M. et al., 2009. Occurrence and relevance of filamentous fungi in respiratory secretions of patients with cystic fibrosis--a review. *Medical mycology*, 47(4), pp.387–97.].

Pohleven, J. et al., 2012. Bivalent carbohydrate binding is required for biological activity of Clitocybe nebularis lectin (CNL), the N,N'-diacetyllactosediamine (GalNAcβ1-

4GlcNAc, LacdiNAc)-specific lectin from basidiomycete C. nebularis. *The Journal of biological chemistry*, 287(13), pp.10602–12..

Rademacher, C. et al., 2012. A Siglec-like sialic-acid-binding motif revealed in an adenovirus capsid protein. *Glycobiology*, 22(8), pp.1086–91.

Raman, R. et al., 2005. Glycomics : an integrated relationships of glycans systems approach to structure-funtion relationships of glycans. *Nature methods*, 824(2), p.2005.

Rankin, N.E., 1953. Dissiminated Aspergillosis and Moniliasis associated with Agranulocytosis and antibiotic therapy. *British medical journal*, pp.918–919.

Rhim, a D. et al., 2001. Terminal glycosylation in cystic fibrosis (CF): a review emphasizing the airway epithelial cell. *Glycoconjugate journal*, 18(9), pp.649–59..

Rhim, A.D. et al., 2004. Altered terminal glycosylation and the pathophysiology of CF lung disease. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*, 3 Suppl 2, pp.95–6.

Roussel, D., Lamblin, G. & Degand, P., 1975. Heterogeneity of the carbohydrate chains of sulfated bronchial glycoproteins Isolated from a Patient suffering from Cystic Fibrosis. *The Journal of biological chemistry*, 250(6), pp.2114–2122.

Rubin, B.K., 2007. Mucus structure and properties in cystic fibrosis. *Paediatric respiratory reviews*, 8(1), pp.4–7.

Rüdiger, H. & Gabius, H., 2011. Breaking the Sugar Code : Six Levels of Affinity Regulation in Glycan-lectin Interaction. , pp.11–28.

Sales-Campos, H. et al., 2013. The immune interplay between the host and the pathogen in Aspergillus fumigatus lung infection. *BioMed research international*, 2013, p.693023.

Scanlin, T.F. & Glick, M.C., 2001. Terminal glycosylation and disease: influence on cancer and cystic fibrosis. *Glycoconjugate journal*, 17(7-9), pp.617–26.

Scanlin, T.F. & Glick, M.C., 1999. Terminal glycosylation in cystic fibrosis. *Biochimica et biophysica acta*, 1455(2-3), pp.241–53.

Schubert, M. et al., 2012. Plasticity of the β-trefoil protein fold in the recognition and control of invertebrate predators and parasites by a fungal defence system. *PLoS pathogens*, 8(5), p.e1002706.

Schulz, B. et al., 2005. Mucin glycosylation changes in cystic fibrosis lung disease are not manifest in submucosal gland secretions. *Biochem. J*, 919(387), pp.911–919.

Seksek, O., Biwersi, J. & Verkman, A.S., 1996. Evidence against Defective trans-Golgi Acidification in Cystic Fibrosis. *Journal of Biological Chemistry*, 271(26), pp.15542–15548..

Shapiro, R.S., Robbins, N. & Cowen, L.E., 2011. Regulatory circuitry governing fungal development, drug resistance, and disease. *Microbiology and molecular biology reviews : MMBR*, 75(2), pp.213–67.

Sharon, N. & Lis, H., 2004. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*, 14(11), p.53R–62R.

Shenoy, S.R. et al., 2002. Multisite and multivalent binding between cyanovirin-N and branched oligomannosides: calorimetric and NMR characterization. *Chemistry & biology*, 9(10), pp.1109–18.

Shenoy, S.R. et al., 2001. Selective interactions of the human immunodeficiency virus-inactivating protein cyanovirin-N with high-mannose oligosaccharides on gp120 and other glycoproteins. *The Journal of pharmacology and experimental therapeutics*, 297(2), pp.704–10.

Da Silva, R.C., Segat, L. & Crovella, S., 2011. Role of DC-SIGN and L-SIGN receptors in HIV-1 vertical transmission. *Human immunology*, 72(4), pp.305–11..

Singh, R.S., Bhari, R. & Rai, J., 2010. Further screening of Aspergillus species for occurrence of lectins and their partial characterization. *Journal of basic microbiology*, 50(1), pp.90–7..

Singh, R.S., Tiwary, A.K. & Bhari, R., 2008. Screening of Aspergillus species for occurrence of lectins and their characterization. *Journal of basic microbiology*, 48(2), pp.112–7.

Slesiona, S. et al., 2012. Persistence versus escape: Aspergillus terreus and Aspergillus fumigatus employ different strategies during interactions with macrophages. *PloS one*, 7(2), p.e31223..

Stein, N., 2008. CHAINSAW : a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography*, 41(3), pp.641–643..

Stevens, D.A. et al., 2003. Allergic Bronchopulmonary Aspergillosis in Cystic Fibrosis — State of the Art. *Cystic Fibrosis Foundation Consensus Conference*, 37(Suppl 3).

Stillmark, H., 1888. Über Ricin, ein giftiges ferment aus den Samen von Ricinus comm. L. und einigen anderen Euphobiaceen. *Thesis Univ.Dorprat*, p.quoted by Kocourek 1986.

Sulzenbacher, G. et al., 2010. Crystal structure of the GalNAc/Gal- specific agglutinin from the phytopathogenic ascomzcete Sclerotinia sclerotiorum reveal novel adaptation of a beta-trefoil domain. , 400(4), pp.715–723.

Sundstrom, P., 2002. Microreview Adhesion in Candida spp. , 4, pp.461–469.

Symoens, F. et al., 2010. Unusual Aspergillus species in patients with cystic fibrosis. *Medical mycology*, 48 Suppl 1, pp.S10–6..

Tang, A.C. et al., 2014. Current concepts: host-pathogen interactions in cystic fibrosis airways disease. *European respiratory review : an official journal of the European Respiratory Society*, 23(133), pp.320–332.

Teotia, S., Mondal, K. & Gupta, M.N., 2006. Integration of Afffinity Precipitation with Partitioning Methods for Bioseparation of Chitin Binding Lectins. *Food and Bioproducts Processing*, 84(C1), pp.37–43.

Thau, N. et al., 1994. rodletless mutants of Aspergillus fumigatus. *Infection and immunity*, 62(10), pp.4380–8..

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 Research*, 22(22), pp.4673–4680.

155

Tiralongo, J. et al., 2009. Inhibition of Aspergillus fumigatus conidia binding to extracellular matrix proteins by sialic acids: a pH effect? *Microbiology (Reading, England)*, 155(Pt 9), pp.3100–9..

Tremblay, M.J., 2010. HIV-1 and pattern-recognition receptors: a marriage of convenience. *Nature immunology*, 11(5), pp.363–5..

Tronchin, G. et al., 1997. Expression and identification of a laminin-binding protein in Aspergillus fumigatus conidia. *Infection and immunity*, 65(1), pp.9–15..

Tronchin, G. et al., 2002. Purification and Partial Characterization of a 32-Kilodalton Sialic Acid-Specific Lectin from Aspergillus fumigatus. *Infection and immunity*, 70(12), pp.6891–6895.

Upadhyay, S.K. et al., 2009. Identification and characterization of a laminin-binding protein of Aspergillus fumigatus: extracellular thaumatin domain protein (AfCalAp). *Journal of medical microbiology*, 58(Pt 6), pp.714–22.

Vakonakis, I. et al., 2008. Solution Structure and Sugar-Binding Mechanism of Mouse Latrophilin-1 RBL: a 7TM Receptor-Attached Lectin-Like Domain. *Structure*, 16(June), pp.944–953.

Varki, A. et al., 2009. *Essentials of Glycobiology*,

Varki, A. & Gagneux, P., 2012. Multifarious roles of sialic acids in immunity. *Annals of the New York Academy of Sciences*, 1253, pp.16–36.

Varrot, A., Basheer, S.M. & Imberty, A., 2013. Fungal lectins: structure, function and potential applications. *Current opinion in structural biology*, 23(5), pp.678–85..

Veelders, M. et al., 2010. Structural basis of flocculin-mediated social behavior in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 107(52), pp.22511–6..

Vermeulen, E., Lagrou, K. & Verweij, P.E., 2013. Azole resistance in Aspergillus fumigatus: a growing public health concern. *Current opinion in infectious diseases*, 26(6), pp.493–500.

De Vrankrijker, a M.M. et al., 2011. Aspergillus fumigatus colonization in cystic fibrosis: implications for lung function? *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 17(9), pp.1381–6.

Warwas, M.L. et al., 2007. Structure and role of sialic acids on the surface of Aspergillus fumigatus conidiospores. *Glycobiology*, 17(4), pp.401–10.

Wasylnka, J. a & Moore, M.M., 2000. Adhesion of Aspergillus species to extracellular matrix proteins: evidence for involvement of negatively charged carbohydrates on the conidial surface. *Infection and immunity*, 68(6), pp.3377–84.

Wasylnka, J. a, Simmer, M.I. & Moore, M.M., 2001. Differences in sialic acid density in pathogenic and non-pathogenic Aspergillus species. *Microbiology (Reading, England)*, 147(Pt 4), pp.869–77.

Willment, J. a & Brown, G.D., 2008. C-type lectin receptors in antifungal immunity. *Trends in microbiology*, 16(1), pp.27–32.

Wimmerova, M. et al., 2003. Crystal structure of fungal lectin: six-bladed beta-propeller fold and novel fucose recognition mode for Aleuria aurantia lectin. *The Journal of biological chemistry*, 278(29), pp.27059–67..

Winn, M.D. et al., 2011. Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography*, 67(Pt 4), pp.235–42.

Wiseman, T. et al., 1989. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical Biochemistry*, 179(1), pp.131–137.

Witt, H., 2003. Chronic pancreatitis and cystic fibrosis. *Gut*, 52 Suppl 2, pp.ii31–41.

Wolf, J.M. et al., 2010. The Candida albicans ESCRT pathway makes Rim101-dependent and -independent contributions to pathogenesis. *Eukaryotic Cell*, 9(8), pp.1203–1215.

Wong, J.Y. et al., 2014. Successful lung transplant in a child with cystic fibrosis and persistent Blastobotrys rhaffinosifermentans infection. *Pediatric transplantation*, 18(5), pp.E169–73.

Xia, B. et al., 2005. Altered O-glycosylation and sulfation of airway mucins associated with cystic fibrosis. *Glycobiology*, 15(8), pp.747–75.

Xiong, S., Fan, J. & Kitazato, K., 2010. The antiviral protein cyanovirin-N: the current state of its production and applications. *Applied Microbiology and Biotechnology*, 86(3), pp.805–812..

Yang, F. et al., 1999. Crystal structure of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. *Journal of molecular biology*, 288(3), pp.403–12.

Zhu, X. et al., 2000. Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry*, 39(37), pp.11194–204.

Žurga, S. et al., 2014. A novel β-trefoil lectin from the parasol mushroom (Macrolepiota procera) is nematotoxic. *The FEBS journal*, 281(15), pp.3489–506.