



HAL
open science

Characterization of lectins from emerging opportunistic fungi and identification of novel lectins from fungal extracts

Dania Martinez-Alarcon

► **To cite this version:**

Dania Martinez-Alarcon. Characterization of lectins from emerging opportunistic fungi and identification of novel lectins from fungal extracts. Biochemistry [q-bio.BM]. Université Grenoble Alpes [2020-..]; Universiteit Utrecht (Utrecht, Nederland), 2021. English. NNT : 2021GRALV043 . tel-04464042

HAL Id: tel-04464042

<https://theses.hal.science/tel-04464042>

Submitted on 18 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Préparée dans le cadre d'une cotutelle *entre
l'Université Grenoble Alpes et l'Université
d'Utrecht*

Spécialité : **CHIMIE - BIOLOGIE**

Arrêté ministériel : le 6 janvier 2005 – 25 mai 2016

Présentée par

Dania MARTINEZ-ALARCON

Thèse dirigée par **Dr. Annabelle VARROT** et **Prof. Roland
J. PIETERS**

préparée au sein des **Laboratoire Centre de Recherche
sur les Macromolécules Végétales et Biologie chimique
et découverte de médicaments**
dans les **Écoles Doctorales Chimie et Sciences du
Vivant, et Drug Innovation**

Caractérisation de lectines de champignons opportunistes émergents et identification de nouvelles lectines à partir d'extraits fongiques

Thèse soutenue publiquement le **17 septembre 2021**
devant le jury composé de :

Dr. Thibaut CREPIN

Université Grenoble Alpes, Président

Prof. Ute KRENGEL

Université d'Oslo, Rapportrice

Dr. Angelina de Palma

Université NOVA de Lisbonne, Rapportrice

Prof. Geert-Jan BOONS

Université d'Utrecht, Examinateur



THESIS

To obtain the degree of

DOCTOR OF GRENOBLE ALPES UNIVERSITY

**Prepared as part of a joint supervision
between the Université Grenoble Alpes and
the University of Utrecht**

Specialty: CHEMISTRY - BIOLOGY

Ministerial decree: January 6, 2005 - May 25, 2016

Presented by

Dania MARTINEZ-ALARCON

Thesis supervised by **Dr. Annabelle VARROT** and **Prof. Roland J. PIETERS**

prepared in the Laboratory **Center for Research on Plant
Macromolecules and Chemical Biology and Drug Discovery**
In the Doctoral Schools of **Chemistry and Life Sciences, and Drug
Innovation**

Characterization of lectins from emerging opportunistic fungi and identification of novel lectins from fungal extracts

Thesis defended on **September 17th, 2021**
in front of the jury composed by:

Dr. Thibaut CREPIN
Grenoble University, President
Dr. Ute KRENGEL
University of Oslo, Reviewer
Dr. Angelina de Palma
NOVA University Lisbon, Reviewer
Prof. Geert-Jan BOONS
Utrecht University, Examiner



The present work was carried out jointly by Grenoble University (France) and Utrecht University (The Netherlands), in the context of the educational training network of the European Joint Doctorate PhD4GlycoDrug.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 765581



I want to dedicate this work to my greatest driving force in life, the person who makes me wish to be a better version of myself every day, and who is always by my side to support me unconditionally.

From my heart to my husband: Rodrigo Escobar Díaz Guerrero.

AKGNOWLEGMENTS

This project was a great challenge full of obstacles that would not have been possible to overcome without the help and collaboration of a huge list of people, to whom I would like to dedicate a few words.

First of all, I would like to thank the members of my thesis committee: **Dr. Thibaut Crépin**, **Dr. Ute Krengel**, **Dr. Angelina de Palma**, and **Prof. Geert-Jan Boons** for taking the time to examine this thesis and join me in the final defense of this work.

I also would like to thank the members of my CSI committee: **Dr. Robert de Vries**, and **Dr. Olivier Lerouxel** for their advice and commentaries during our meetings. Without a doubt, your critical thinking and questions were key pieces in the search for answers beyond the obvious.

To our collaborators, **Prof. Muriel Cornet**, **Dr. Julia Novion** and **Dr. Liu Yan**, for their great disposition and their faith in this project. Without your contributions, none of this would have been possible.

To **Dr. Serge Perez** and **Dr. Anne Imberty**, thank you for sharing your great knowledge and for always bringing interesting discussions to our meetings.

A very special thank you also to **Valérie Chazalet** and **Emilie Gillon**, for sharing all their experience with me and for always being there when I had a question. Thanks to both of you, because working in such an organized environment made my days at CERMAV much more relaxed and productive. But above all, thank you because your enthusiasm and great personality make the lab a wonderful place to work. It's great to always start the day with your smiles.

To my sisters **Diana** and **Nadia**, for their support and patience throughout the development of this project and for never losing faith in me. For being my source of inspiration and the engine that drives me to keep going every day.

I want to thank my husband **Rodrigo Escobar**, for being my greatest support in the adverse moments and letting me share with him my joys and achievements in the good moments. For his immense patience and love, for being always present (even at distance), for being willing to cross the world just to make me smile. Rodrigo, it is difficult to resume all the reasons that I am grateful to you. So, I will mention only the most important one: thank you for being in my life!

To the **Priem De la Rosa family**, for having welcomed me with open arms since I arrived at France. **Aida, Bernard, Emma and Rafa**, thank you for making me feel part of the family, for all those afternoons full of games, laughter, and delicious food. But more than anything, thank you because I never felt alone in difficult times, you were always there to lend a hand when I needed it most.

To my colleagues and friends who were present in the most representative stages of this project, always ready to support me and encourage me to move forward. Thank you, **Simona, Dylan, Jalaa, Nathan, Isaty, Olga, Aurore, Oriane, Roosmarijn, Torben, Vito, Gael and Luca**, for the shared experiences, your support, and your contributions, for making lunchtime a very nice experience to wait for during the morning. Thank you for all the laughs, and for all the nice moments we shared.

To the board members of the **PhD4GlycoDrug** consortium, for making this great project possible, and for being a constant source of inspiration to me. I also want to thank you for your technical support and for your interest in our professional and personal development. A very special thanks to **Dr. Marko Anderluh** who has accompanied me throughout this journey in a unique manner, and who is always there to support me in undertaking new challenges. Marko, I sincerely appreciate that you care about every ESR and you are always looking new opportunities for us.

Now that I mentioned the PhD4GlycoDrug, I want to especially thank the other ESRs who have shared this great adventure with me and who, without a doubt, have become part of my family: **Rafael, Kanhaya, Margherita, Nives, Mujtaba, Sjors, Benedetta, Cyril, Elena, Tiago, and Gabriele**. Thank you very much because this adventure would not have been the same without your company and support. Our adventures will be forever in my memory, and you all have a very special place in my heart. I consider myself a very lucky person because, well, what are the possibilities to bring together such a particular group of people? jajaja. I cannot imagine what our advisers would have thought during the recruitment, but I am very happy for having met you.

I know this is getting too long, but I really need to thank **Elena Loi** and **Gabriele Conti** for their technical advice in chemistry, for staying up late in the lab to discuss and help me interpret results. Also, thanks for your patience while teaching me the basics of chemistry.

From all the people I have thanked, **Margherita Duca** has a very very very special place on the list. Somehow our projects are interconnected and together we live the joys and also the failures of the other. Maybe for this reason we created a very strong bond (not hydrogen, jaja terrible joke). Marghe was my teacher and the person who guided me through the part of the project that involved organic synthesis, in which I had no experience. When I arrived at the laboratory, I didn't know how to do a liquid extraction, I didn't know how to interpret an NMR, I didn't know how to clamp the material in the fumehood, I didn't know anything at all. Nevertheless, Marghe was always there to teach me and to avoid I died. But farther than an incredibly patient teacher, she is also an exceptional friend and ABSOLUTELY NOTHING would have been the same without her company in this journey. Thank you forever my Margarita, for supporting me in difficult times, for your unconditional love. I am only sure of

one thing, wherever I will go from now on, you will always be an important part of my life.

To **Dr. Rafael Bermeo Malo**... well, maybe you will never realize the watershed that your presence in this world has meant for my life. But it doesn't matter, the only thing you need to know is that I will be eternally grateful to you because there is a Dania before and after Rafael. I really hope that your lessons will accompany me for the rest of my life, because they'll be essential to achieve happiness. Thank you for having crossed this path by my side. We started this together, and we finished it together. Until forever Mr. Malo!

Finally, I would like to express my most sincere gratitude to my advisors **Dr. Annabelle Varrot** and **Prof. Roland Pieters** for giving me the opportunity to be a part of their teams, and for all their guidance during these years. Words are not enough for me to express the joy of having been your student. I hope one day to be lucky enough to work together again.

Honorable Professor Pieters, XD!

Thank you very much for welcoming me to Utrecht. Without a doubt, the time I spent in your laboratory was a great opportunity to learn and develop new skills both in the work and personal aspects. I appreciate the opportunity to work together and your willingness to contribute to the maturation of my scientific skills, but above all, I want to thank you for trusting me.

Although transitory, my stay in Utrecht was magnificent and I definitively learned a lot more than I would have ever thought possible in a year. I also enjoyed our personal talks and your always active sense of humor. So, I am very grateful to you for giving to me this opportunity. If there is something that I will always remember about my stay with you, it is that "*Sometimes the answer to the most complicated question is found in the simplest details*".

Annabelle,

I didn't tell you this before, but when I arrived to your lab I was terrified. I had the impression that the project was too ambitious and that it would hardly be possible for me to acquire the necessary skills to be up to the task. Then, the first challenges arrived almost at the beginning of the project, with SapL1. I spent entire nights awake wondering if I was the right person to overcome the situation? If you made the right decision by giving me this opportunity? But you seemed pretty confident about it. You may not have realized it at the time, but your support made me feel deeply grateful and I was determined not to disappoint you. So, I kept trying, harder and harder until we were finally able to overcome that obstacle. Later, as other challenges came (because things did not go very often as were expected), you always stayed by my side. Your technical advice and support helped me to get through all the difficulties of this project, and, one by one, we overcome all of them.

It's funny because during the "good times", when I thought I had done a magnificent job, you always had something new to teach me. I must admit that on many occasions I felt frustrated because I was totally sure I had done my best. But precisely there, when I thought it was impossible to improve, you always pushed me beyond what I thought was my limit. I personally believe that you have a brilliant mind an enormous talent to enhance the skills of your students. But, beyond the scientific field, I admire you because you are a great woman, full of strength but with a great heart.

There are so many other things that I would like to tell you, but words hardly could express how grateful I am, and will always be, to you. So, I'll just say: Thank you Annabelle, for sharing your research experience with me and for being there when I needed you. For never leaving me alone, for pushing me through both, difficult and good times. But above all, I want to thank you for trusting me and my ideas.

Once upon a time there was a bunny, sitting in the sun and munching on some brownies.

A fox comes along and asks the bunny what she's up to. The bunny says: "I'm writing my dissertation on how bunnies eat foxes". "What?! Are you serious?" says the fox. "Everyone knows foxes eat bunnies!". "Come and see", says the bunny, and the fox follows her into her rabbit hole. A few minutes later, out comes the bunny, with a few small stains of blood on her fur, which she promptly cleans up and sits down to munch on brownies.

A wolf comes along, and asks the bunny what she's up to. The bunny says: "I'm writing my dissertation on how bunnies eat wolves". "What?! Are you serious?" says the wolf. "Everyone knows wolves eat bunnies!". "Come and see", says the bunny, and the wolf follows her into her rabbit hole. A few minutes later, out comes the bunny, using a toothpick to clean between her teeth. She sits down to munch on brownies.

A tiger comes along, and asks the bunny what she's up to. The bunny says: "I'm writing my dissertation on how bunnies eat tigers". "What?! Are you serious?" says the tiger. "Everyone knows tigers eat bunnies!". "Come and see", says the bunny, and the tiger follows her into her rabbit hole.

The tiger follows the bunny into the burrow, where a lion sits up as soon as they arrive, kills the tiger with one stroke, and waits patiently for the bunny to collect herself and go back outside.

The moral of the story is that it does not matter what your thesis topic is; it only matters who your advisor is.

TABLE OF CONTENTS

SUMMARY	27
1. GENERAL INTRODUCTION	33
1.1 Fungal infections	33
1.2 <i>Scedosporium apiospermum</i> an emerging fungal pathogen	33
1.2.1 <i>Emerging opportunistic fungal pathogens</i>	33
1.2.2 <i>Scedosporium complex</i>	34
1.2.3 <i>Scedosporium spp pathogeny</i>	35
1.2.4 <i>Distribution and epidemiological overview</i>	39
1.2.5 <i>Cystic Fibrosis; a high-risk population for scedosporiosis</i>	40
1.2.6 <i>Clinical treatment</i>	41
1.2.7 <i>Host-binding; the first step towards infection</i>	42
1.3 Lectins	43
1.3.1 <i>Definition and generalities</i>	43
1.3.2 <i>Carbohydrate-binding properties of lectins</i>	44
1.3.3 <i>Specificity of lectins</i>	45
1.3.4 <i>Classification of lectins</i>	47
1.3.5 <i>Fungal lectins: roles and biotechnological applications</i>	48
1.4 The anti-adhesive approach.....	49
1.4.1 <i>Host-binding; a prerequisite for infection.....</i>	49
1.4.2 <i>Anti-adhesive therapy.....</i>	51
1.4.3 <i>Carbohydrates and glycomimetics as microbial anti-adhesives.....</i>	52
1.5 <i>Aspergillus fumigatus</i> lectins.....	54
2 AIM OF THE THESIS AND OBJECTIVES	56
3. PROJECT 1: Biochemical and structural studies of target lectin SapL1 from the emerging opportunistic microfungus <i>S. apiospermum</i>.....	59
3.1 Summary	59
3.2 Publication I.....	61

3.3 Supplementary information	72
4. PROJECT 2: Biochemical and structural characterization of a Cyanovirin-like lectin from human opportunistic fungi	81
4.1 Summary	81
4.2 Introduction	81
4.2.1 <i>Cyanovirin-N</i>	82
4.2.2 <i>The CV-N fold</i>	83
4.2.3 <i>Cyanovirin-N homologs (CVNHs)</i>	83
4.2.4 <i>CVNHs carbohydrate binding properties</i>	85
4.2.5 <i>CVNHs biological roles</i>	86
4.2.6 <i>CVNHs subcellular localization</i>	87
4.2.7 <i>AFL6</i>	87
4.3 Material and methods	88
4.3.1 <i>Genetic construction</i>	88
4.3.2 <i>Production of SapL6</i>	89
4.3.3 <i>Purification of SapL6</i>	89
4.3.4 <i>Size exclusion chromatography</i>	90
4.3.5 <i>Dynamic light scattering (DLS)</i>	90
4.3.6 <i>Thermal Shift Assay (TSA)</i>	90
4.3.7 <i>Isothermal Titration Calorimetry (ITC)</i>	91
4.3.8 <i>Hemagglutination assay</i>	91
4.3.9 <i>Mammalian Glycan array assay (CFG)</i>	91
4.3.10 <i>Crystallization and data collection</i>	92
4.3.11 <i>Structure determination</i>	93
4.4 Results and discussion	93
4.4.1 <i>Identification of SapL6</i>	93
4.4.2 <i>Production and purification of SapL6</i>	95
4.4.3 <i>Molecular size and oligomeric stage of SapL6</i>	96
4.4.4 <i>Thermal stability of SapL6</i>	97
4.4.5 <i>Hemagglutination activity of SapL6</i>	98
4.4.6 <i>SapL6 carbohydrate binding properties ITC</i>	99
4.4.7 <i>SapL6 carbohydrate binding properties; glycan array</i>	100
4.4.8 <i>SapL6 crystallization</i>	102
4.4.9 <i>SapL6 Structure</i>	103

4.4.10 Comparison of SapL6 with other CVNHs	105
4.4.11 Structural insights of putative carbohydrate-binding pockets in SapL6.....	107
4.5 Conclusion	110
4.6 Contributions	110
4.7 Supplementary information	112
5. PROJECT 3: Identification of new native lectins from protein extracts of medical relevant fungal strains.....	113
5.1 Summary	113
5.2 Introduction	114
5.3 Material and methods	117
5.3.1 Fungal strains.....	117
5.3.2 Sample preparation	117
5.3.3 Lectins purification	117
5.3.4 SDS-PAGE.....	117
5.3.5 Proteins sequencing	118
5.4 Results and discussion	118
5.4.1 Purification of native lectins.....	118
5.4.2 Sequencing of native lectins	121
5.4.3 Identification of SapCP from <i>S. apiospermum</i>	124
5.4.4 Identification of PVL-like lectins from <i>S. apiospermum</i> and <i>A. fumigatus</i>	128
5.4.5 Identification of the fucose-binding lectins FleA and SapL1	133
5.4.6 Identification of AfFG-GAP repeat protein from <i>A. fumigatus</i>	136
5.4.7 Identification of CcMBL-homologs in <i>A. fumigatus</i> and <i>S. apiospermum</i>	140
5.4.8 Identification of SapFKBP12 from <i>S. apiospermum</i>	144
5.5 Conclusions	152
5.6 Contributions	152
5.7 Supplementary information	153
6. PROJECT 4: GCFs- a new mini glycanarray for identification and purification of lectins.....	165

6.1 Summary	165
6.2 Introducton	165
6.2.2 <i>Common strategies for lectins purification.....</i>	166
6.2.3 <i>Drawbacks of sugar-based affinity chromatography</i>	167
6.2.4 <i>Brief description of the device</i>	169
6.3 Material and methods.....	170
6.3.1 <i>Synthesis of 2'-Chloroethyl 2,3,4-Tri-O-acetyl-α-L-fucopyranoside</i>	170
6.3.2 <i>Synthesis of 2-(2-cloroethoxy)ethylper-O-acetylated glucosides.....</i>	171
6.3.3 <i>General procedures for azidation of glycosides</i>	171
6.3.4 <i>General procedures for deprotection of glycosides</i>	171
6.3.5 <i>General procedures for the reduction of azido glycosides</i>	171
6.3.6 <i>Hydroxylation of borosilicate fibers</i>	172
6.3.7 <i>Silane coupling (Self-assembled monolayer)</i>	172
6.3.8 <i>Monodeposition of glycosides on GPS-coated filters</i>	172
6.3.9 <i>Multi-spot deposition of glycosides on GPS-coated filters</i>	173
6.3.10 <i>Lectin binding to glycoside-coated filters</i>	173
6.3.11 <i>Protein denaturing electrophoresis (SDS-PAGE).....</i>	173
6.4 Results and discussion	173
6.4.1 <i>Glycosides synthesis</i>	173
6.4.2 <i>Borosilicate surface modification</i>	174
6.4.3 <i>Specific interaction between lectins and glycosides-coated filters</i>	175
6.4.4 <i>Mini chip for detection of protein-glycan interactions on unlabeled samples .</i>	177
6.4.5 <i>Recovery of lectins from complex protein mixtures.....</i>	179
6.5 Conclusions	180
6.6 Contributions	181
7. GENERAL CONSLUSIONS AND PERSPECTIVES.....	183
REFERENCES.....	191
APPENDIX.....	220
Publication II.....	221

LIST OF FIGURES

Figure 1.1: <i>Scedosporium</i> genus and <i>S. apiospermum</i>	35
Figure 1.2: <i>Scedosporium</i> spp pathogeny.....	38
Figure 1.3. Global distribution and relative burden of scedosporiosis.....	39
Figure 1.4: <i>Scedosporium apiospermum</i> in the Cystic Fibrosis context.....	41
Figure 1.5: Carbohydrates-lectin interactions.....	43
Figure 1.6: Carbohydrate recognition domain of lectins	44
Figure 1.7: Schematic representation of lectins symmetry	45
Figure 1.8: Examples of glycan determinants recognized by different lectins... ..	47
Figure 1.9: Roles and potential applications of fungal lectins.....	49
Figure 1.10: Fungal adhesion to epithelial cells.....	51
Figure 1.11: Overall structure of FleA and AFL6.	55
Figure 2: Thesis results overview	57
Figure 3.1: Graphical abstract of project 1.....	60
Figure 4.1: Graphical abstract of project 2.....	82
Figure 4.2. CV-N Structure.....	83
Figure 4.3. Phylogenetic tree of the CVNH Family.	84
Figure 4.4. CVNHs sugar binding specificity	86
Figure 4.5: Comparison of CVNH tandem repeats	94
Figure 4.6: Phylogenetic tree of CVNHs	95
Figure 4.7. SapL6 expression and purification	96
Figure 4.8. SapL6 oligomerization stage	97

Figure 4.9. SapL6 thermal Stability	98
Figure 4.10. Hemagglutination activity of SapL6	98
Figure 4.11. SapL6 interaction with the mammalian glycan array of CFG	100
Figure 4.12. optimization of SapL6 purification.....	102
Figure 4.13. SapL6 crystallization.....	103
Figure 4.14. SapL6 Structure	104
Figure 4.15. Comparison of SapL6 with other CVNHs.....	106
Figure 4.16. Structural comparison of CV-N and SapL6 carbohydrate BP	108
Figure 4.17. Structural comparison of the chimera-LKAMG and SapL6 carbohydrate BP.....	109
Figure 4.18: Multiple alignment of sequences of the CVNH for phylogeny	112
Figure 5.1: Graphical abstract of project 3.....	116
Figure 5.2: Native lectins identified from <i>S. apiospermum</i> mycelium cultures..	120
Figure 5.3: Native lectins identified from <i>A. fumigatus</i> mycelium cultures	120
Figure 5.4: Native lectins identified from fungal cultures 20X concentrated ...	121
Figure 5.5: SapCP	124
Figure 5.6: CPPs structural analysis.....	127
Figure 5.7: Alignment of the peptides identified from B1 and the PVL.....	129
Figure 5.8. Alignment of the peptides identified from B7 and the PVL.....	130
Figure 5.9: <i>Psathyrella velutina</i> Lectin (PVL).....	130
Figure 5.10: Evolutionary tree of the major groups of fungi.....	132
Figure 5.11: Phylogenetic analysis of PVL-like lectins	133

Figure 5.12. Alignment of the peptides identified from B10, B31, B32 and the FleA sequence.....	135
Figure 5.13: Alignment of the peptides identified from B45 and B46 and FleA sequences	136
Figure 5.14: Alignment of the peptides identified from B23/B28 and AfFG-GAP sequence.....	137
Figure 5.15: Alignment of the peptides identified from B25 and AfFG-GAP sequence	138
Figure 5.16: Structural model of AfFG-GAP.....	139
Figure 5.17: CcMBL	140
Figure 5.18 Alignment of the peptides identified from B2, B3, B47 and B48 with the sequence of CcMBL.....	141
Figure 5.19: Alignment of the peptides identified from B8 and B29 with the sequence of CcMBL	142
Figure 5.20: Alignment of the peptides identified from B15 and B4 with the sequence of CcMBL	143
Figure 5.21: Alignment of the peptides identified from B4 and the SapFKBP12 sequence	145
Figure 5.22: Ambivalent effect of CaN inhibition in fungi and human cells	146
Figure 5.23: The role of FKBP12 in the panfungal strategy	148
Figure 5.24: FKBP12 from pathogenic fungi: alignment and structure of complexes.....	149
Figure 5.25: Alignment of sequences of highly conserved fungal FKBP12.....	150
Figure 6.1: Schematic representation of the device.....	170
Figure 6.2: Glycosides synthesis	174
Figure 6.3: Borosilicate surface functionalization and glycosides coating	175

Figure 6.4: Electrophoretic profile of fractions recovered from the GCFs after protein incubation 176

Figure 6.6. Electrophoretic profile of fractions recovered from the GCFs after incubation with total proteins extracts 180

LIST OF TABLES

Table 1.1. Carbohydrates that prevent infection <i>in vivo</i>	53
Table 4.2. Data-collection statistics.....	104
Table 5.1. Summary of mass spectrometry hits.....	123
Table 5.2. List of peptides identified by MS for sample B1	153
Table 5.3. List of peptides identified by MS for sample B2.	154
Table 5.4. List of peptides identified by MS for sample B3.	154
Table 5.5. List of peptides identified by MS for sample B4.	155
Table 5.6. List of peptides identified by MS for sample B7.	155
Table 5.7. List of peptides identified by MSy for sample B8.	156
Table 5.8. List of peptides identified by MS for sample B10.	156
Table 5.9. List of peptides identified by MS for sample B15	156
Table 5.10. List of peptides identified by MS for sample B23.	157
Table 5.11. List of peptides identified by MS for sample B25	158
Table 5.12. List of peptides identified by MS for sample B28	158
Table 5.13. List of peptides identified by MS for sample B29	159
Table 5.14. List of peptides identified by MS for sample B31.	159
Table 5.15. List of peptides identified by MS for sample B32	160
Table 5.16. List of peptides identified by MS for sample B40.	160
Table 5.17. List of peptides identified by MS for sample B41	161
Table 5.18. List of peptides identified by MS for sample B43.	161
Table 5.19. List of peptides identified by MS for sample B45	162
Table 5.20. List of peptides identified by MS for sample B46	163
Table 5.21. List of peptides identified by MS for sample B47	163

Table 5.22. List of peptides identified by MS for sample B48.	164
Table 5.23. List of peptides identified by MS for sample B53	164
Table 6.1. Carbohydrate-based matrix for lectins purification.....	187

LIST OF ABBREVIATIONS

AAL	<i>Aleuria auratia</i> lectin
AFL/FleA	<i>Aspergillus fumigatus</i> fucose binding lectin
ARN	Ribonucleic acid
APS	Ammonium persulfate
Å	Ångström
αMeFuc	Methyl alpha-L-fucopyranoside
BGA	Human blood group antigens
BPs	Binding pockets
BSA	Bovine serum albumin
CaN	Phosphatase calcineurin
CcMBL	<i>Coprinopsis cinerea</i> Mucose Binding lectin
CF	Cystic fibrosis
CFG	Consortium for Functional Glycomics
CGD	Chronic granulomatous disease
CNS	Central nervous system
ConA	Concanavalin A
COPD	Chronic obstructive pulmonary disease
CnA	Calcineurin catalytic subunit
CnB	Calcineurin regulatory subunit
CP	Cerato platanin lectin
CPPs	Cerato-platanin proteins
CRD	Carbohydrate recognition domain
CRZ1	Calcineurin-responsive zinc finger transcription factor
CSM	Cell surface molecules
CV	Column volumes
CyPs	Cyclosporin-binding cyclophilins
DLS	Dynamic light scattering
OD	Optical density
EDTA	Ethylenediaminetetraacetic acid
FITC	Fluorescein Isothiocyanate
FKBPs	FK506-binding proteins
Fuc	Fucose
GAGs	Glycosaminoglycans
Gal	Galactose

GalNAc	N-acetylgalactosamine
CGFs	Glycosides coated filters
GlcNAc	N-acetylglucosamine
Glu	Glucose
GMO	Genetically modified organisms
GNA	<i>Galanthus nivalis</i> lectin
GPS-SAM	GPS-Self Assembled Monolayer
HMOs	Human breast milk oligosaccharides
HSCT	Hematopoietic stem cell transplant
HPLC	High-performance liquid chromatography
IMAC	Immobilized metal affinity chromatography
IgG	Immunoglobulins
IL-2	Interleukin-2
IPTG	Isopropyl-b-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
kDa	Kilodalton
LacI	Lac repressor
MAMP	Microbe-associated a molecular pattern
Man	Mannose
MES	2-(N-morpholino)ethanesulfonic acid
NeuAc	N-acetyl-neuraminic acid
NFAT	Nuclear factor of activated T cells
NMR	Nuclear magnetic resonance
OD	Optical density
ORF	Open reading frame
PDB	Protein data bank
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PPIases	Peptidyl-prolyl isomerases
PRMs	Peptidorhamnomannans
PVL	<i>Psathyrella velutina</i> lectin
PT	Palladium
Rh	Rhamnose
RFU	Relative fluorescent units
RT	Room temperature
SAM	Self assembled monolayer

SapL1	<i>Scedosporium apiospermum</i> lectin 1
SapL6	<i>Scedosporium apiospermum</i> lectin 6
SapCP	<i>Scedosporium apiospermum</i> cerato platanin
SapFKBP12	<i>Scedosporium apiospermum</i> lectin FK506 binding lectin
SEC	Size exclusion chromatography
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate electrophoresis
SOT	Solid organ transplant
TEMED	N,N,N',N'-Tetramethylethylenediamine
TEV	Tobacco etch virus protease
T_m	Melting temperature
Trc	Trc promoter
T7p	T7 promoter
Trx	Thioredoxin reductase
TSA	Thermal shift assay
WT	Wild type
ΔH	Enthalpy

SUMMARY

Scedosporium apiospermum is an emerging opportunistic fungal pathogen responsible for life-threatening infections in immunocompromised patients. This fungus exhibits limited susceptibility to all current antifungals and displays high mortality rates. *S. apiospermum* represents a therapeutic challenge since; on one hand, the options for its treatment are limited when the infection has progressed and, on the other hand, its prevention by antifungal prophylaxis carries significant risks for health. In this context, therapies aimed at precluding its binding to human tissues seem to represent a better alternative for high-risk populations. Unfortunately, due to its emerging character, *S. apiospermum* pathogenicity and virulence factors remain largely unknown.

Some lectins (carbohydrate-binding proteins) are important mediators of host-pathogen interactions during the early stages of infections. Therefore, they are considered valuable therapeutic targets and several approaches have suggested their inhibition as a strategy to prevent infections. The most popular one, proposes the use of sugar-based inhibitors and glycomimetics specifically designed to block the binding site of those proteins as a strategy to avoid microbial adhesion to the host cells. Nevertheless, in the case of *S. apiospermum*, the development of these therapies is hindered because its lectins remain unidentified.

The main aim of this work was the identification and characterization of lectins from this emerging opportunistic fungal pathogen. This has been addressed by two different approaches; the first strategy comprises the *in-silico* prediction of putative lectins codified on its genome by data mining, while the second strategy lies in the identification of native lectins directly from fungal protein extracts. Through the *in-silico* approach, we have identified two novel lectins from *S. apiospermum* genome, SapL1 and SapL6. Both proteins were produced in

Escherichia coli and their full biochemical/structural characterization was achieved. For the second strategy, we have analysed the intracellular proteins and secretomes of *S. apiospermum* in conjunction with another opportunistic fungal pathogen of medical relevance, *Aspergillus fumigatus*. The analysis of peptides obtained by mass spectrometry led to the identification of several proteins, from which 4 corresponded to new hypothetical proteins with carbohydrate-binding activity from *S. apiospermum* (PVL-like lectin, SapCP, CcMBL-like lectin and SapFKBP12). The preliminary analysis of those proteins revealed that some might be attractive pharmacological targets, and others could find application as biotechnological tools.

The second objective of this thesis was the development of a new tool that allows the identification and purification of unknown lectins directly from crude protein extracts. This tool assembles the fusion of the three classic steps of lectin purification train: clarification, concentration, and separation, in a single operating unit. It also allows the simultaneous purification of lectins with different specificities in a quick and simple process. In addition, it can be used as a "mini glycan array" for the study of carbohydrate lectin interactions, without the need for fluorescent labeling. Hence, we consider that this device might greatly contribute to the study of the lectinome of species.

Overall, this work represents a general strategy to address the study *S. apiospermum* lectinome. Our findings contribute to the understanding of glycosylated surface recognition by this emerging opportunistic pathogen and, we expect, will contribute to design targeted therapies against it. Finally, we hope that the information provided here encourages the study and characterization of the new proteins discovered, in the search for biomedical applications and new drug targets.

RÉSUMÉ

Scedosporium apiospermum est un pathogène fongique opportuniste émergent responsable d'infections potentiellement mortelles chez les patients immunodéprimés. Ce champignon présente une sensibilité limitée à tous les antifongiques actuels et affiche un taux de mortalité élevé. *S. apiospermum* représente un défi thérapeutique puisque ; d'une part, les options de traitement sont limitées lorsque l'infection a progressé et, d'autre part, sa prévention par prophylaxie antifongique comporte des risques importants pour la santé. Dans ce contexte, les thérapies visant à empêcher sa liaison aux tissus humains semblent représenter une meilleure alternative pour les populations à risque. Malheureusement, en raison de son caractère émergent, les facteurs de pathogénicité et de virulence de *S. apiospermum* restent largement inconnus.

Certaines lectines (protéines de liaison aux glucides) sont d'importants médiateurs des interactions hôte-pathogène au cours des premiers stades de l'infection. Par conséquent, elles ont été considérées comme des cibles thérapeutiques précieuses et plusieurs approches ont suggéré leur inhibition comme stratégie de prévention des infections. La plus populaire propose l'utilisation d'inhibiteurs à base de sucre et de glycomimétiques spécifiquement conçus pour bloquer le site de liaison de ces protéines comme stratégie pour éviter l'adhésion microbienne aux cellules hôtes. Néanmoins, dans le cas de *S. apiospermum*, le développement de ces thérapies est entravé car ses lectines restent non identifiées.

L'objectif principal de ce travail était l'identification et la caractérisation des lectines de ce pathogène fongique opportuniste. Cela a été traité selon deux approches : la première stratégie comprend la prédiction *in-silico* de lectines putatives codés dans son génome par "data mining", tandis que la seconde stratégie réside dans l'identification de lectines natives directement à partir d'extraits de protéines fongiques. Grâce à l'approche *in-silico*, nous avons

identifié deux nouvelles lectines du génome de *S. apiospermum*, SapL1 et SapL6. Les deux protéines ont été produites dans *Escherichia coli* et leur caractérisation biochimique/structurale complète a été réalisée. Pour la deuxième stratégie, nous avons analysé les protéines intracellulaires et les sécrétomes de *S. apiospermum* en conjonction avec un autre pathogène fongique opportuniste d'importance médicale, *Aspergillus fumigatus*. L'analyse des peptides obtenus par spectrométrie de masse a conduit à l'identification de plusieurs protéines, dont 4 correspondaient à de nouvelles protéines hypothétiques à activité de liaison aux glucides de *S. apiospermum* (lectine de type PVL, SapCP, lectine de type CcMBL et SapFKBP12). L'analyse préliminaire de ces protéines a révélé que certaines pourraient être des cibles pharmacologiques intéressantes, et d'autres pourraient trouver des applications en tant qu'outils biotechnologiques.

Le deuxième objectif de cette thèse était le développement d'un nouvel outil permettant l'identification et la purification de lectines inconnues directement à partir d'extraits bruts de protéines. Cet outil rassemble la fusion des trois étapes classiques lors de la purification des lectines : clarification, concentration et séparation, en une seule unité opératoire. Il permet également la purification simultanée de lectines avec différentes spécificités dans un processus simple et rapide. De plus, il peut être utilisé comme un "mini glycane array" pour l'étude des interactions entre les glucides et les lectines, sans nécessiter de marquage fluorescent. Par conséquent, nous considérons que ce dispositif pourrait grandement contribuer à l'étude du lectinome des espèces.

Dans l'ensemble, ce travail représente une stratégie générale pour aborder l'étude du lectinome de *S. apiospermum*. Nos découvertes contribuent à la compréhension de la reconnaissance de surface glycosylée par ce pathogène opportuniste émergent et, nous l'espérons, contribueront à la conception de nouvelles thérapies. Enfin, nous espérons que les informations fournies ici encourageront

l'étude et la caractérisation des nouvelles protéines découvertes, dans la recherche d'applications biomédicales et de nouvelles cibles thérapeutiques.

1. GENERAL INTRODUCTION

1.1 Fungal infections

Fungal infections are an important worldwide cause of human morbidity and mortality that is often overlooked due to their underreporting. This is largely due to limited access to accurate and inexpensive tests for the detection of fungal species during co-infections by other pathogens, such as bacteria [1]. Therefore, it is considered that the actual number of deaths attributable to those microorganisms could be greatly underestimated. Despite this, their epidemiological overview points to a dramatic increase in the occurrence of life-threatening infections since 1980, primarily in immunosuppressed patients [2].

Although it is true that these incidences can be correlated to the growing size of risk populations, strongly reinforced by the use of immunosuppressive agents and corticosteroids, these are not the major triggers. This phenomenon has been mainly attributed to the appearance of new species of yeasts and molds, previously harmless to humans, which have developed the capacity to generate invasive infections [1–3]. These species collectively cause enormous damage to human health and affect billions of people each year [1].

1.2 *Scedosporium apiospermum* an emerging fungal pathogen

1.2.1 *Emerging opportunistic fungal pathogens*

During the last decades, an increased incidence of invasive infections, especially in immunosuppressed patients, has been caused by previously rare fungal species [4]. These organisms are called “emerging opportunistic fungal pathogens” and represent a major health problem due to the scarce information about their virulence factors, host-binding mechanisms and treatment. Among those emerging organisms, species from the *Scedosporium* complex are particularly dangerous since their mortality rate may be over 80% [5].

1.2.2 *Scedosporium* complex

The genus *Scedosporium* comprises more than ten species (*Scedosporium angustum*, *S. apiospermum*, *S. aurantiacum*, *S. boydii*, *S. cereisporum*, *S. dehoogii*, *S. desertorum*, *S. ellipsoideum*, *S. fusoideum*, *S. minutisporum*, *S. rarisporum* and *S. sanyaense*) whose taxonomy has been previously described based on molecular phylogenetic by using the sequences of four genetic loci [5,6] (Figure 1.1A). However, only the genome of *S. boydii*, *S. aurantiacum*, *S. dehoogii* and *S. apiospermum* have been completely sequenced to date [7–9].

Scedosporium infections mainly affect immunocompromised patients such as solid organ transplant (SOT) or hematopoietic stem cell transplant (HSCT) recipients or patients with chronic granulomatous disease (CGD). However, these pathogens also represent a high risk for patients with other immune deficiencies such as cancer and cystic fibrosis (CF). Finally, although with low frequency, *Scedosporiosis* infections have also been reported in healthy persons [10,11].

Within the genus, the main responsible for human's infections is *S. apiospermum*, formerly considered as the asexual form of *Pseudallescheria boydii* (Figure 1.1B-E). This cosmopolitan microfungus is responsible for localized to severe or fatal disseminated infections and is one of the most common *Scedosporium* species (with *S. boydii* and *S. aurantiacum*) capable to chronically colonize the lungs of CF patients [12,13].

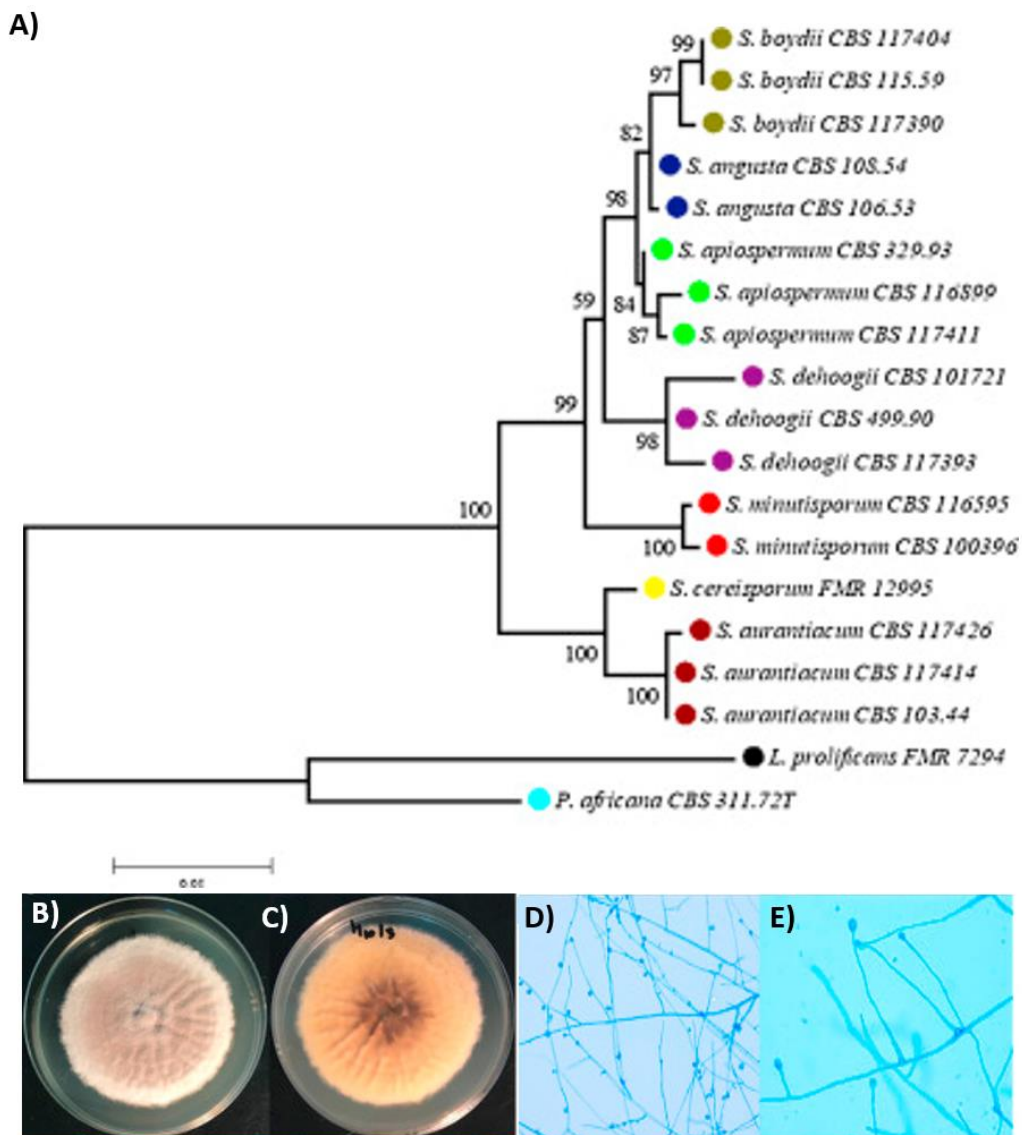


Figure 1.1: *Scedosporium* genus and *S. apiospermum*. A) Molecular phylogenetic maximum likelihood analysis of the BT2 gene of *Scedosporium* complex species. B, C) Macroscopic appearance of *S. apiospermum* (top view and reverse, respectively). D, E) Microscopic appearance conidiogenous cells and conidia (magnification, 40 X and 100 X, respectively). Adapted with permission from Luplertlop, 2018 [7].

1.2.3 *Scedosporium* spp pathogeny

Infections caused by any member of the genus *Scedosporium* are commonly known as scedosporiosis and originate predominantly in the lungs and soft tissues, with occasional spread to bones. The range of diseases caused by these infections is wide, ranging from localized cutaneous infections to full dissemination in both, immunocompromised and immunocompetent hosts

[10,11,13]. In humans scedosporiosis can be divided into 4 main categories according to the clinical conditions associated with the pathologies: (i) mycetoma, (ii) colonization of the respiratory tract, (iii) nonmycetoma deep infections and (v) disseminated infections (Figure 1.2A).

Mycetomas are subcutaneous infections characterized by the formation of compact aggregates of mycelium. These structures are known as granules or sclerotia and their main function is to allow these opportunistic pathogens to evade the defenses of the host's immune system [14]. The color and morphology of the sclera can vary according to the species, but they are usually granules of 2 to 20 μm in diameter. These type of infections are relatively uncommon in immunocompromised hosts [15].

Colonization of the respiratory tract is very common in the presence of pathologies that affect bronchial secretions, such as tuberculosis, chronic obstructive pulmonary disease (COPD), CF, Job's syndrome, etc. This is because the abnormally viscous mucus favors the anchorage of pathogens and interfere with the clearance mechanisms of the lungs. *Scedosporium* colonization can be transient or permanent. Although, some patients with chronic colonization may have minimal or no symptoms, others develop pulmonary infiltrates or allergic reactions [11,16].

Nonmycetoma deep infections comprises all those local infections that do not present clinical manifestations characteristic of mycetomas (such as sclerotia formation). These diseases, in turn, can be subclassified into sinopulmonary and extrapulmonary infections [11]. **Sinopulmonary infections**, as its name indicates, are located at lungs. They are originated from conidia germination that results in an invasion of the lower respiratory tract by hyphae, typically producing a necrotizing pneumonia. This affection is very rare in healthy individuals; however, invasive bronchopulmonary infections with dissemination

to the central nervous system (CNS) have been reported in immunocompetent individuals after near drowning in polluted water. On the other hand, **extrapulmonary infections** comprises all the non-systemic infections that aren't mycetomas and affect any tissue or organ except the lungs (*e.g.*, skin, bones, sinuses, muscle, heart, CNS, kidney, liver, eyes, outer ear canal, nails, prostate, esophagi, prostate, etc.).

Disseminated infections are a serious progression of a disease wherein an infection, originally localized, spreads from one area of the body to other organs or systems. Their last stage, often result affecting the entire body (systemic infections) [17]. The respiratory tract and skin lesions are the most common routes of entry for pathogens that can lead towards disseminated infections. Those affectations are usually seen among immunocompromised patients; however, they have also been reported in immunocompetent patients [18]. In both cases, dissemination carries a bad prognosis.

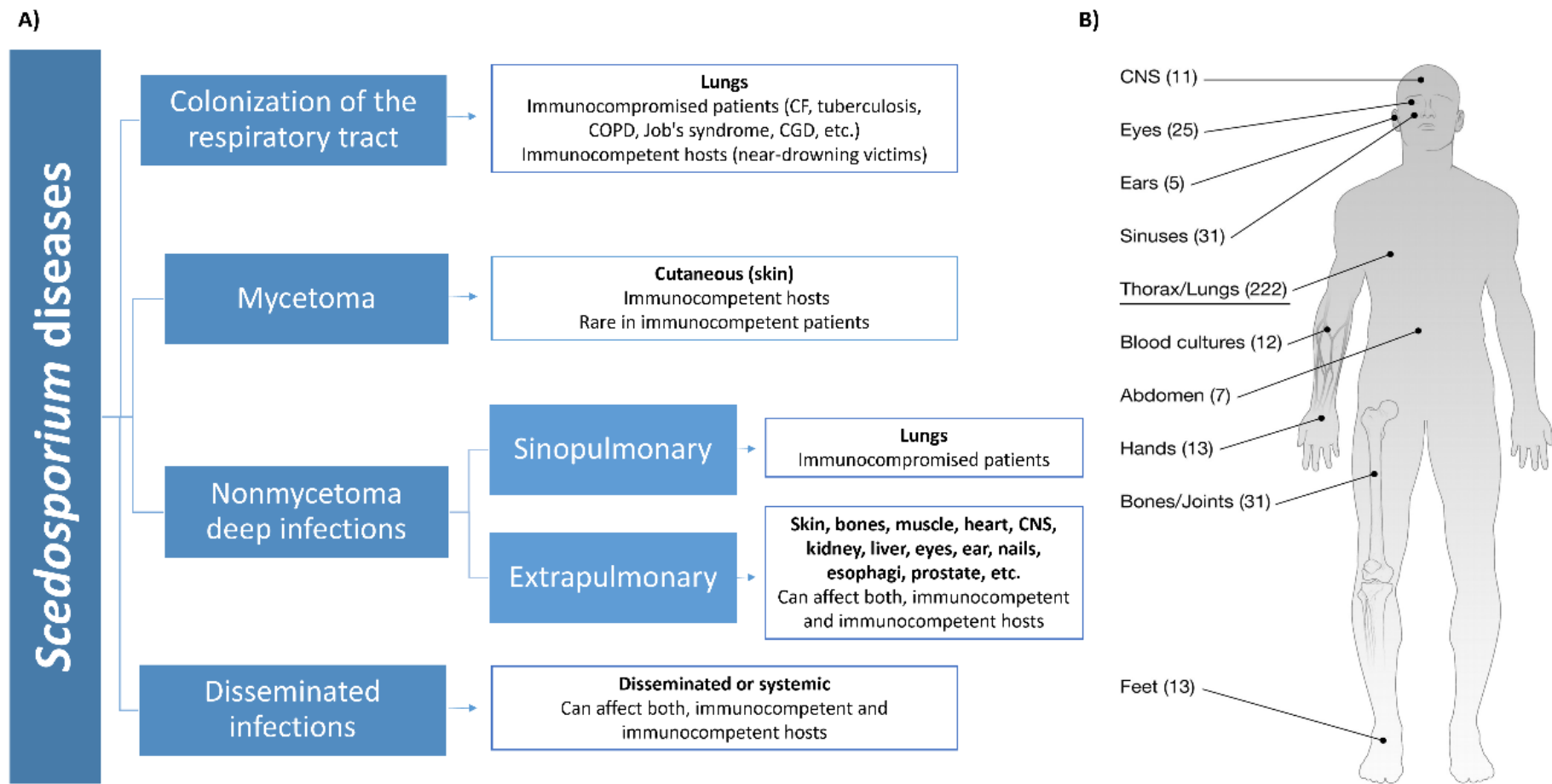


Figure 1.2: *Scedosporium spp* pathogeny. A) Diagram of diseases caused by *Scedosporium* complex species in humans. B) Anatomical origins (sites of infection) of 370 isolates submitted to the Fungus Testing Laboratory at the University of Texas Health Science System from January 2000 to May 2007. Panel B was adapted with permission from Cortez *et al.*, 2008 [11].

1.2.4 Distribution and epidemiological overview

Scedosporium apiospermum is a globally distributed soil fungus that is capable of assimilating numerous aromatic or polycyclic hydrocarbons. This is possibly why it is found mainly in areas affected by men, such as industrial areas, urban parks, roadsides and agricultural lands [10]. Its role in the environment is still limited and there is no real knowledge of its geographical distribution, however, the main affected countries have been identified based on the incidence of infections caused by this pathogen (Figure 1.3) [19].

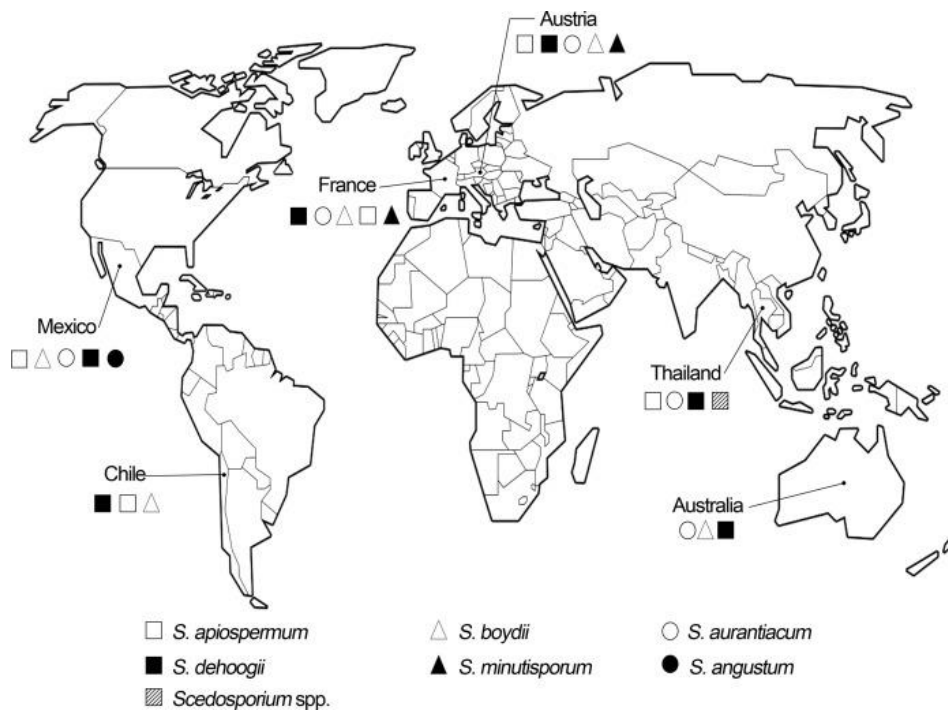


Figure 1.3. Global distribution and relative burden of scedosporiosis. Reprinted with permission from Luplertlop, 2018 [10]

Although there is a large number of diseases caused by *scedosporium* spp, by far the most common are infections of the airways (sinopulmonary infections). In a study published in 2008 by Cortez and colleagues, they analyzed all the isolates of *Scedosporium* from infected patients that were submitted to the Fungus Testing Laboratory of Texas University over a period of 7 years. From the 370 samples considered, 60% (222) correspond to isolates from the thorax/lungs, while the

remaining 40% were divided between CNS, eyes, ears, sinuses, blood cultures, abdomen, hands, bones, joints and feet (Figure 1.2B).

Sinopulmonary infections mainly affect immunosuppressed patients and often result in systemic dissemination. Together with *Fusarium* species and some dematiaceous fungi, the *Scedosporium* complex is responsible for approximately 10% of the mycosis caused by filamentous fungi in hematopoietic cell transplant recipients and up to 19% in solid organ transplant patients [13]. In addition, after *Aspergillus fumigatus*, *Scedosporium* spp are the most common among the filamentous fungi that colonize the respiratory tract of people with cystic fibrosis (CF). Furthermore, because of their propensity to disseminate in case of immune deficiency, the airway colonization by those fungi is considered as a contraindication for lung transplantation, which, in many CF patients, remains as the ultimate line of treatment [20]. Hence, CF patients are considered as one of the main risk populations for scedosporiosis and there is growing concern about them [5].

1.2.5 Cystic Fibrosis; a high-risk population for scedosporiosis

Cystic Fibrosis (CF) is a serious condition caused by an autosomal recessive disorder of the CF Transmembrane Regulator (CFTR). CFTR is an integral membrane glycoprotein that acts as a chloride channel in epithelial cells, and whose dysfunction leads to the production of abnormally viscous bronchial secretions [16,21]. This pathology not only hinders the natural process of mucociliary clearance but also affects the glycosylation pattern of pulmonary mucins [22,23]. In conjunction, those effects favor the binding of pathogens and trigger a colonization process that usually starts in the adolescence and becomes chronic in up to 19% of the patients. This colonization may also lead to a chronic inflammation and sometimes to an allergic bronchopulmonary mycosis [5].

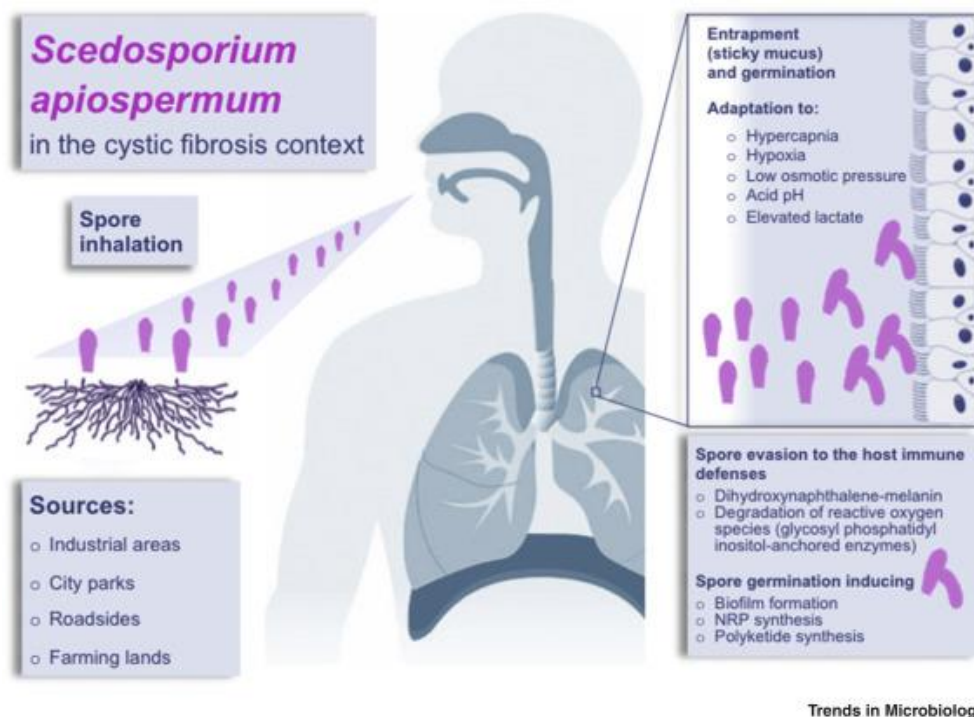


Figure 1.4: *Scedosporium apiospermum* in the Cystic Fibrosis context. Reprinted with permission from Bouchara and Papon, 2019 [19]

1.2.6 Clinical treatment

The treatment of *Scedosporium* infections is challenging because its efficacy depends on timely diagnosis, which is based on morphological detection by microscopy, histological analysis and culture on selective media. This process is time consuming and may lead to false negative results, especially from respiratory secretions of patients with CF because of the common co-colonization by other fungi or bacteria [5]. Furthermore, *Scedosporium* spp display a primary resistance to classical antifungals such as 5-flucytosine, amphotericin B and the first-generation triazole drugs, fluconazole or itraconazole and exhibit a limited susceptibility to the newest generations of antifungal drugs, *e.g.*, echinocandins, voriconazole and isavuconazole [5,24]. Nowadays, the first-line treatment for *Scedosporium* infections involves a combination therapy that includes the use of voriconazole in conjunction with other antifungals [24,25]. However, due to common recurrences, even without interruption of treatment, the recovery rates are poor and mortality remains over 65% while it is almost 100% when

dissemination occurs, a reason why it has aroused special attention despite its emerging character [25,26].

Due to its high incidence and recurrence, the constitutive administration (prophylaxis) of antifungals in immunocompromised people has been proposed. However, some of the side effects of these drugs are dangerous during long-term administration. In this context, therapies aimed to prevent the development scedosporiosis represent a better alternative for these patients.

1.2.7 Host-binding; the first step towards infection

In its adult stage, the *S. apiospermum* mycelium is composed of a network of elongated and cylindrical cells that are surrounded by a cell wall. These fungal filaments are called hyphae and originate from the germination of oval spores (2 to 5 μm) known as conidia when they originate from asexual reproduction [12].

Scedosporium infections begin with conidia adherence to tissues, followed by germination and hyphal elongation [11]. This adherence process allows it to avoid cleansing mechanisms aimed to eradicate the invading pathogens and represents the initial step towards infection [11,27–29].

Conidial adhesion is mediated by cell surface molecules (CSM), including different types of carbohydrates such as polysaccharides and glycoconjugates. The presence and/or abundance of these molecules on the cell surface varies according to the stage of development and is of great relevance to understand fungal pathobiology [30]. Some of the most important carbohydrate CSM described to date for *Scedosporium* species include peptidorhamnomannans (PRMs) [30] α -glucan [31], melanin [32], ceramide monohexosides [30,33], N-acetyl-D-glucosamine-containing molecules [34] and mannose/glucose-rich glycoconjugates [30]. Some carbohydrate binding proteins (lectins) display an essential role during fungal pathogenesis and can be considered CSMs.

1.3 Lectins

1.3.1 Definition and generalities

Lectins are ubiquitous proteins that include at least one domain with the ability of binding, in a reversible and specific way, to glycans without modifying them [35,36]. The spectrum of carbohydrates that can be recognized by lectins is considerably broad and ranges from simple monosaccharides to complex structures containing polysaccharides and glycoconjugates, such as glycoproteins, glycolipids, proteoglycans or glycosaminoglycans (GAGs) (Figure 1.5) [37]. Carbohydrates represent the third alphabet of life and cover the surface of every cell. The information encoded in carbohydrate called glycode can be deciphered by the lectins.

Since lectins are generally capable of agglutinating erythrocytes, they are often referred to as hemagglutinins. However, it is important to clarify that this is an oversimplification, since not all lectins are hemagglutinins nor all hemagglutinins are lectins. This is, some lectins do not display hemagglutination activity because their ligands are not present on the cell surface of red blood cells and not all hemagglutinins agglutinate erythrocytes through the reversible binding of sugars found on their cell surface.

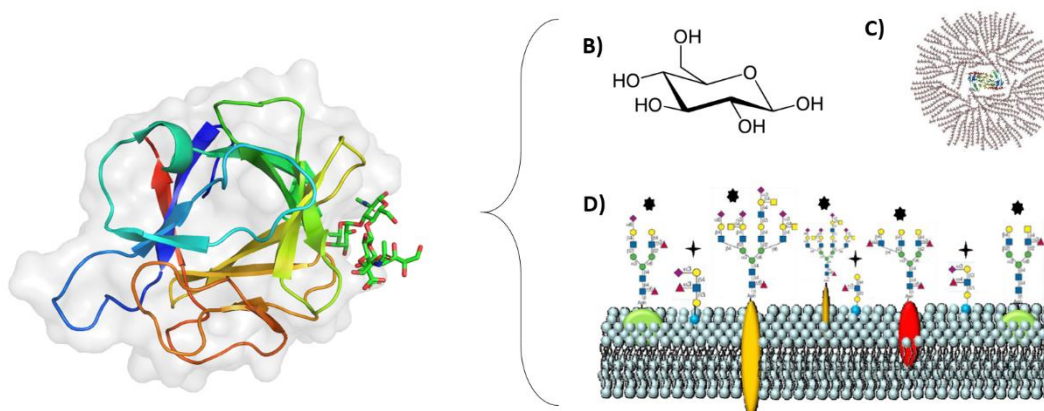


Figure 1.5: Carbohydrates-lectin interactions. A) Structure of the CCL2 lectin from *Coprinopsis cinerea* in complex with sialyl lewis X (PDB 4USO). B-D) glycosyl ligands of lectins, monosaccharide, glycogen and membrane glycoconjugates (from right to left, respectively).

1.3.2 Carbohydrate-binding properties of lectins

Most of the lectins described to date have complex structures and some of them may even include prosthetic groups, *e.g.*, ions and glycosylations. In some cases, these prosthetic groups are essential to exert their biological activity [38], however, it is important to highlight that the carbohydrate-binding properties of lectins are essentially attributed to the Carbohydrate Recognition Domain (CRD), where the binding pocket (BP) is embedded [39]. The selectivity of this interactions is given by i) a hydrogen bond network established with the hydroxyl groups of the sugar and ii) π -CH stacking interactions and iii) a Van der Waals packing, which often includes the coverage of the hydrophobic portion of the sugar by apolar residues of the protein (Figure 1.6) [38,39].

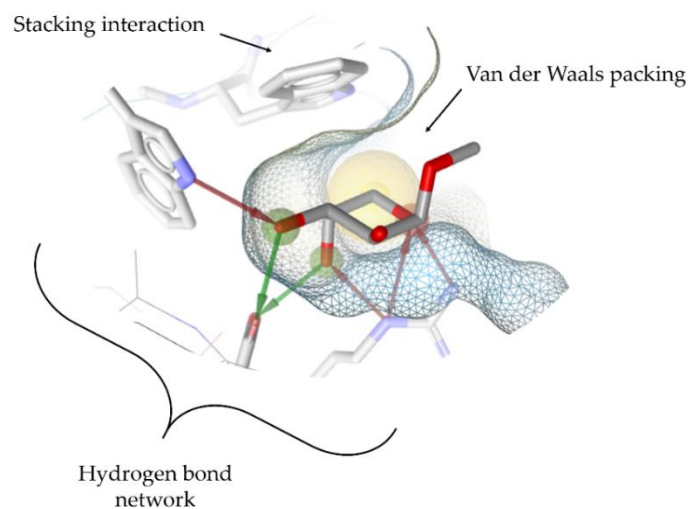


Figure 1.6: Carbohydrate recognition domain of lectins. The main forces that influence CRD selectivity and the hydrophobic pocket are indicated in the diagram. Figure created using LigandScout 4.3.

Lectins tend to present a shallow binding site resulting, usually, in low affinity in the millimolar order in particular for monosaccharides [38]. This low affinity can be compensated by multivalency, meaning the lectins present multiple binding sites. Upon multivalent interactions, there is indeed an extensive enhancement of binding, *i.e.*, avidity, that is a well-known phenomenon for

lectin-glycan interactions. Multivalency can be achieved by the presence of several tandem repeats, by oligomerization of monomers or by a mix of those two (Figure 1.7). A good example of this behaviour is represented by the lectin FleA from *A. fumigatus*. This lectin presents six tandem repeats and forms homodimers of 6-bladed β -propellers with 6 binding sites each (Pseudo D6, Figure 1.7) [40].

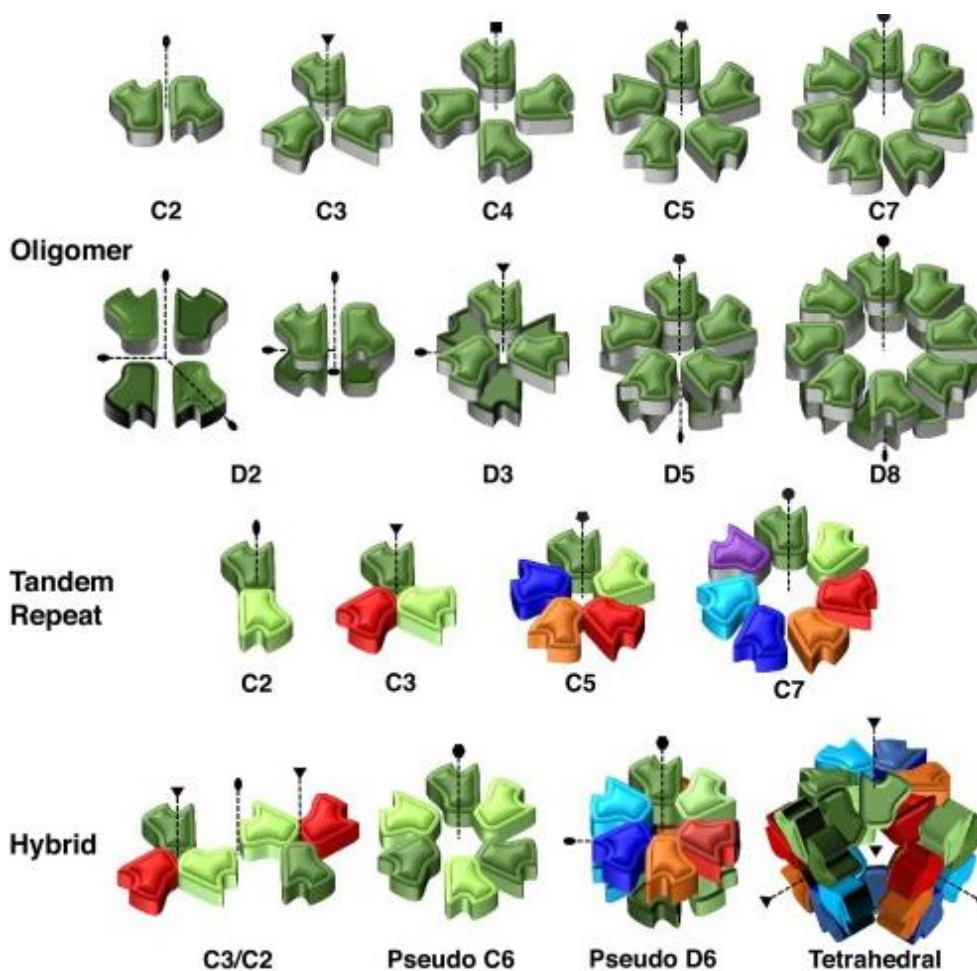


Figure 1.7: Schematic representation of different types of symmetry observed in lectins. The symmetry family and symmetry axis are noted for each schematic representation. Reprinted with permission from Notova *et al.*, 2020 [41]

1.3.3 Specificity of lectins

The set of glycans that lectins can recognize is very wide. Some of them have affinity for single monosaccharides such as the case of GNA from *Galanthus nivalis*, which recognize mannose [42], while in other cases, a specific

arrangement of carbohydrate linkages is necessary to achieve lectin binding. PHA:L from the common bean (*Phaseolus vulgaris*) is a clear example of this condition since its minimal structural unit for high-affinity binding is the pentasaccharide Gal(β 1-4)GlcNAc(β 1-2)[Gal(β 1-4)GlcNAc(β 1-6)]Man [43]. There are also some lectins with affinity for more than one monosaccharide, such as the members of the “Glucose/mannose” or “Galactose/N-acetyl-D-galactosamine” subfamilies [44]. In those examples, lectins specificity can be considered degenerated since the interaction with different sugars is established through the same binding site. On the other hand, the recently discovered subfamily of “superlectins” also displays affinity for more than one type of sugar, however, their dual recognition is due to the presence of multiple CRDs with specificity for different sugars [45]. BC2L-C from *Burkholderia cenocepacia* was the first superlectin to be characterized [46]. Figure 1.8 shows some additional examples of glycans recognized by different lectins.

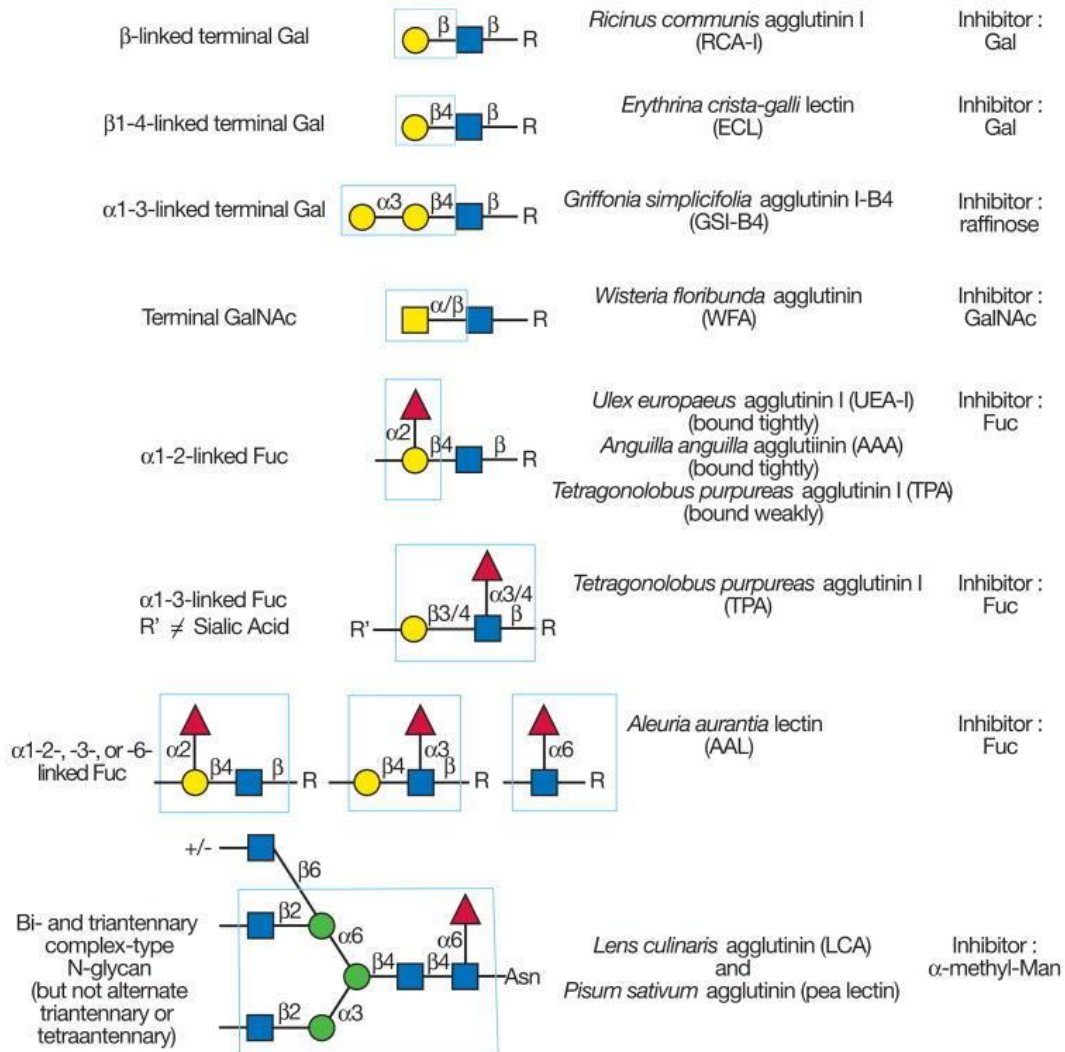


Figure 1.8: Examples of glycan determinants recognized by different lectins. The determinants required for binding are indicated in the boxed areas. Yellow circles represent galactose, blue squares represent N-acetylglucosamine, red triangles represent fucose, green circles represent mannose. Reprinted with permission from Varki *et al.*, 2009 [62].

1.3.4 Classification of lectins

Lectins are distributed among all kingdoms including viruses, protozoa, bacteria, animals, plants and fungi [47,48]. They are a very heterogeneous group of proteins, that vary greatly from each other, both in structure and in function [39]. Therefore, their classification is complex to address and there is not a global consensus in the scientific community. Originally, lectins classification was based on taxonomic origin (*e.g.*, plants lectins, animal lectins, viral lectins, bacterial lectins and fungal lectins). However, incompatible functions of this classification have aroused the need for more functional classifications. Over the last years,

different proposals for classification emerged based on common characteristics such as cell location, carbohydrate recognition specificity, functions, structure, among others [49]. For some of those, there are very well defined parameters that allow consider lectins from all living organisms, however, some others, are usually established for a defined set of organisms [50].

With the dramatic increase in the number of lectin structures released over the last decade, the recently created UniLectin3D platform has proposed a classification based on their fold. This is particularly interesting since there is a wide variety of fold and arrangements very well defined among organisms that, in addition to a classification criterion, allow the prediction of specificity and biological functions based on homology. UniLectin3D comprises a curated database of 3D lectin structures that provides a regularly updated classification and predicts new hypothetical lectins from genome sequence databases [51].

1.3.5 Fungal lectins: roles and biotechnological applications

Fungi are an invaluable source for novel lectins with unique specificities that, are essential for their survival and development. Although, they are poorly studied, in comparison those from plants and animals, a broad spectrum of biotechnological and biomedical applications is emerging for those proteins.

Fungal lectins participate in diverse biological roles (Figure 1.9). For example some are involved on defense mechanisms due their toxicity against insects, virus and worms [52], while others display functions related to metabolism or are key elements during the stages of dormancy, growth and morphogenesis [53]. They might also serve as storage proteins or participate in the process of anchoring, necessary for both, ectomycorrhizal symbiosis and infection by pathogenic fungi [53,54]. They play an important role in fungal innate immunity, the sole line of defense to recognize foreign epitopes from invaders.

On the other hand, fungal lectins are also valuable tools in glycobiology, since they can be coupled to resins for the isolation of industrially-relevant glycans [55,56]. Due to their high specificity, some of those proteins are being used as biomarkers to detect/monitor the changes that occur in the glycoconjugates of the surface of cell membranes during physiological and pathological processes [52] and, some others, have shown immunological and antitumor activity [57,58]. Furthermore, since some fungal lectins are important mediators of host-pathogen interactions, they represent a valuable tool for biomedical research on antifungals development. [28,59–66].







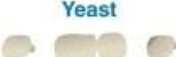





ROLES	ORIGIN	RESEARCH AND BIOTECHNOLOGICAL APPLICATIONS
<ul style="list-style-type: none"> ➤ Storage proteins  ➤ Growth and morphogenesis ➤ Parasitism/infections <ul style="list-style-type: none"> ➤ Host recognition  ➤ Adhesion ➤ Molecular recognition <ul style="list-style-type: none"> ➤ Mycorrhization ➤ Lichens ➤ Defense   ➤ Cell flocculation/Mating process 	<p style="text-align: center;">Mushrooms</p>  <p style="text-align: center;">Microfungi</p>  <p style="text-align: center;">Yeast</p> 	<ul style="list-style-type: none"> ➤ Glycoproteins and carbohydrates purification  ➤ Glycomics studies ➤ Biomarkers ➤ Cancer research <ul style="list-style-type: none"> ➤ Markers and diagnosis  ➤ Immunostimulating ➤ Antiproliferative/Antitumor ➤ Antiviral  ➤ Insecticide/Vermicide  ➤ Targeted drug delivery 

Figure 1.9: Roles and potential applications of fungal lectins. Reprinted with permission from Varrot, *et al.*, 2013 [52].

1.4 The anti-adhesive approach

1.4.1 Host-binding; a prerequisite for infection

Host recognition and anchoring are the first steps in infection processes. They are carried out through the recognition of determinants exposed on the surface of human tissues by microbial surface molecules and is critical for colonization [67]. The main function of adhesion is to help pathogens to avoid the human clearance mechanisms that aim at their eradication, such as their elimination through urine, sweat and mucous membranes. In addition, it also provides them with better

access to sources of nutrition and creates the right conditions for their growth [48]. In fungi, this process is fundamental for differentiation of the conidia towards the hypha and, in some cases, it is also a universal prerequisite for starting the transcription of their virulence factors of second line, including delivery of toxins and biofilm formation [28,68].

Because host-binding is so crucial for infection, pathogens have devised a vast repertoire of attachment mechanisms. In general, the host-anchoring process can be visualized in three very well-defined stages adsorption, initial adhesion and high-affinity interactions. During the first stage, the microorganisms bind to the host cell surface through weak nonspecific interactions. These interactions are transient and are determined by general physicochemical properties of cell surfaces, such as charge and hydrophobicity. Then, the process is reinforced by the initial adhesion and high affinity interactions, which are mediated by specific interactions involving components exposed on the cellular surfaces of both, host and pathogen cells [28].

Since carbohydrates are abundantly distributed in human cell membranes, microbial lectins are often involved in the initial mechanism of host-adhesion. In fact, their role is so fundamental that it has been observed that mutants of pathogenic bacterial strains with lectin deficiency are often unable to initiate their infectious process [69]. Therefore, lectins can be considered as virulence factors indispensable for host-pathogen interactions during the early stages of infection and represent promising therapeutical targets for the development of anti-adhesives molecules.

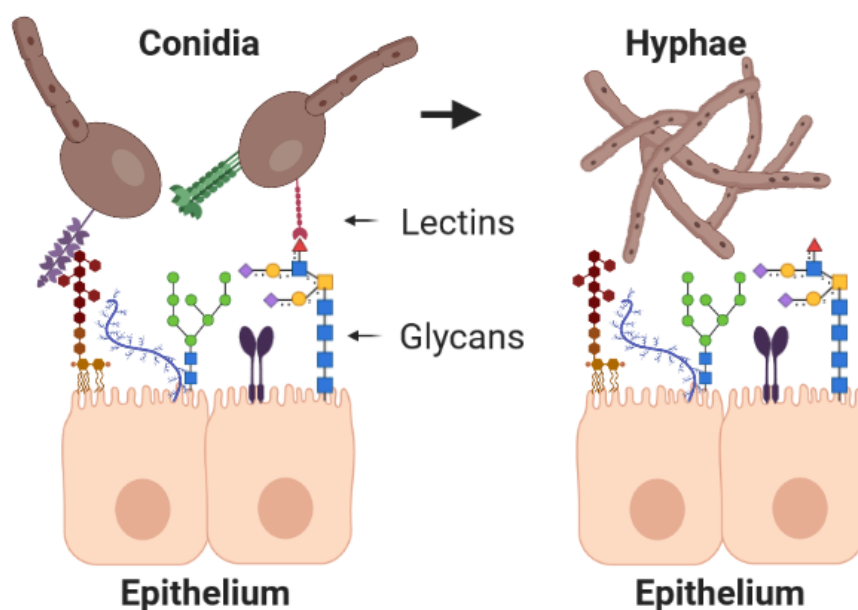


Figure 1.10: Fungal adhesion to epithelial cells. Figure generated using Biorender.com.

1.4.2 Anti-adhesive therapy

The alarming increase in the incidence of multidrug-resistant pathogens is an urgent public health problem. Infections are becoming one of the leading causes of death worldwide and it is increasingly difficult to develop new generations of drugs at the speed that the situation demands [70]. Thus, there is an urgent need for therapeutic strategies capable of resisting the ever-increasing repertoire of microbial resistance mechanisms [28].

Since attachment is strict condition for virulence, it is logical to assume that the blocking of host-pathogen interactions will prevent the development of infections. Hence, a new approach that proposes the use of antagonists of such interactions has emerged as a valuable alternative for infectious diseases. This has been called the anti-adhesive therapy.

Anti-adhesive therapy represents an efficient substitute and/or complement to conventional antimicrobials because it explores different pharmacological routes and the development of resistance is rare and slow [69]. This is probably one of its biggest advantages and is due to the fact that its mechanism of action does not

intended to kill the pathogen or stop its cell cycle. Consequently, there is no selective pressure for resistant pathogens to survive and multiply, inheriting their resistance to the progeny [28]. Furthermore, this approach is particularly promising as a treatment for fungal infections because the administration can be local and directed, therefore the expected side effects on non-target tissues are less than those caused by conventional antifungals. This may allow its implementation as prophylactic therapy for immunosuppressed patients [29,67].

1.4.3 Carbohydrates and glycomimetics as microbial anti-adhesives

Since it is well known that some microbial lectins participate in the initial adhesion to human epithelial surfaces, it has been proposed that carbohydrates could be used to inhibit pathogen anchoring infection during the early stages of infection. The first indications of the protective effect of sugars against pathogens, dates back to the late 19th century, when it was observed the mortality rate of breast-fed infants was 7 times lower than infants fed with cow's milk [71]. Later, it was discovered that this survival phenomenon was due to the fact that the oligosaccharides present in human breast milk (HMOs) hinder the anchoring of pathogens and protect infants during lactation [72]. Nowadays, the protective role of HMOs in newborn babies is well established and they are widely used in the formulation of nutritional products.

The use of sugars as protective therapy against infections in animals was first reported in 1979 by Aronson and coworkers [73]. In their published work, they report that the administration of α -D-mannopyranoside to mice, prevented the colonization of their urinary tract by *Escherichia coli*. Subsequently, many studies have confirmed the feasibility of using saccharides to prevent infections caused by different pathogenic microorganisms. Table one collects some examples of these studies.

Table 1.1. Carbohydrates that prevent infection *in vivo*

Pathogen	Target tissue	Inhibitor	Reference
<i>C. jejuni</i>	Intestinal	milk oligosaccharides	Ruiz <i>et al.</i> , 2003 [74]
<i>E. coli</i> (type 1 fimbriated)	Urinary	Man α Man	Aronson <i>et al.</i> , 1979 [73]
<i>E. coli</i> (P fimbriated)	Urinary	Gal α 4Gal	Edén <i>et al.</i> , 1982 [75]
<i>E. coli</i> K99	Intestinal	NeuAc(α 2-3)Gal β 4Glc	Mouricout <i>et al.</i> , 1990 [76]
<i>K. pneumoniae</i>	Respiratory	Man	Fader <i>et al.</i> , 1980 [77]
<i>H. pylori</i>	Stomach	sialyl-3P-LacNAc	Mysore <i>et al.</i> , 1999 [78]
<i>S. flexneri</i>	Eye	Man	Izhar <i>et al.</i> , 1982 [79]
<i>S. pneumoniae</i>	Respiratory	sialyl-3P-GalL(1-4)LacNAc	Idänpään <i>et al.</i> , 1997 [80]
<i>S. sobrinus</i>	Oral cavity	oxidized α 1,6glucan	Wang <i>et al.</i> , 1996 [81]
<i>S. pyogenes</i>	Pharynx	hyaluronan	Cywes <i>et al.</i> , 2000 [82]

Modified from Ofek *et al.*, 2003 [29]

Although sugars have been shown to be effective in hindering microbial adhesion, lectin-carbohydrate interactions are generally very weak and some of these molecules are susceptible to enzymatic degradation [68]. Therefore, an evolved approach to lectin-targeted antiadhesive therapy propose the use of carbohydrate mimics (glycomimetics) as a replacement for natural sugars. These compounds mimic the bioactive function of carbohydrates while displaying better drug-like properties, which leads to higher metabolic stability and higher selectivity [83]. However, the development of efficient and specific glycomimetics requires prior knowledge about the lectin that is being targeted. This is particularly problematic in the case of emerging pathogens whose anchor lectins remain anonymous and evidences the imperative need for their identification.

1.5 *Aspergillus fumigatus* lectins

A. fumigatus is also an ubiquitous opportunistic saprophytic mold that is the most prevalent airborne fungal pathogen causing infections in humans. It is responsible for 16 million bronchopulmonary infections a year in immunocompromised patients with high mortality rate and resistance to actual antifungals [84]. Its lectinome has been under study in our research group for a decade. Two lectins named FleA and AFL6 have been especially under study.

FleA is a lectin found on the surface of the *A. fumigatus* conidia with a strong pro-inflammatory effect on human bronchial cells. Its structure is a six-bladed β -propeller homodimer with 6 binding pockets (BP) per protomer (Figure 1.11A) [54]. FleA is specific for fucosylated carbohydrates, notably, human blood group antigens (BGA). Those can be found on airway mucins and macrophage glycoproteins and it has been suggested that their recognition by FleA would mediate the binding of *A. fumigatus* conidia to host cells. The BPs all present a triad necessary for fucose recognition and have diverged in FleA to have low and high affinity BPs [34]

AFL6 belongs to the Cyanovirin-N homologue lectin subfamily (CNVH) in fungi (Figure 1.11B, unpublished work). The biological roles of this subfamily remain unclear, but it has been suggested that some of them play an important role in pathogenesis [85]. It seems to have a biological relevance as it is constitutively expressed during the whole fungal life cycle. CNVHs have caught the attention of the scientific community for being homologous to Cyanovirin-N, a powerful antiviral protein isolated from the cyanobacteria *Nostoc ellipsosporum* [86]. CN-V presents a tandem repeat leading to the formation of two pseudo-symmetrical domains with one BPs per domain recognising mannosides but in fungal CNVH, usually only on BP seems functional and the specificity is still under debate [59].

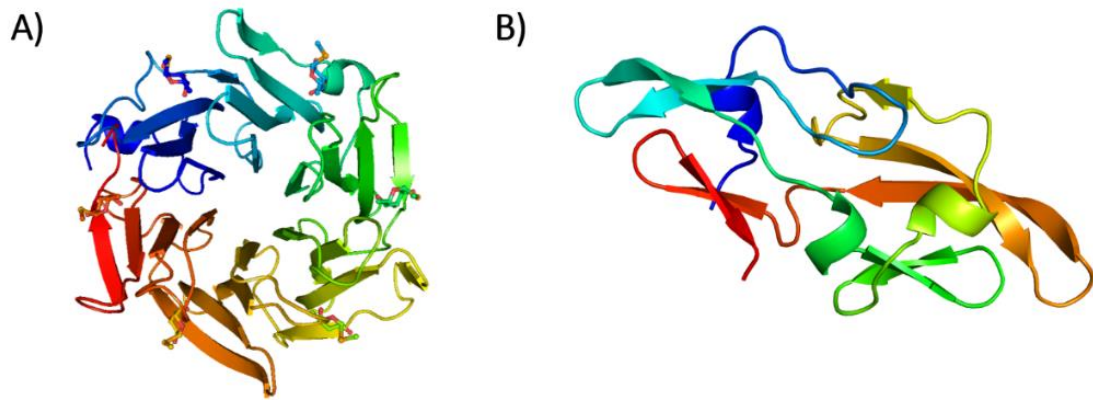


Figure 1.11: Overall structure of FleA and AFL6. A) crystal structure recombinant FleA from *A. fumigatus* in complex with alpha-L-fucopyranoside (PDB 4U4D). B) crystal structure recombinant AFL6 from *A. fumigatus* (unpublished data). Figure created using PyMOL Version 1.8.4 (Schrödinger).

2. AIM OF THE THESIS AND OBJECTIVES

The overall goal of this work was to contribute to the general understanding of the host-pathogen interactions by the identification and characterization of new lectins from the emerging opportunistic fungal pathogen *Scedosporium apiospermum*. This has been addressed by two different strategies; the first strategy comprises the *in-silico* prediction of putative lectins encoded in the *S. apiospermum* genome by data mining, followed by their recombinant production and subsequent characterization. The second strategy lies in the identification of native lectins directly from fungal protein extracts of *S. apiospermum*. In this latter approach, we have also included the analysis of protein extracts from another medically relevant pathogenic fungi, *Aspergillus fumigatus*.

A second aim of this project was to develop a new tool that allows the identification and purification of unknown lectins directly from crude extracts of proteins without the need for fluorescent labelling. This tool assembles the fusion of the three classic steps of lectin purification train: clarification, concentration and purification, in a single operating unit. It also allows the simultaneous purification of lectins with different specificities in a simple and fast process. It can also be used as a "mini glycan array" for the study of carbohydrate lectin interactions.

To achieve those main goals, 4 objectives were defined:

- 1- To identify putative lectins from the genome of *S. apiospermum* by data mining.
- 2- To carry out the heterologous production of the lectins identified and proceed with their biochemical and structural characterization.
- 3- To identify native lectins from protein extracts of *S. apiospermum*.
- 4- To develop a new tool for purification and identification of lectins from crude extracts.

RESULTS

Chapters III-VII

The following five chapters describe the main results of the project, which are schematically summarized in the following diagram.

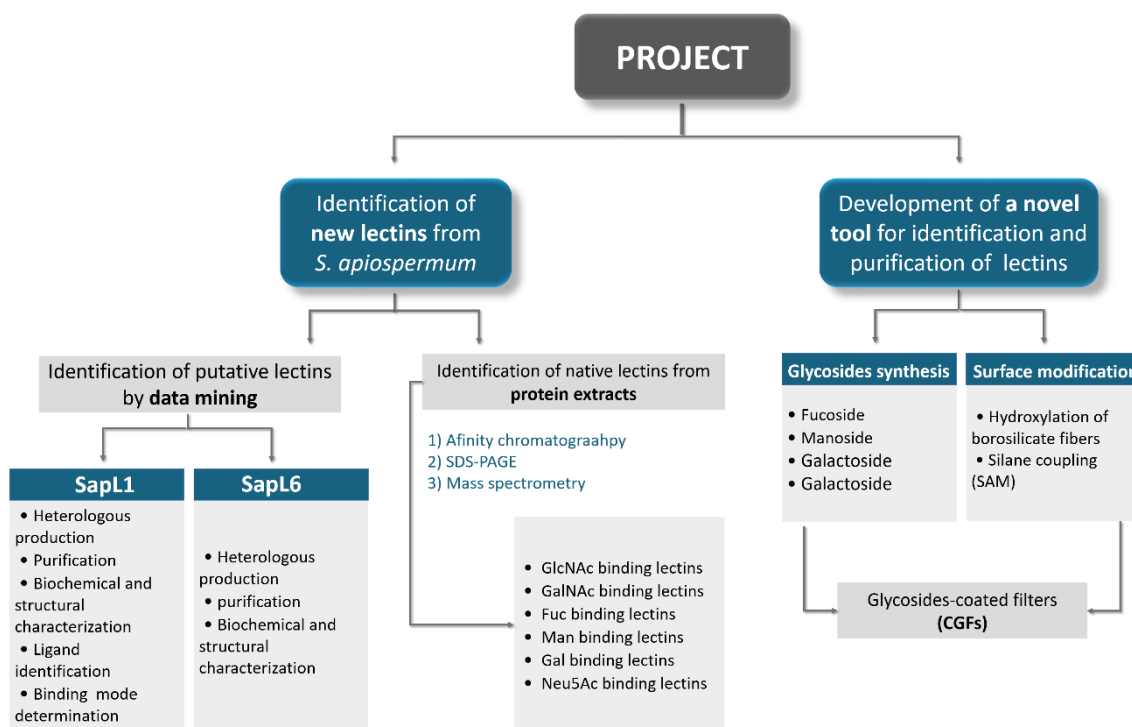


Figure 2: Thesis results overview. Figure created with Microsoft Office 365.

3. BIOCHEMICAL AND STRUCTURAL STUDIES OF TARGET LECTIN SapL1 FROM THE EMERGING OPPORTUNISTIC MICROFUNGUS *S. apiospermum*

3.1 Summary

In this chapter, we present the first report of the identification and characterization of a lectin from *S. apiospermum*. This lectin has been named SapL1 and it is homologous to the conidial surface lectin from *Aspergillus fumigatus* (FleA), which is known to be involved in the adhesion to host glycoconjugates present in human lung epithelium. Therefore, the discovery and characterization of SapL1 contribute to the understanding of glycosylated surface recognition by this fungal pathogen and might guide the development of antiadhesive glycodrugs by targeting its inhibition.

The main findings of this part of the project are described in a paper that has been accepted for publication in the Scientific Reports Journal. This work includes a detailed strategy to achieve the soluble expression of SapL1 in *E. coli*, its biochemical characterization, an analysis of its specificity and affinity by Glycan array and Isothermal Titration Calorimetry (ITC), as well as the structural characterization of its binding mode by X-ray crystallography. The publication of this work was used for this chapter. A graphical abstract is shown in figure 3.1.

SapL1: A NEW TARGET FOR THE DEVELOPMENT OF ANTIADHESIVE GLYCODRUGS AGAINST *S. apiospermum*



START

1. Problem: *S. apiospermum* bronchopulmonary infections are life-threatening.

2. Need: a new therapy to avoid its adhesion to the lungs by blocking its host binding lectins.

Anti-adhesive therapy to block anchoring lectins

Blocking lectins
Block adhesion
Avoid infections

3. Problem #2: *Scedosporium* lectins are unknown.

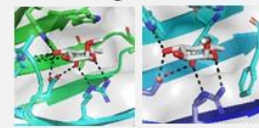
Which lectins?

4. Our strategy: identify such lectins.

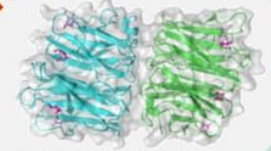
✓4.6 Guided design of inhibitors



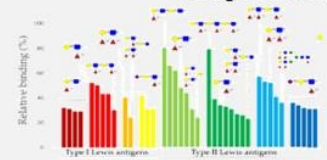
✓4.5 Binding mode determination



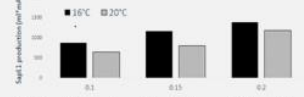
✓4.4 Structure (X-Ray)



✓4.3 Ligand identification



✓4.2 Production and characterization



✓4.1 Target (host binding lectin) identification

Conclusions

- ❑ We identified and characterized SapL1: a new lectin from *S. apiospermum* involved in host recognition.
- ❑ SapL1 is specific for fucosylated carbohydrates and recognizes all blood group types.
- ❑ We have identify the specific interactions responsible for SapL1 binding specificity.
- ❑ This information contributes to understand the recognition of human glycosylated surfaces by *S. apiospermum* and now is leading the design of a novel therapy against this pathogen



Figure 3.1: Graphical abstract of project 1. Production and characterization of SapL1. Figure created with Microsoft Office 365.



OPEN

Biochemical and structural studies of target lectin SapL1 from the emerging opportunistic microfungus *Scedosporium apiospermum*

Dania Martínez-Alarcón^{1,2}, Viviane Balloy³, Jean-Philippe Bouchara⁴, Roland J. Pieters² & Annabelle Varrot^{1,5✉}

Scedosporium apiospermum is an emerging opportunistic fungal pathogen responsible for life-threatening infections in humans. Host-pathogen interactions often implicate lectins that have become therapeutic targets for the development of carbohydrate mimics for antiadhesive therapy. Here, we present the first report on the identification and characterization of a lectin from *S. apiospermum* named SapL1. SapL1 was found using bioinformatics as a homolog to the conidial surface lectin FleA from *Aspergillus fumigatus* known to play a role in the adhesion to host glycoconjugates present in human lung epithelium. In our strategy to obtain recombinant SapL1, we discovered the importance of osmolytes to achieve its expression in soluble form in bacteria. Analysis of glycan arrays indicates specificity for fucosylated oligosaccharides as expected. Submicromolar affinity was measured for fucose using isothermal titration calorimetry. We solved SapL1 crystal structure in complex with α -methyl-L-fucoside and analyzed its structural basis for fucose binding. We finally demonstrated that SapL1 binds to bronchial epithelial cells in a fucose-dependent manner. The information gathered here will contribute to the design and development of glycodrugs targeting SapL1.

During the last decades, an increased incidence of invasive infections, especially in immunosuppressed patients, has been caused by previously rare fungal pathogens such as species from the *Scedosporium* genus^{1,2}. This genus comprises more than ten worldwide distributed soil saprophyte species whose taxonomy has been previously described based on molecular phylogenetic^{3,4}. *Scedosporium* species may lead to bronchitis and allergic bronchopulmonary mycoses as well as severe disseminated infections through inhalation of conidia. They rank second among the filamentous fungi colonizing the respiratory tract of cystic fibrosis (CF) patients, after *Aspergillus fumigatus*⁴. Together with *S. boydii* and *S. aurantiacum*, *S. apiospermum* is the most common *Scedosporium* species able to chronically colonize the lungs of CF patients^{4,5}. Because of their propensity to disseminate in case of immune deficiency, this fungal colonization of the airways is considered in some centers as a contraindication to lung transplantation that is the ultimate treatment in CF⁶. Besides, infections by *Scedosporium* species have also been reported in immunocompetent patients⁷.

The treatment of *Scedosporium* infections is challenging because its efficacy depends on the biological diagnosis which is a time-consuming process. The common co-colonization by other fungi or bacteria may lead to false negative results, especially from respiratory secretions of patients with CF⁴. Furthermore, *Scedosporium* species display a primary resistance to most current antifungals such as amphotericin B or first-generation triazole drugs as fluconazole or itraconazole and exhibit a limited susceptibility to the newest generations of antifungal drugs, *i.e.* echinocandins and voriconazole^{4,8}. Nowadays, the first-line treatment involves combination therapy but following common recurrences and even without interruption of treatment, the recovery rates

¹Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France. ²Utrecht University, 3584 CG Utrecht, The Netherlands. ³Sorbonne Université, Inserm, Centre de Recherche Saint-Antoine, F-75012 Paris, France. ⁴UNIV Angers, Université de Bretagne Occidentale, SFR ICAT, CHU, Host-Pathogen Interaction Study Group (GEIHP, EA 3142), Institut de Biologie en Santé-IRIS, Angers, France. [✉]email: annabelle.varrot@cermav.cnrs.fr

are poor and mortality remains over 65% while it is almost 100% when dissemination occurs, reason why it has aroused special interest^{8–10}.

Scedosporium infections begin with conidial adherence to tissues, followed by germination and hyphal elongation¹¹. The adherence allows it to avoid cleansing mechanisms aimed to eradicate the invading pathogens and represents the initial step towards infection^{11–14}. Conidial adhesion is mediated by cell surface molecules (CSM), including carbohydrates where some of the most important described to date for *Scedosporium* species include peptidorhamnomannans (PRMs)^{15–17}, α -glucans¹⁸, melanin¹⁹, ceramide monohexosides^{17,20}, N-acetyl-D-glucosamine-containing molecules²¹ and mannose/glucose-rich glycoconjugates¹⁷. Their presence and/or abundance on the cell surface vary according to the stage of development and is of great relevance to understand fungal pathobiology¹⁷. The carbohydrate binding proteins known as lectins also act as CSMs and were shown to have an essential role during pathogenesis in the host recognition and adhesion process²². They became drug targets for the development of carbohydrates related molecules as antiadhesives drugs^{12–14,23–26}. The anti-adhesive therapy approach is particularly promising since it does not kill the pathogen nor arrest its cell cycle. Consequently, resistance frequencies are very low^{13,23} and expected side effects in non-target tissues are lower than those caused by conventional antifungal compounds. This may allow its implementation as a prophylactic therapy for immunocompromised patients¹⁴.

Due to the emerging character of *S. apiospermum*, there is very limited information on its mechanisms of recognition and anchoring to the host. Furthermore, lectins from this microorganism have not been characterized to date, which hinders the development of an anti-adhesive therapy. Conversely, other filamentous fungal pathogens were investigated leading to the identification and characterization of their host binding modes. For example, in *A. fumigatus*, which is a saprophytic mold also responsible for bronchopulmonary infections in receptive hosts, the lectin FleA (or AFL) was identified and revealed to play a role in host–pathogen interactions²⁷. FleA is a six-bladed β -propeller homodimer located on the conidial surface that recognizes human blood group antigens and mediates *A. fumigatus* binding to airway mucins and macrophages glycoproteins in a fucose-dependent manner²⁸. In healthy individuals, this anchorage is critical for the mucociliary clearance process and the macrophagy; in fact, it has been described that fleA-deficient (Δ fleA) conidia are even more pathogenic than wild type (WT) conidia, both in healthy and chemically immunocompromised mice^{28,29}. However, CF patients represent a very particular scenario because the mucus in their lungs is thicker, in relation with mucin overproduction and its high content in calcium ions, which modulates the supramolecular organization of mucin MUC5B by protein cross-linking^{29,30}. This contributes to the suboptimal transport properties of mucus and compromises the pathogen clearance mechanisms²⁹. Furthermore, the aberrant glycosylation in CF patients causes, among other things, an increase in the abundance of sialyl-Lewis X and Lewis X determinants in lung mucins^{31,32}, which is translated as an increase in the fucose content. Therefore, in this context, the FleA (and homologous proteins) anchoring to the mucus layer plays an essential role in the colonization of the CF lungs by *A. fumigatus*.

Here, we have used the recently sequenced genome of *S. apiospermum*³³ to identify a putative homologue of FleA that we have called SapL1 for *Scedosporium apiospermum* Lectin 1. The present report comprises SapL1 identification, its expression in soluble form in bacteria, an analysis of the fine specificity and affinity of the recombinant protein, as well as its structural characterization by X-ray crystallography and fucose-dependent binding to bronchial epithelial cells by fluorescence microscopy.

Results

Production and purification of SapL1. The hypothetical protein XP_016640003.1 (EMBL accession number), encoded by SAPIO_CDS9261 and from now on referred as SapL1, was identified through data mining using the FleA (pdb entry 4D4U³⁴) sequence as bait into the genome of the reference strain *S. apiospermum* IHEM 14462³³. During the sequence analysis, we found that the first 74 amino acids of the putative protein (Uniprot A0A084FYP2) represent a very disordered region that is not present in any other related protein. We suspected that this peptide could be due to an error during the genome annotation. To verify it, we tried to get the gene from total mRNAs extracted from mycelium by 5' Rapid Amplification of cDNA Ends (5' RACE) using the following primers GSP1-SapL1 5'-TTAAGCAGGGGGCAGAACAGC-3' and GSP2-SapL1 5'-CCA ACGACCCAGCCAGAGTTCC-3' but we were unsuccessful. This could be due to the fact that the gene was not expressed in the conditions tested. We did not have access to total mRNAs from other morphological stages of the microfungus. Thus, we decided to focus on the identified carbohydrate recognition domain (CRD) of SapL1 starting from methionine-75 (Fig. S1). The SapL1 coding sequence (75–369) was fused to an N-terminal 6xHis tag cleavable by the Tobacco Etch virus protease (TEV) under regulation of *trc* and *T7* promoters into pProNde and pET-TEV vectors, respectively (Fig. 1A). Expression was performed in *Escherichia coli* and purification was carried out using immobilized metal affinity chromatography (IMAC). Unfortunately, the original expression yield with both vectors (~ 0.35 mg·L⁻¹ of culture) was too low to proceed with characterization studies. Therefore, we explored new alternatives to enhance the expression. First, the thioredoxin protein (Trx) as well as 6-His tag and TEV cleavage site were fused at the N-terminus of SapL1 by subcloning into the pET32-TEV vector (Fig. 1A). This strategy substantially increased the SapL1 production yield but most of the protein remained insoluble as part of inclusion bodies (Fig. S2).

A wide range of expression conditions were subsequently assayed to achieve sufficient soluble expression by modification of various parameters such as growth temperature, host strain, inducer concentration, optical density of the culture at induction, culture duration, etc. Sixty-nine different sets of parameters were assayed (Table S1) and it was possible to improve the yield up to 4 mg·L⁻¹ (Fig. 1B). The best set of conditions for SapL1 expression was using *E. coli* strain TRX, pProNde vector, LB medium, growth at 37 °C and 160 rpm until OD₆₀₀ = 0.4, before a switch of the temperature to 16 °C and overnight induction at OD 0.8 with 0.05 mM IPTG and 1% L-rhamnose (Rh).

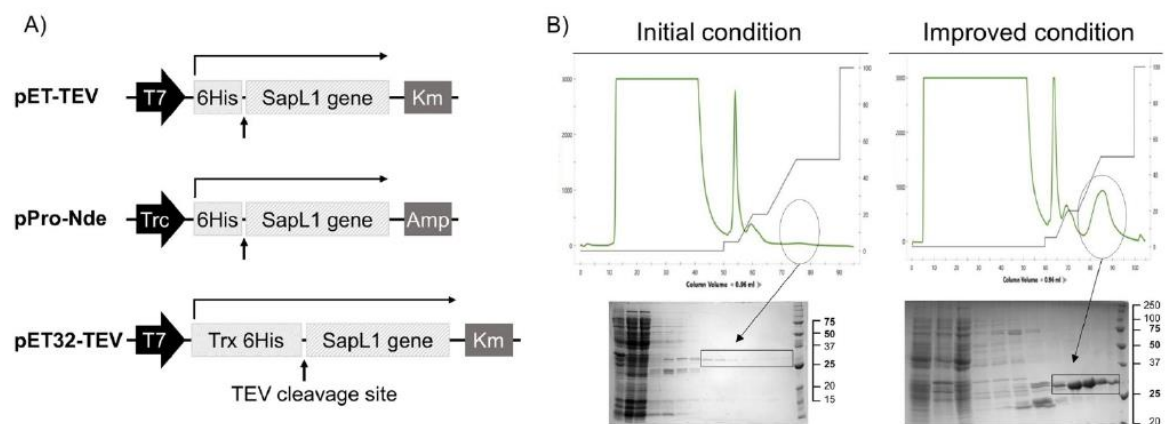


Figure 1. SapL1 production. (A) Schematic representation of the genetic constructs used for the expression of SapL1. (B) Representative chromatograms of SapL1 purification before and after the process with their respective profile on 15% SDS–polyacrylamide gels (insoluble fraction, soluble fraction, flow through, washing and elution and molecular weight marker, from left to right respectively). The fractions containing SapL1 are delimited.

Rhamnose influence on SapL1 solubility. Interestingly, during the optimization of the expression conditions, we found that rhamnose (Rh) plays an essential role in the solubility and stability of SapL1. Therefore, we performed a new set of experiments to demonstrate the correlation between the amount of protein recovered after purification and the Rh concentration in the media. For this, we investigated SapL1 expression under the previously described conditions modified by supplementation of the culture medium with different concentrations of Rh (0.1%, 0.15%, 0.2%, 0.5% and 1%). Purification parameters were set to obtain high purity product and kept identical for all experiments. Figure 2 shows the chromatograms obtained for representative concentrations accompanied by their respective SDS-PAGE profile. To quantify SapL1 expression in these experiments, we integrated the area under peaks corresponding to the protein (Fig. 2, black bars). Thus, we confirmed that SapL1 recovery is directly proportional to the Rh concentration in the culture medium. Then, the experiment was repeated at three different concentrations (0.1%, 0.15%, 0.2%) using 20 °C as the induction temperature, and production behavior was similar to that found at 16 °C, but with lower performance.

Additional experiments also showed that SapL1 expression can be induced at high concentration of rhamnose in a dose-dependent manner, even without addition of IPTG when the pProNde vector is used (data not shown).

Biochemical characterization. We assayed the thermal stability of SapL1 in 26 different buffers in a pH range of 5 to 10 through Thermal Shift Assay (TSA). The most suitable condition for this protein was MES buffer 100 mM pH 6.5, where a single denaturing event at T_m of 55 °C was observed (Fig. S3). Then, to estimate the molecular size and oligomerization state of the native protein, we performed size exclusion chromatography using an ENrich™ SEC 70 column (Bio-Rad) and 20 mM MES, 100 mM NaCl, pH 6.5 as mobile phase. However, the protein displayed strong non-specific interactions with the matrix of the column and SapL1 could not be eluted even using 5 M of NaCl. Interestingly, it could be recovered when the buffer was supplemented with 20 mM α -methyl fucoside or L-rhamnose, evidencing similar effects on SapL1 elution for both sugars. Due to the impossibility to estimate the molecular weight of SapL1 by size exclusion chromatography on this resin, we performed measurements in solution using Dynamic Light Scattering (DLS). We obtained a monodisperse peak corresponding to a protein of 72 ± 29.4 kDa corroborating that SapL1 forms dimers as FleA and other proteins of this family (monomer MW: 40 kDa, data not shown)²⁷. The range of the standard deviation also suggested an ellipsoidal shape, which is characteristic for the dimers in this lectin family^{27,35}.

Carbohydrate binding properties. A hemagglutination assay showed that recombinant SapL1 agglutinated rabbit red blood cells at $0.97 \mu\text{g}\cdot\text{mL}^{-1}$ (Fig. 3A). It confirms that the recombinant lectin is active and that its heterologous production in *E. coli* did not alter its hemagglutinating properties.

To identify the potential ligands of SapL1 on epithelial cell surfaces, we submitted SapL1 to the glycan array version 5.4 of the Consortium for Functional Glycomics (USA) consisting of 585 mammalian glycans. It was labelled with Fluorescein Isothiocyanate (FITC) in a molar ratio of 0.426 and its binding properties were analysed at two different concentrations (5 and $50 \mu\text{g}\cdot\text{mL}^{-1}$). As expected from its homology with FleA, SapL1 recognizes fucosylated oligosaccharides independently of the fucose linkage. The α 1,2 and α 1,3/4 linked fucosides displayed the highest affinity whilst the lowest was seen with the α 1,6 linked ones. The weakest interactions with fucosylated compounds were reported for branched oligosaccharides (Figs. 3C and S4).

The affinity of SapL1 for L-fucose and α -methyl-fucoside was determined by Isothermal Titration Calorimetry (ITC) and the K_d was found to be $225 \pm 1.52 \mu\text{M}$ and $190 \pm 1.44 \mu\text{M}$, respectively with stoichiometry fixed to 1 since the measurements were done in the presence of an excess of ligand (Fig. 3B). These values are in agreement

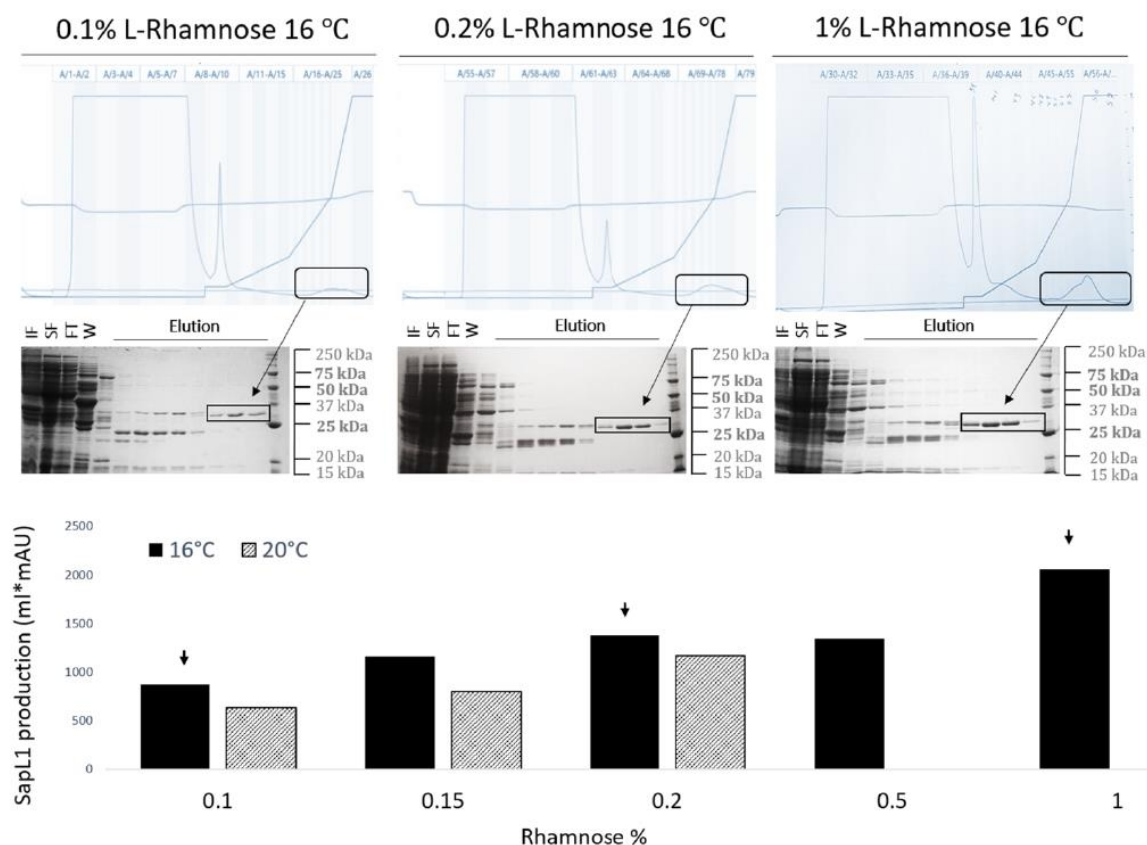


Figure 2. Rhamnose influence on SapL1 solubility. (A) SapL1 purification chromatograms at 0.1%, 0.2%, and 1% rhamnose with their corresponding SDS-PAGE profile. IF: Insoluble fraction, SF: Soluble fraction, FT: Flow through, W: wash and elution, from left to right respectively. Black rectangles indicate the elution peaks and elution fractions containing SapL1 on SDS-polyacrylamide gels. (B) Numerical values derived from the integration of areas under the peaks corresponding to SapL1 for each experiment. Black and striped bars represent the results for experiments performed at 16 °C and 20 °C, respectively.

with the affinity constant (around 110 μM) reported for FleA for α -methyl-fucoside³⁶. No binding interaction was observed for SapL1 with rhamnose by ITC (data not shown).

Analysis of the glycans constituting blood group determinants revealed that SapL1 binds to all epitopes with a preference for H type 2 blood group (Fuc α 1-2Gal β 1-4GlcNAc β) then Lewis a (Gal β 1-3(Fuca1-4)GlcNAc β) and Lewis X (Gal β 1-4(Fuca1-3)GlcNAc β). However, most of the recognized branched oligosaccharides contained the core fucose Fuca1-6. Epitopes with two fucose units, such as Lewis b and fucosylated polylactosamine, were also well recognized. Addition of a galactose or a GalNAc as in blood group B or A antigens did not impair Fuca1-2 recognition (Fig. 3C). Spacers used to join the carbohydrates to the chip also display a strong influence on binding. It is remarkable that 70% of the 90 positive binders contained either the spacer Sp0 (CH₂CH₂NH₂) or Sp8 (CH₂CH₂CH₂NH₂). Those spacers also display a strong influence on binding, especially for small glycans such as Gala1-3(Fuca1-2)Gal β 1-4(Fuca1-3)GlcNAc β which was recognized when attached to Sp0 but not to Sp8. This may be due to a steric hindrance caused by the modification of carbohydrate presentation on the surface of the chip.

Overall structure of SapL1. SapL1 was co-crystallized with α -methyl-fucoside and the structure of the complex was solved by molecular replacement at 2.3 Å resolution in the P₂₁ space group using the coordinates of FleA (PDB code 4D4U³⁴) as the search model. The asymmetric unit contained two monomers, assembled as a dimer with all 295 amino acids visible apart of the N-terminal methionine. See data collection and refinement statistics in Table 1. SapL1 folds into the canonical six-bladed β -propeller with six-binding sites at the interface between blades typical for this family of lectin (Fig. 4). A fucose moiety was found in 5 and 4 of the six binding pockets of chains A and B, respectively while glycerol originating from the crystallizing solution was found in the other binding sites.

The overall fold of SapL1 and FleA as well as the overall dimer are very similar with a rmsd of 1.2 and 1.26 Å, respectively. They share 43% of sequence identity and both proteins present almost the same distribution of β -strands except for an extra strand in blade 2 (β 10), the lack of the last strand of blade 4 in SapL1, and the

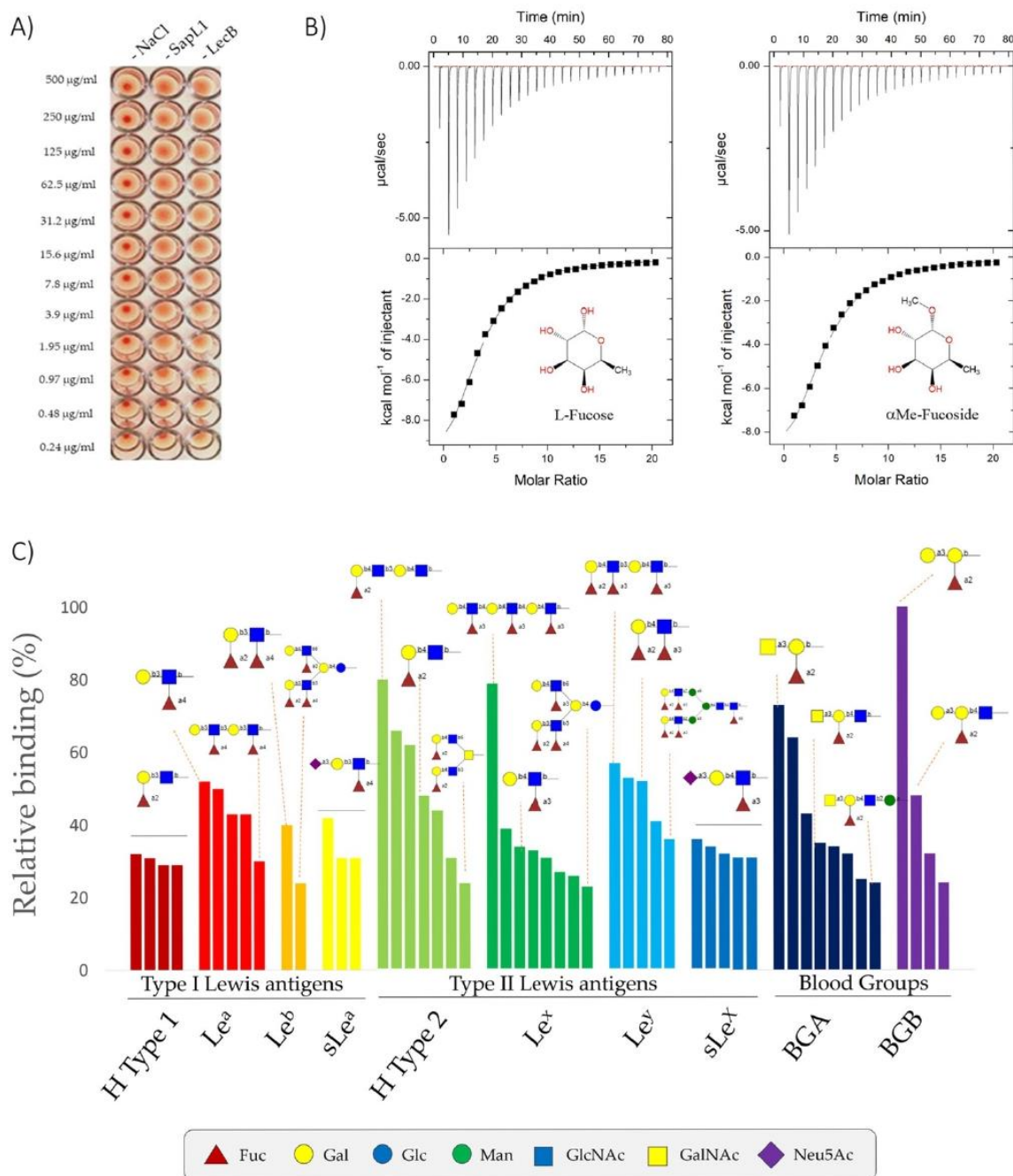


Figure 3. SapL1 carbohydrate binding properties. (A) Hemagglutination assay of SapL1 on fresh rabbit erythrocytes. Negative and positive controls consist of 150 mM NaCl and the lectin LecB from *Pseudomonas aeruginosa*, respectively. (B) Titration of SapL1 with L-fucose and α -methyl-fucoside with the thermogram and the integration displayed at the top and bottom, respectively. (C) Analysis of the interactions of SapL1 to glycans containing Lewis and ABH blood group epitopes. The graph shows the relative binding of SapL1 to glycans containing Lewis and ABH blood group antigens from the 90 hits identified as binders.

Data collection		
Beamline	SOLEIL Proxima 1	
Wavelength (Å)	0.97857	
Space group	P2 ₁	
Unit cell dimensions a, b, c (Å), α, β, γ (°)	76.06, 45.66, 83.48, 90, 105.05, 90	
No. of monomers in ASU	2	
Resolution (Å)	40.0–2.4 (2.46–2.4)	
<i>R</i> _{merge}	0.085 (0.428)	
<i>R</i> _{pim}	0.07 (0.356)	
Mean <i>I</i> /σ (<i>I</i>)	5.3 (1.6)	
Completeness (%)	98.2 (98.4)	
Multiplicity	2.7 (2.7)	
CC1/2	0.991 (0.7)	
No. reflections /No. Unique reflections	58305/ 21570	
Refinement		
Resolution (Å)	40.0–2.4	
No. of reflections in working set / Free set	21561 / 1089	
<i>R</i> work/ <i>R</i> free	17.5 / 24.0	
R.m.s Bond lengths (Å)	0.017	
Rmsd Bond angles (°)	1.94	
Rmsd Chiral (Å ³)	0.087	
No. atoms / Bfac (Å ²)	Chain A	Chain B
Protein	2222/32.6	2220/40.8
Ligand and heterogen	89/31.6	66/36.9
Waters	111/31.0	69/33.2
Ramachandran Allowed / Favored/ Outliers (%)	96/4/ 0	95/4/1
PDB Code	6TRV	

Table 1. Data-collection and refinement statistics *Values in parentheses are for the outer shell.

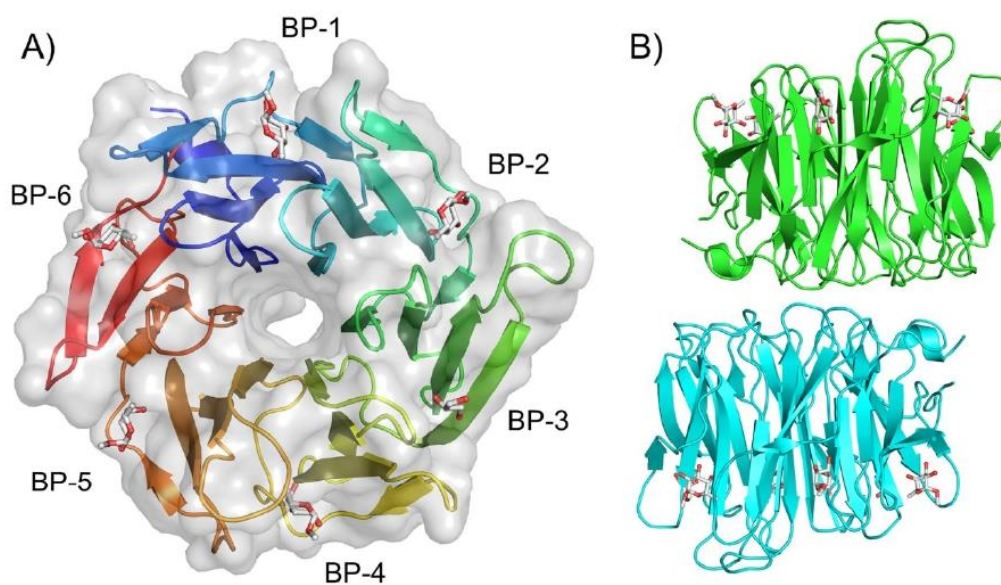


Figure 4. SapL1 overall structure. (A) Surface and cartoon representation of SapL1 monomer colored from blue (N-terminal end) to red (C-terminal end). BP: Binding pockets. (B) Representation of SapL1 dimer colored by monomer with ligand depicted in sticks.

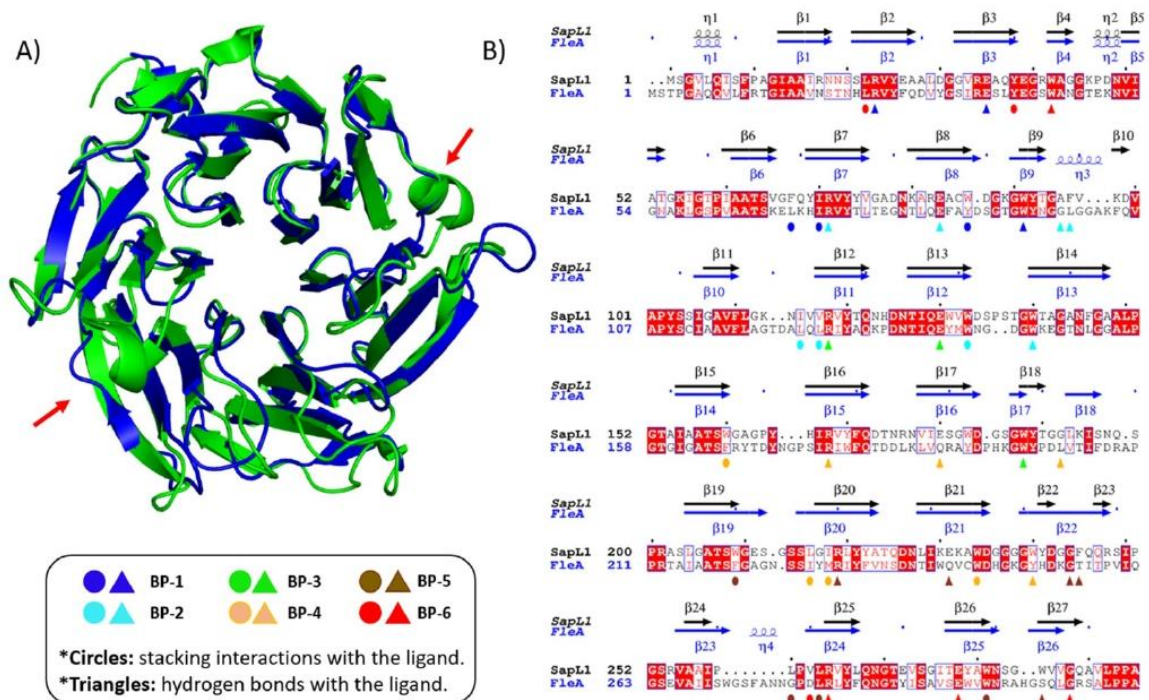


Figure 5. SapL1/FleA comparison. (A) Overlay of SapL1 (blue, PDB: 6TRV) and FleA (green, PDB: 4D4U) structures. Red arrows indicate the main differences between structures. (B) Sequence alignment of SapL1 and FleA with display of their secondary structure elements. The residues involved in ligand interactions of each pocket are indicated by circles for hydrophobic and stacking interactions and by triangles for hydrogen bonds colored according to the binding pocket: BP-1, blue; BP-2 cyan; BP-3 green; BP-4, orange; BP-5, brown; BP-6, red.

external face of blade 5, in which FleA displays an elongated β -strand (β 22) that is split in two in SapL1 (β 22/23). It is remarkable that the structure of the first 3 blades is highly conserved in both proteins, while the second half (blades 3–6) displays the largest discrepancies. Furthermore, FleA also presents two additional small α -helices (η 3 and η 4), located in the loops between sheets that serve as connectors for blades (2/3 and 5/6, respectively, Fig. 5). Similar conclusions can be drawn when comparing SapL1 to its homologue AOL in *Aspergillus oryzae*³⁷.

Protein–ligand interactions. Due to divergence in the tandem repeat sequence forming each blade of the propeller, the six binding sites of SapL1 monomer are not equivalent, but they share important conserved features. Hydrophobic interactions are observed between the C6 of the fucose and at least three residues of the protein (mainly isoleucine, tryptophan/tyrosine and leucine). The O2 and O3 hydroxyls make strong hydrogen bonds with a conserved triad of amino acids consisting of an arginine, a glutamic acid and a tryptophan. In the cases where glycerol was found in the binding pocket, its own hydroxyls mimic the interactions of fucose with these same residues (Fig. 6). It is to be noted that SapL1 binding sites are more conserved than FleA binding sites where a glutamine can replace the glutamic acid and a tyrosine the tryptophan in the triad (Fig. S5).

The O2 hydroxyl seemed to be the most versatile position, since it established interactions with an adjacent loop only present in four of the six binding sites (BPs 2, 4, 5 and 6). This is particularly interesting since the binding pockets that do not contain this loop (BPs 1 and 3) were mostly occupied by glycerol instead of fucose, indicating that those interactions could be responsible for enhancing the affinity and should be explored for development of inhibitors.

Binding to epithelial cells. In order to investigate the role of SapL1 in host–pathogen interactions and especially in adhesion, fluorescence microscopy was used. FITC-labelled SapL1 was incubated at two different concentrations with BEAS-2B human bronchial epithelial cells. The microscopy images, obtained for two separate experiments, clearly show binding of the lectin to those cells (Fig. 7). Fluorescence was observed all around the cell surface but also concentrated in some part of the nucleus. A strong fluorescence signal was already observed at the lowest concentration used ($5 \mu\text{g mL}^{-1}$). The binding of SapL1 was inhibited in the presence of its cognate ligand: α -methyl-fucoside (Fig. 7E–G). Some background signal was still observed in the first experiment (Fig. 7E, G) whilst it was totally abolished in the duplicate (Fig. 7E, H). These data show that SapL1 binding to the cells is dependent on the recognition of fucosylated carbohydrates.

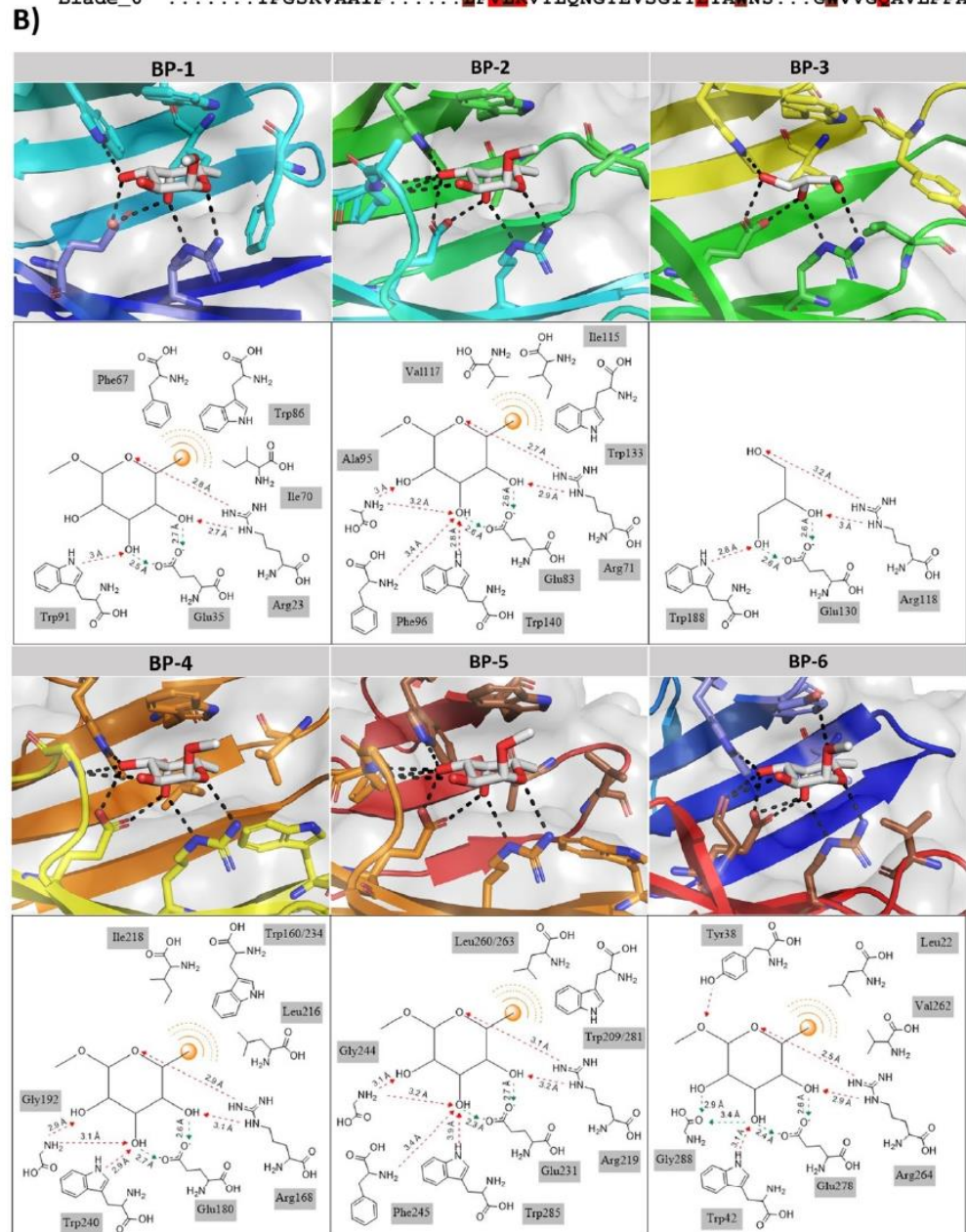
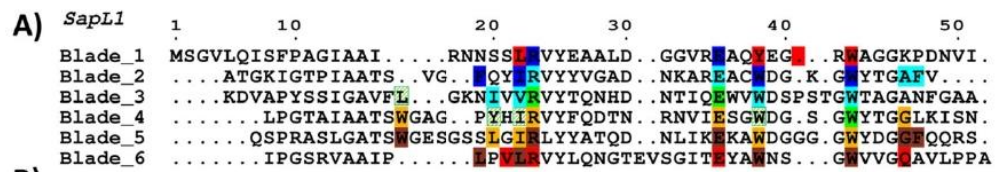


Figure 6. SapL1-ligand interactions. (A) The sequence alignment of the 6 blades of SapL1. The residues involved in ligand binding interactions are indicated in solid boxes: BP-1, blue; BP-2 cyan; BP-3 green; BP-4, orange; BP-5, brown; BP-6, red. Striped green boxes indicate the four additional residues expected to be involved in fucose binding within BP-3 that were not visible in the structure since only glycerol was attached to this pocket. (B) Zoom on the interactions of ligand with each binding site of SapL1 and their schematic representation.

here by the addition of rhamnose to the culture medium^{41–49}. This hypothesis was later supported by the finding that the addition of glycerol, instead of rhamnose, displays the same effect on production and solubility of SapL1 (data not shown). Plasmolysis may also play an essential role in SapL1 stabilization, since it has been shown that macromolecular crowding has positive effects on protein folding and in some cases it can drive self-association of improperly folded proteins into functional oligomers⁴⁵. These findings, and the extensive research carried out to produce SapL1 in its soluble form, provide important insights for the heterologous expression of eukaryotic lectins tending to be produced as insoluble proteins in *E. coli*. Besides, our findings highlight the useful role of organic osmoregulators during heterologous expression of proteins to favour their proper folding and to avoid the formation of inclusion bodies.

We demonstrated that SapL1 is strictly specific for fucosylated carbohydrates and recognized all blood group types present in the glycan array screening. These results are particularly interesting since it has been shown that there is six times more fucose $\alpha 1,3/4$ linked in glycoproteins in the CF airways³¹. This phenomenon is mainly due to an increased expression of the $\alpha 1,3$ -fucosyltransferase, which is involved in the synthesis of sialyl-Lewis X and Lewis X determinants attached to bronchial mucins³². Therefore, the fact that SapL1 recognizes equally the $\alpha 1,2$ and $\alpha 1,3/4$ -linked fucosides may explain the high incidence of scedosporiosis in CF patients. Besides, it correlates with the presence of *SapL1* gene in all the pathogenic strains of *Scedosporium* whose genome has been sequenced to date^{33,50,51}. However, there is no evidence suggesting that blood group phenotypes may have an influence in this recognition conversely to previous reports on other pathogens like *Pseudomonas aeruginosa* and *Haemophilus influenzae*^{52–54}.

SapL1 adopts a the 6-bladed β -propeller fold and forms dimer leading to two fucose binding surfaces. It belongs to ProLec6A lectin family whose biological role is still unknown, in particular for the mushroom members³⁵. The first member structurally characterized in this lectin family was AAL from the orange peel mushroom *Aleuria aurantia*⁵⁵. It presents differences on both sides of the β -propeller with the microfungus members of this family: SapL1, AOL and FleA mainly at the level of the surface loops. On one side, this leads to a different dimerization interface and the dimer cannot overlay. On the other side, where sugar binding occurs, the differences in size and sequence of the surface loops lead to change in the architecture of the related binding sites and affect fucose binding since AAL has only five and not six functional binding sites^{27,37,55}. The specificity and affinity of each one of the six binding pockets of SapL1 have been deeply analyzed to highlight the features that must be explored for the design of efficient inhibitors. Within our analysis, we found that the binding pockets are non-equivalent but they all share the features necessary for fucose recognition. This example of divergence can also be found in the other lectins of this family with FleA exhibiting the greatest differences between pockets known to date³⁴. It is remarkable that although SapL1 has a relatively low sequence identity with other members of this family, its specificity and affinity are very similar to those reported for FleA from *A. fumigatus*³⁶, AAL from *A. aurantia*⁵⁵ and the bacterial lectins BambL and RSL from *Burkholderia ambifaria* and *Ralstonia solanacearum*, respectively^{56,57}.

Finally, fluorescence microscopy experiments allowed us to get preliminary results on SapL1 function. The lectin is able to recognize and to bind to bronchial epithelial cells, and we demonstrated that this binding is dependent on the recognition of fucosylated glycoconjugates at the surface of the target cells as it is abolished in the presence of fucose. These results corroborate a role of SapL1 in mediating the recognition and adhesion of *S. apiospermum* to host cells as found for other lectins from pathogenic microorganisms and in particular for its homolog FleA^{22,27}. Localisation of the lectin in the microfungus would also help to better understand its role but no SapL1 antibodies are yet available. Together with the crystallographic data, our study shows that the structure and function of lectins belonging to the 6-bladed β -propeller fucose specific family recently named PropLec6A are highly conserved, settling the possibility for development of a broad-spectrum therapy⁵⁶.

Overall, our research has revealed the first insights about the recognition of fucosylated human glycoconjugates by *S. apiospermum* lectin SapL1 and contributes to the general understanding of the host-binding process during the early stages of infection. The lectin is found in all *Scedosporium* strains sequenced. The detailed information exposed here places SapL1 as a promising target for treatment of *Scedosporium* infections although it is clear that more functional data are still required. It will be of great value to guide the development of anti-adhesive glycodrugs against this pathogen.

Methods

Production. The coding sequence for SapL1 (75–369) was optimized for expression in *Escherichia coli* and ordered at Eurofins Genomics (Ebersberg, Germany) before cloning into the expression vectors pET-TEV⁵⁸, pET32-TEV³⁶ and pProNde. pProNde is a homemade vector where the NcoI restriction site of pProEx HTb (EMBL, Heidelberg) has been replaced by a NdeI restriction site by PCR. Then, plasmids were introduced into *E. coli* strains by thermal shock at 42 °C and different expression parameters were assayed (Table S1) until soluble expression was achieved. Finally, SapL1 was expressed using pProNde vector in *E. coli* BL21trxB (DE3) strain with the cells grown at 37 °C and 160 rpm until an OD₆₀₀ of 0.4. The temperature was then lowered to 16 °C and when OD_{600nm} reached 0.8, the induction was carried out overnight by the addition 0.05 mM of IPTG and 1% L-rhamnose.

Purification. Cells were subsequently harvested by centrifugation at 5000 g for 10 min and resuspended in buffer A (50 mM Tris-HCl, 500 mM NaCl, pH 8.5). After addition of 1 μ L of Denarase (C-LEcta GmbH, Leipzig, Germany) and moderate agitation during 30 min at room temperature, cells were disrupted at 1.9 kbar using a cell disruptor (Constant Systems Ltd, UK). Cell debris were removed by centrifugation at 22,000 g for 30 min. The supernatant was filtered using 0.45 μ m membranes (PES, ClearLine) and the protein purification was carried out by IMAC using 1 mL His-Trap FF columns (Cytiva) and a NGC chromatography system (Bio-Rad).

After loading the supernatant, the column was rinsed thoroughly with buffer A until stabilization of the baseline. Bound proteins were eluted through the addition of buffer B (50 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, pH 8.5) in an 0–500 mM imidazole gradient over 20 mL. Fractions containing SapL1 were pooled and concentrated by ultrafiltration (Pall, 10 kDa cut-off) prior buffer exchange to 50 mM Tris-HCl, 100 mM NaCl, pH 8.5 using PD10 desalting columns (Cytiva). Then, the fusion was cleaved off overnight at 19 °C using the TEV protease produced in the lab in 1:50 ratio and addition of 0.5 mM EDTA and 0.25 mM TCEP. The sample was loaded on the His-trap column to separate the cleaved protein collected in the flow-through from the TEV protease and potential uncleaved sample retained and eluted with imidazole. SapL1 containing fractions were concentrated by centrifugation to the desired concentration.

Size exclusion chromatography. Size exclusion chromatography (SEC) was performed using a High-Resolution ENrich SEC 70 column (Bio-Rad) on NGC chromatography system (Bio-Rad). Column was equilibrated with 50 mL of buffer D (20 mM MES, 100 mM NaCl, pH 6.5) and 200 μ L of sample at 10 mg mL⁻¹ were injected into the system followed by 40 mL isocratic elution on buffer D supplemented with 20 mM α -methyl-L-fucoside (TCI Europe), 0.1% L-rhamnose, 5 M NaCl, 2 M NaCl or 500 mM NaCl, according to the experiment. Fractions were monitored by the absorbance at 280 nm and 0.5 mL fractions were collected in the resolving region of the column.

Dynamic light scattering (DLS). DLS analyses were performed using a Zetasizer Nano ZS (Malvern Panalytical) with a 40 μ L quartz cuvette. Measurements were performed in triplicate on protein sample at 1 mg mL⁻¹ in buffer C (50 mM Tris-HCl, 100 mM NaCl, pH 8.5) after centrifugation.

Thermal shift assay (TSA). The thermal stability of SapL1 was analyzed by TSA with the MiniOpticon real-time PCR system (Bio-Rad). Prior assay, buffer stocks at 100 mM and a mixture containing 70 μ L of SapL1 at 1 mg mL⁻¹, 7 μ L of 500 \times Sypro Orange (Merk Sigma-Aldrich), and 63 μ L of ultrapure H₂O were prepared. Then, 7.5 μ L of H₂O, 12.5 μ L of the corresponding buffer and 5 μ L of the protein/Sypro mixture were mixed in 96-well PCR microplates. The heat exchange test was then carried out from 20 to 100 °C with a heating rate of 1 °C min⁻¹. Fluorescence intensity was measured with Ex/Em: 490/530 nm and the data processing was performed with the CFX Manager software.

Isothermal titration calorimetry (ITC). Experiments were performed using a Microcal ITC200 calorimeter (Malvern Panalytical) with 40 μ L of L-fucose 5 mM in the syringe and 200 μ L of protein 0.05 mM in the sample cell. Both, the protein and sugars were dissolved in a buffer composed of 20 mM Tris HCl pH 8.0 and 100 mM NaCl. A total of 26 injections of 1.5 μ L of ligand were added to the sample cell at intervals of 180 s while stirring at 850 rpm. Experimental data were adjusted to a theoretical titration curve by the Origin ITC Analysis software. All experiments were performed at least by duplicated and the stoichiometry was fixed to 1.

Hemagglutination assay. Agglutination test was performed with fresh rabbit erythrocytes (bioMérieux, Lyon) in U 96-well plates (Nalgene). For the test, 150 mM NaCl was used as a negative control and 1 mg mL⁻¹ LecB of *P. aeruginosa* as a positive control. 50 μ L of sample was prepared at 0.1 mg mL⁻¹ and submitted to serial double dilutions. 50 μ L of rabbit erythrocytes 3% were added to each well prior incubation of the plate at room temperature. After 2 h, the result of the experiment was evaluated and agglutination activity was calculated according to the dilution of the protein.

Glycan arrays. Protein was labelled with Fluorescein Isothiocyanate (FITC, Merk, Sigma-Aldrich) according to the supplier's instructions with slight modifications. Briefly two milligrams of protein were dissolved in 1 mL of buffer E (100 mM Na₂CO₃, 100 mM NaCl, pH 9); then, 40 μ L of FITC at 1 mg mL⁻¹, previously dissolved in DMSO, were gradually added to the protein solution and the mixture was gently stirred at room temperature overnight. Next day, the solution was supplemented with NH₄Cl to a final concentration of 50 mM and free FITC was removed using PD10 column with PBS as mobile phase. Protein concentration was determined at A₂₈₀ and FITC at A₄₉₀ using a NanoDrop 200 (Thermo Scientific) and Fluorescein/Protein molar ratio (F/P) was estimated by the following formula:

$$\text{Molar } \frac{F}{P} = \frac{MW}{389} \times \frac{\frac{A_{495}}{195}}{[A_{280} - (0.35 \times A_{495})]E^{0.1\%}}$$

where MW is the molecular weight of the protein, 389 is the molecular weight of FITC, 195 is the absorption E 0.1% of bound FITC at 490 nm at pH 13.0, (0.35 X A495) is the correction factor due to the absorbance of FITC at 280 nm, and E 0.1% is the absorption at 280 nm of a protein at 1.0 mg mL⁻¹. Being an ideal F/P should be 0.3 > 1.

Labelled lectin was sent to the Consortium for Functional Glycomics (CFG; Boston, MA, USA) and its binding properties were assayed at 5 and 50 μ g mL⁻¹ on a "Mammalian Glycan Array version 5.4" which contain 585 glycans in replicates of 6. The highest and lowest signal of each set of replicates were eliminated and the average of the remaining data was normalized to the percentages of the highest RFU value for each analysis. Finally, the percentages for each glycan were averaged at different lectin concentrations.

Crystallization and data collection. Crystal screening was performed using the hanging-drop vapour diffusion technique by mixing equal volumes of pure protein at 5 mg mL⁻¹ and precipitant solutions from com-

mercial screenings of Molecular Dimensions (Newmarket, UK). 2 µL drops were incubated at 19°C until crystals appeared. A subsequent optimization of positive conditions for SapL1 crystallization was carried out and crystals suitable for X-ray diffraction analysis were obtained under solution containing 100 mM Bicine pH 8.5, 1.5 M ammonium sulfate (NH₄)₂SO₄ and 12% v/v glycerol. Crystals were soaked in mother liquor supplemented with 10% (v/v) glycerol, prior to flash cooling in liquid nitrogen. Data collection was performed on PX1 beamline at SOLEIL Synchrotron (Saint Aubin, France) using an Eiger2 X 9M pixel detector (Dectris Ltd, Switzerland).

Structure determination. Data were processed using XDS⁵⁹ software and were converted to structure factors using the CCP4 program package v.6.1⁶⁰, with 5% of the data reserved for Rfree calculation. The structure was determined using the molecular-replacement method with Phaser v.2.5⁶¹, using the structure of FleA dimer (PDB entry 4D4U³⁴) as starting model. Model refinement was performed using REFMAC 5.8⁶² alternated with manual model building in Coot v.0.7⁶³. Sugar residues and other compounds that were present were placed manually using Coot and validated using Privateer⁶⁴. The final model has been validated and deposited in the PDB Database with accession number 6TRV.

Cell culture and fluorescence microscopy. Human bronchial epithelial cells (BEAS-2B cell line) were maintained and serially passaged in F-12 culture medium supplemented with 10% fetal calf serum (FCS), 1% penicillin and streptomycin, and 10 mM HEPES in 75-cm² culture flasks. For microscopy experiments, BEAS-2B cells were grown to confluency on coverslips (precision cover glasses thickness No. 1.5H; Marienfeld, Lauda-Königshofen, Germany). To test interaction of SapL1 lectin with epithelial cells, BEAS-2B cells were incubated with different concentrations of FITC-lectin (5 or 10 µg mL⁻¹ in F-12) for 1 h at 37 °C. To validate the specificity of this interaction, SapL1-FITC was co-incubated in presence of 2 mM methyl-α-L-fucopyranoside for 30 min at 37 °C and then added to the cells for 1 h. Supernatant was then removed, cells washed 3 times with F-12, once with PBS and then fixed with 4% paraformaldehyde for 15 min. After 3 washes with PBS, nuclei were stained with 4'-6-diamidino-2-phenylindole dihydrochloride (DAPI, 1:1000 in PBS) for 5 min. Coverslip were mounted with Prolong Glass Antifade Mountant (Invitrogen) on Superfrost glass slides (Thermo Fisher Scientific). Images were acquired with an upright Olympus BX43 microscope.

Figures were created using PyMOL Version 1.8.4 (Schrödinger), ChemDraw Version 15, ESPript Version 3.0⁶⁵ and PowerPoint 16.

Accession numbers

The structure of SapL1 was deposited as PDB ID: 6TRV.

Received: 17 December 2020; Accepted: 16 July 2021

Published online: 09 August 2021

References

1. Tuite, N. L. & Lacey, K. Vol. 968 (eds Louise O'Connor & Barry Glynn) 1–23 (Humana Press, Totowa, NJ, 2013).
2. Thornton, C. R. Detection of the “Big Five” mold killers of humans: *Aspergillus*, *Fusarium*, *Lomentospora*, *Scedosporium* and *Mucormycetes*. *Adv. Appl. Microbiol.* **110**, 1–61. <https://doi.org/10.1016/bs.aambs.2019.10.003> (2020).
3. Gilgado, F., Cano, J., Gené, J. & Guarro, J. Molecular phylogeny of the *Pseudallescheria boydii* species complex: proposal of two new species. *J. Clin. Microbiol.* **43**, 4930–4942. <https://doi.org/10.1128/JCM.43.10.4930-4942.2005> (2005).
4. Ramirez-Garcia, A. et al. *Scedosporium* and *Lomentospora*: an updated overview of underrated opportunists. *Med. Mycol.* **56**, S102–S125. <https://doi.org/10.1093/mmy/myx113> (2018).
5. Bouchara, J.-P. et al. Advances in understanding and managing *Scedosporium* respiratory infections in patients with cystic fibrosis. *Expert Rev. Respir. Med.* **14**, 259–273. <https://doi.org/10.1080/17476348.2020.1705787> (2020).
6. Rammaert, B. et al. Perspectives on *Scedosporium* species and *Lomentospora prolificans* in lung transplantation: results of an international practice survey from ESCMID fungal infection study group and study group for infections in compromised hosts, and European Confederation of Medical Mycology. *Transpl. Infect. Dis.* **21**, 1–8. <https://doi.org/10.1111/tid.13141> (2019).
7. Luplertlop, N. *Pseudallescheria/Scedosporium* complex species: from saprobic to pathogenic fungus. *J. Mycol. Med.* **28**, 249–256. <https://doi.org/10.1016/j.mycmed.2018.02.015> (2018).
8. Tortorano, A. M. et al. ESCMID and EECMM joint guidelines on diagnosis and management of hyalohyphomycosis: *Fusarium* spp., *Scedosporium* spp. and others. *Clin. Microbiol. Infect.* **20**, 27–46. <https://doi.org/10.1111/1469-0691.12465> (2014).
9. Guarro, J. et al. *Scedosporium apiospermum*: changing clinical spectrum of a therapy-refractory opportunist. *Med. Mycol.* **44**, 295–327. <https://doi.org/10.1080/13693780600752507> (2006).
10. Troke, P. et al. Treatment of scedosporiosis with voriconazole: clinical experience with 107 patients. *Antimicrob. Agents Chemother.* **52**, 1743–1750. <https://doi.org/10.1128/AAC.01388-07> (2008).
11. Cortez, K. J. et al. Infections caused by *Scedosporium* spp. *Clin. Microbiol. Rev.* **21**, 157–197. <https://doi.org/10.1128/CMR.00039-07> (2008).
12. Theuretzbacher, U. & Piddock, L. J. V. Non-traditional antibacterial therapeutic options and challenges. *Cell Host Microbe* **26**, 61–72. <https://doi.org/10.1016/j.chom.2019.06.004> (2019).
13. Krachler, A. M. & Orth, K. Targeting the bacteria-host interface strategies in anti-adhesion therapy. *Virulence* **4**, 284–294. <https://doi.org/10.4161/viru.24606> (2013).
14. Ofek, I., Hasty, D. L. & Sharon, N. Anti-adhesion therapy of bacterial diseases: prospects and problems. *FEMS Immunol. Med. Microbiol.* **38**, 181–191. [https://doi.org/10.1016/S0928-8244\(03\)00228-1](https://doi.org/10.1016/S0928-8244(03)00228-1) (2003).
15. Figueiredo, R. T. et al. TLR4 recognizes *Pseudallescheria boydii* conidia and purified rhamnomannans. *J. Biol. Chem.* **285**, 40714–40723. <https://doi.org/10.1074/jbc.M110.181255> (2010).
16. Pinto, M. R., Gorin, P. A. J., Wait, R., Mulloy, B. & Barreto-Bergter, E. Structures of the O-linked oligosaccharides of a complex glycoconjugate from *Pseudallescheria boydii*. *Glycobiology* **15**, 895–904. <https://doi.org/10.1093/glycob/cwi084> (2005).
17. Pinto, M. R. et al. Involvement of peptidorhamnomannan in the interaction of *Pseudallescheria boydii* and HEp2 cells. *Microbes Infect.* **6**, 1259–1267. <https://doi.org/10.1016/j.micinf.2004.07.006> (2004).
18. Bittencourt, V. C. B. et al. An α-glucan of *Pseudallescheria boydii* is involved in fungal phagocytosis and Toll-like receptor activation. *J. Biol. Chem.* **281**, 22614–22623. <https://doi.org/10.1074/jbc.M511417200> (2006).

19. Ghamrawi, S. *et al.* Cell wall modifications during conidial maturation of the human pathogenic fungus *Pseudallescheria boydii*. *PLoS ONE* **9**, e100290. <https://doi.org/10.1371/journal.pone.0100290> (2014).
20. Rollin-Pinheiro, R., Liporagi-lobes, L. C., Meirelles, J. V. D. & Souza, L. M. D. Characterization of *Scedosporium apiospermum* glucosylceramides and their involvement in fungal development and macrophage functions. *PLoS ONE* **9**, e98149. <https://doi.org/10.1371/journal.pone.0098149> (2014).
21. De Mello, T. P. *et al.* *Scedosporium apiospermum*, *Scedosporium aurantiacum*, *Scedosporium minutisporium* and *Lomentospora prolificans*: a comparative study of surface molecules produced by conidial and germinated conidial cells. *Mem. Inst. Oswaldo Cruz* **113**, 1–8. <https://doi.org/10.1590/0074-02760180102> (2018).
22. Imberty, A. & Varrot, A. Microbial recognition of human cell surface glycoconjugates. *Curr. Opin. Struct. Biol.* **18**, 567–576. <https://doi.org/10.1016/j.sbi.2008.08.001> (2008).
23. Sharon, N. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochim. Biophys. Acta* **1760**, 527–537. <https://doi.org/10.1016/j.bbagen.2005.12.008> (2006).
24. Tamburrini, A., Colombo, C. & Bernardi, A. Design and synthesis of glycomimetics: recent advances. *Med. Res. Rev.* **40**, 495–531. <https://doi.org/10.1002/med.21625> (2020).
25. Sattin, S. & Bernardi, A. Glycoconjugates and glycomimetics as microbial anti-adhesives. *Trends Biotechnol.* **34**, 483–495. <https://doi.org/10.1016/j.tibtech.2016.01.004> (2016).
26. Pieters, R. J. Carbohydrate mediated bacterial adhesion. *Adv. Exp. Med. Biol.* **715**, 227–240. https://doi.org/10.1007/978-94-007-0940-9_14 (2011).
27. Houser, J. *et al.* A soluble fucose-specific lectin from *Aspergillus fumigatus* conidia: structure, specificity and possible role in fungal pathogenicity. *PLoS ONE* **8**, e83077. <https://doi.org/10.1371/journal.pone.0083077> (2013).
28. Kerr, S. C. *et al.* FleA expression in *Aspergillus fumigatus* is recognized by fucosylated structures on mucins and macrophages to prevent lung infection. *PLOS Pathog.* **12**, e100555. <https://doi.org/10.1371/journal.ppat.1005555> (2016).
29. Sakai, K., Hiemori, K., Tateno, H., Hirabayashi, J. & Gono, T. Fucose-specific lectin of *Aspergillus fumigatus*: binding properties and effects on immune response stimulation. *Med. Mycol.* **57**, 71–83. <https://doi.org/10.1093/mmy/myx163> (2019).
30. Thornton, D. J., Rousseau, K. & McGuckin, M. A. Structure and function of the polymeric mucins in airways mucus. *Annu. Rev. Physiol.* **70**, 459–486. <https://doi.org/10.1146/annurev.physiol.70.113006.100702> (2008).
31. Glick, M. C., Kothari, V. A., Liu, A., Stoykova, L. I. & Scanlin, T. F. Activity of fucosyltransferases and altered glycosylation in cystic fibrosis airway epithelial cells. *Biochimie* **83**, 743–747. [https://doi.org/10.1016/s0300-9084\(01\)01323-2](https://doi.org/10.1016/s0300-9084(01)01323-2) (2001).
32. Lamblin, G. *et al.* Human airway mucin glycosylation: a combinatorial of carbohydrate determinants which vary in cystic fibrosis. *Glycoconj. J.* **18**, 661–684. <https://doi.org/10.1023/A:1020867221861> (2001).
33. Vandeputte, P. *et al.* Draft genome sequence of the pathogenic fungus *Scedosporium apiospermum*. *Genome Announc.* **2**, e00988–e914. <https://doi.org/10.1128/genomea.00988-14> (2014).
34. Houser, J. *et al.* Structural insights into *Aspergillus fumigatus* lectin specificity: AFL binding sites are functionally non-equivalent. *Acta Crystallogr. D Biol. Crystallogr.* <https://doi.org/10.1107/S1399004714026595> (2015).
35. Bonnardel, F. *et al.* Architecture and evolution of blade assembly in β -propeller lectins. *Structure* **27**, 764–775. <https://doi.org/10.1016/j.str.2019.02.002> (2019).
36. Lehot, V. *et al.* Multivalent fucosides with nanomolar affinity for the *Aspergillus fumigatus* lectin FleA prevent spore adhesion to pneumocytes. *Chem. Eur. J.* **24**, 19243–19249. <https://doi.org/10.1002/chem.201803602> (2018).
37. Makyio, H. *et al.* Six independent fucose-binding sites in the crystal structure of *Aspergillus oryzae* lectin. *Biochem. Biophys. Res. Commun.* **477**, 477–482. <https://doi.org/10.1016/j.bbrc.2016.06.069> (2016).
38. Hogan, L. H., Klein, B. S. & Levitz, S. M. Virulence factors of medically important fungi. *Clin. Microbiol. Rev.* **9**, 469–488. <https://doi.org/10.1128/CMR.9.4.469-488.1996> (1996).
39. Debourgogne, A., Dorin, J. & Machouart, M. Emerging infections due to filamentous fungi in humans and animals: only the tip of the iceberg?. *Environ. Microbiol. Rep.* **8**, 332–342. <https://doi.org/10.1111/1758-2229.12404> (2016).
40. Vandeputte, P. *et al.* Comparative transcriptome analysis unveils the adaptive mechanisms of *Scedosporium apiospermum* to the microenvironment encountered in the lungs of patients with cystic fibrosis. *Comput. Struct. Biotechnol. J.* **18**, 3468–3483. <https://doi.org/10.1016/j.csbj.2020.10.034> (2020).
41. Frank, D. E. *et al.* Thermodynamics of the interactions of Lac repressor with variants of the symmetric Lac operator: effects of converting a consensus site to a non-specific site. *J. Mol. Biol.* **267**, 1186–1206. <https://doi.org/10.1006/jmbi.1997.0920> (1997).
42. Capp, M. W. *et al.* Compensating effects of opposing changes in putrescine (2+) and K+ concentrations on lac repressor-lac operator binding: *in vitro* thermodynamic analysis and *in vivo* relevance. *J. Mol. Biol.* **258**, 25–36. <https://doi.org/10.1006/jmbi.1996.0231> (1996).
43. Cayley, S., Lewis, B. A., Guttman, H. J. & Record, M. T. Characterization of the cytoplasm of *Escherichia coli* K-12 as a function of external osmolarity implications for protein-DNA interactions *in vivo*. *J. Mol. Biol.* **222**, 281–300. [https://doi.org/10.1016/0022-2836\(91\)90212-o](https://doi.org/10.1016/0022-2836(91)90212-o) (1991).
44. Richeys, B. *et al.* Variability of the intracellular ionic environment of *Escherichia coli* differences between *in vitro* and *in vivo* effects of ion concentrations on protein-DNA interactions and gene expression. *J. Biol. Chem.* **262**, 7157–7164 (1987).
45. Harries, D., Rosgen, J. Vol. 84 679–735 (Elsevier, 2008).
46. Massiah, M. A., Wright, K. M. & Du, H. Obtaining soluble folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS: application to LB and M9 minimal media. *Curr. Protoc. Protein Sci.* **84**, 1–24. <https://doi.org/10.1002/0471140864.ps0613s84> (2016).
47. Pettitt, B. M. & Bolen, D. W. Protein folding, stability, and solvation structure in osmolyte solutions. *Biophys. J.* **89**, 2988–2997. <https://doi.org/10.1529/biophysj.105.067330> (2005).
48. Singh, L. R. *et al.* Forty years of research on osmolyte-induced protein folding and stability. *J. Iran. Chem. Soc.* **8**, 1–23 (2011).
49. Ajito, S., Iwase, H., Takata, S.-I. & Hirai, M. Sugar-mediated stabilization of protein against chemical or thermal denaturation. *J. Phys. Chem. B* **122**, 8685–8697. <https://doi.org/10.1021/acs.jpcc.8b06572> (2018).
50. Pérez-Bercoff, Á. *et al.* Draft genome of Australian environmental strain WM 09.24 of the opportunistic human pathogen *Scedosporium aurantiacum*. *Genome Announc.* **3**, e01526–01514. doi:<https://doi.org/10.1128/genomeA.01526-14> (2015).
51. Duvaux, L. *et al.* Draft genome sequence of the human-pathogenic fungus *Scedosporium boydii*. *Genome Announc.* **5**, e00871–e817. <https://doi.org/10.1128/genomeA.00871-17> (2017).
52. Kuo, K. C., Kuo, H. C., Huang, L. T., Lin, C. S. & Yang, S. N. The clinical implications of ABO blood groups in *Pseudomonas aeruginosa* sepsis in children. *J. Microbiol. Immunol. Infect.* **46**, 109–114. <https://doi.org/10.1016/j.jmii.2012.01.003> (2013).
53. Scanlin, T. F. & Glick, M. C. Terminal glycosylation in cystic fibrosis. *Biochim. Biophys. Acta* **1455**, 241–253. [https://doi.org/10.1016/S0925-4439\(99\)00059-9](https://doi.org/10.1016/S0925-4439(99)00059-9) (1999).
54. Taylor-Cousar, J. L. *et al.* Histo-blood group gene polymorphisms as potential genetic modifiers of infection and cystic fibrosis lung disease severity. *PLoS ONE* **4**, e4270. <https://doi.org/10.1371/journal.pone.0004270> (2009).
55. Wimmerova, M., Mitchell, E., Sanchez, J. F., Gautier, C. & Imberty, A. Crystal structure of fungal lectin: six-bladed β -propeller fold and novel fucose recognition mode for *Aleuria aurantia* lectin. *J. Biol. Chem.* **278**, 27059–27067. <https://doi.org/10.1074/jbc.M302642200> (2003).
56. Audfray, A. *et al.* Fucose-binding lectin from opportunistic pathogen *Burkholderia ambifaria* binds to both plant and human oligosaccharidic epitopes. *J. Biol. Chem.* **287**, 4335–4347. <https://doi.org/10.1074/jbc.M111.314831> (2012).

57. Kostlanova, N. *et al.* The fucose-binding lectin from *Ralstonia solanacearum*: a new type of β -propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose, and plant xyloglucan. *J. Biol. Chem.* **280**, 27839–27849. <https://doi.org/10.1074/jbc.M505184200> (2005).
58. Houben, K., Marion, D., Tarbouriech, N., Ruigrok, R. W. H. & Blanchard, L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J. Virol.* **81**, 6807–6816. <https://doi.org/10.1128/jvi.00338-07> (2007).
59. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132. doi:<https://doi.org/10.1107/S0907444909047337> (2010).
60. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242. <https://doi.org/10.1107/S0907444910045749> (2011).
61. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674. <https://doi.org/10.1107/S0021889807021206> (2007).
62. Murshudov, G. N. & Nicholls, R. A. REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367. <https://doi.org/10.1107/S0907444911001314> (2011).
63. Emsley, P. & Lohkamp, B. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501. <https://doi.org/10.1107/S0907444910007493> (2010).
64. Aguirre, J. *et al.* Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834. <https://doi.org/10.1038/nsmb.3115> (2015).
65. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, 320–324. <https://doi.org/10.1093/nar/gku316> (2014).

Acknowledgements

We would like to thank to Valérie Chazalet and Emilie Gillon for their technical assistance. We are grateful to the Consortium for Functional Glycomics for providing the glycan array resource. We also would like to thank for access to the beamlines BM30A-FIP at the European Synchrotron Radiation Facility (ESRF), Grenoble, France where initial tests were performed and Proxima 1 at SOLEIL Synchrotron, Saint Aubin, France (Proposal Number 20170827), where final data were collected. Thanks to our local contacts Jean-Luc Ferrer, Serena Sirigu, and Pierre Legrand for their assistance and technical support.

Author contributions

D.M.A. performed cloning, protein expression, protein purification, protein labelling, hemagglutination, isothermal microcalorimetry measurements. V.B. performed the fluorescence microscopy experiments. D.M.A. participated in all steps of protein crystallography and data analysis, wrote the original draft of the manuscript and prepared all figures. J.-P.B. provided the genome data for the bioinformatics studies and contributed to revise the manuscript. R.J.P. participated in the supervision of D.M.A., funding acquisition and revision of the manuscript. A.V. administered the project, conceived the design of the study, obtained funding, evaluated the results, contributed to data analysis, supervised D.M.A., helps in the preparation of the manuscript and contributed to revise the manuscript and the figures. All authors have read and agreed to the published version of the manuscript.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant (H2020-MSCA-ITN-2017-EJD-765581), from the French cystic fibrosis association Vaincre la Mucoviscidose. The project was also supported by Glyco@Alps (ANR-15-IDEX02).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-95008-4>.

Correspondence and requests for materials should be addressed to A.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Supplementary information

Biochemical and structural studies of target lectin SapL1 from the emerging opportunistic microfungus *Scedosporium apiospermum*

**Dania Martínez-Alarcón^{1,2}, Viviane Balloy³, Jean-Philippe Bouchara⁴, Roland J. Pieters²
Annabelle Varrot[†].**

1 Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France.
dania.martinez.alarcon@gmail.com, annabelle.varrot@cermav.cnrs.fr

2 Utrecht University, 3584 CG Utrecht, The Netherlands. R.J.Pieters@uu.nl

3 Sorbonne Université, UPMC Univ. Paris 06, Inserm, Centre de Recherche Saint-Antoine Paris, Paris, France. viviane.balloy@inserm.fr

4 Host-Pathogen Interaction Study Group, GEIHP, EA 3142, SFR ICAT 4208, UNIV Angers, UNIV Brest, Institut de Biologie en Santé, IRIS, CHU d'Angers, Angers, France. jean-philippe.bouchara@univ-angers.fr

* Corresponding author

Email: annabelle.varrot@cermav.cnrs.fr

S1 Table. SapL1 expression conditions assayed.

Strain	vector	Media	Inductor	Inductor concentration	A ₆₀₀ at induction	Induction temperature	Expression length	
BL21(DE3)	pET-TEV	LB	IPTG	100 μM	0.8	16°C	Overnight	
	pProNde			250 μM		25°C	36 h	
				100 μM				
				250 μM				
				150 μM				
	pET32-TEV			100 μM		16°C	1 h	
				50 μM			2 h	
							3 h	
							4 h	
							5 h	
							6 h	
							25 μM	Overnight
				10 μM				
5 μM								
Tuner™(DE3)	pProNde	100 μM	2	16°C	Overnight			
		250 μM	1.5					
		250 μM	2					
		100 μM	2.5					
		250 μM	0.8					
	pET32-TEV	10 μM	0.8			1		
		100 μM						
		250 μM						
		100 μM					0.8	
		250 μM						
BL21Star(DE3) pLysS	pProNde	100 μM	1	16°C	36 h			
		25 μM	0.8					
		100 μM	2					
Rosetta™(DE3) pLysS	pET32-TEV	100 μM	2	16°C	Overnight			
			1.7					
			1.7					
Rosetta-gami 2 (DE3)	pET32-TEV	100 μM	2	16°C	Overnight			
			1.7					
KRX (DE3)	pProNde	LB	Rh	25 μM	0.8	20°C	Overnight	
				100 μM				
				10 μM				
				0.1%				
				0.5%				
				1%				
				2.5%				
				0.1%				
				0.15%				
				0.2%				
0.25%								
BL21trxB (DE3)	pET32-TEV	LB	Rh	0.1%	0.8	16°C	Overnight	
				0.5%				
				0.1%				
				0.15%				
				0.2%				
	0.25%							
	pProNde			0.1%				2
				0.15%				
				0.2%				
				0.25%				
0.5%								
BL21(DE3)	pET32-TEV	LB	Rh	0.75%	0.8	16°C	Overnight	
				1%				
BL21trxB (DE3)	pProNde	LB	IPTG/Rh	50 μM/ 1%	0.8	16°C	Overnight	
				100 μM/ 1%				
BL21(DE3)	pProNde	LB	IPTG/Rh	50 μM	0.8	16°C	Overnight	
				IPTG/Gly				
BL21(DE3)	pProNde	LB	IPTG/Rh	50 μM/ 1%	0.8	16°C	Overnight	
				IPTG/Gly				

```

SAPIO_CDS9261 MVDLGSMTEAATFLIKKYRMIFAEITKGDRLRGGEKSRRRQLEKYGLISQTQDTSSAKSNYSSESLSPQNQLAMSGVLQ 80
FleA          -----MSTPGAQQ 6
                :: * . *

SAPIO_CDS9261 ISFPAGIAAIRNNSLRVYEALDGGVREAQYEGRWAGGKPDNVIATGKIGTPIAATSVGFQYIRVYVYGADNKAREACW 160
FleA          VLFRTGIAAVNSTNHLRVYFQDVYGSIRESLYEGSWANGTEKNVIGNAKLGSVAATSKEKELHIRVYTLTEGNTLQEFAY 88
                : * :*****:..... ***** : *::***: *** **.*. .***...*:*:***** ::***** : .*. :* ..

SAPIO_CDS9261 -DGKGWYTGAFV---KDVAPYSSIGAVFLGK--NIVVRVYTNHDNTIQEWVWDSPTGWTAGANFGAALPGTAIAATSW 234
FleA          DSGTGWYNGGLGGAKFQVAPYSCIAAVFLAGTDALQLRIYAQKPDNTIQEYMWNG--DGWKEGTLNGLGALPGTGIGATSF 166
                .*.***.*.: :*****.*.*****. :*:*:*: *****:*.:. ** .*:*:*.*****.*.***:

SAPIO_CDS9261 GAGPY---HIRVYFQDTNRNVIESGWD-GSGWYTGGLKISN-QSPRASLGATSWGESGSSLGIRLYYATQDNLIKEKAWD 309
FleA          RYTDYNGPSIRIWFQTDLKLVRAYDPHKGWYDPLVTIFDRAPPRTAIAATSFGAGNSSIYMRIYFVNSDNTIWWQVCWD 246
                * **::** : :::: .* .*** . :* : **::*.***.* .** :*:*.***.* : .**

SAPIO_CDS9261 GGGGWYDGGFQQR SIPGSRVAIP-----LPVLRVYLQNGTEVSGITEYAWNSG--WVVGQAVLPPA 342
FleA          HGKGYHDKGTITPVIQGSEVAIISWGSFANNGPDLRLYFQNGTYISAVSEWVWNRAGHSQ LGRSALPPA 286
                * *:* * * * **.* ** * * **:*:**** :*::*:** . :*:*.****

```

Figure S1. Identification of SapL1. Alignment of the SAPIO_CDS9261 and FleA sequences. In red, the first 74 residues that were removed from the sequence of SapL1 synthetic gene. Black arrow indicates the first methionine of the recombinant SapL1

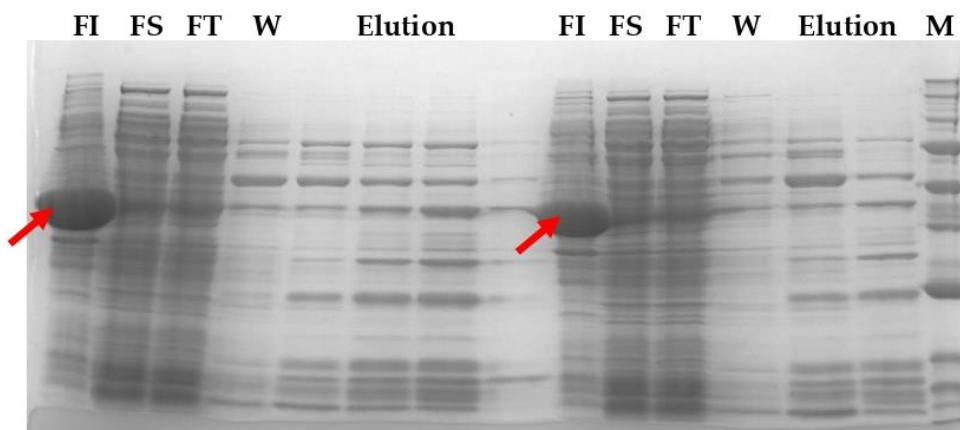


Figure S2. SapL1 production. SDS-PAGE of fractions collected from SapL1 purification using pET32-TEV vector. From right to left: insoluble fraction(IF), Soluble fraction (SF), flow-through (FT), wash (W) and elution fractions, respectively ending with the molecular weight marker (M). The left half of the figure shows the purification fractions in an experiment performed with 0.025 mM of IPTG as inducer of SapL1 expression, while the right half of the gel shows the same distribution of samples from an experiment performed with 0.05 mM of IPTG. Molecular weight marker bands: 250, 150,100,75, 50, 37, 25, 20, 15, 10 kDa from top to bottom. Red arrow indicates inclusion bodies from insoluble fractions.

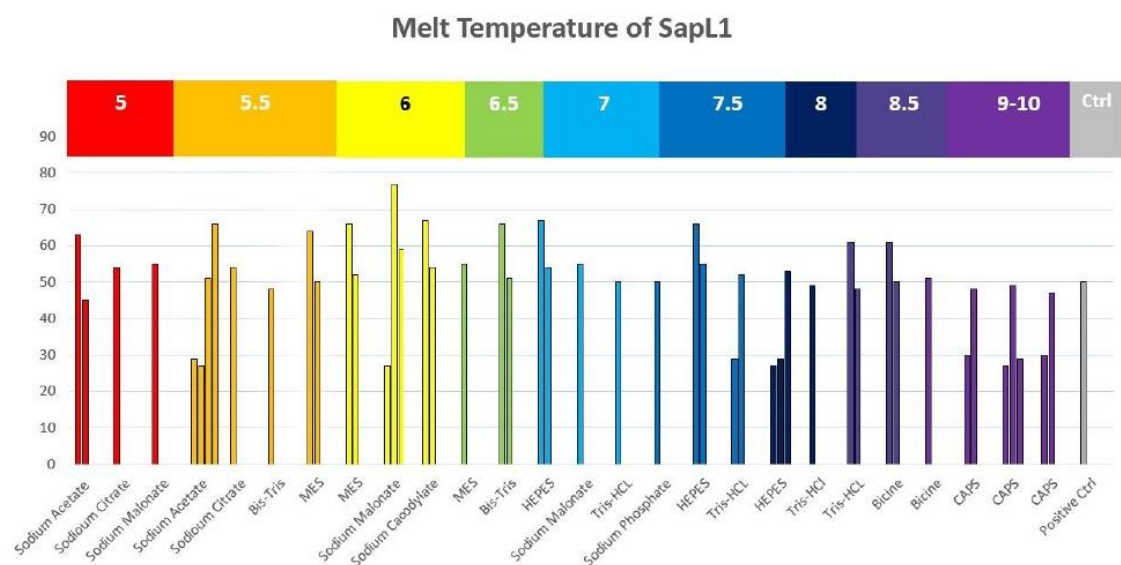


Figure S3. SapL1 thermal Stability. Melting temperatures of SapL1 obtained through the Thermal shift assay (TSA). A temperature gradient from 20 to 100°C in was applied under 26 different buffer conditions with a pH range from 5 to 10.

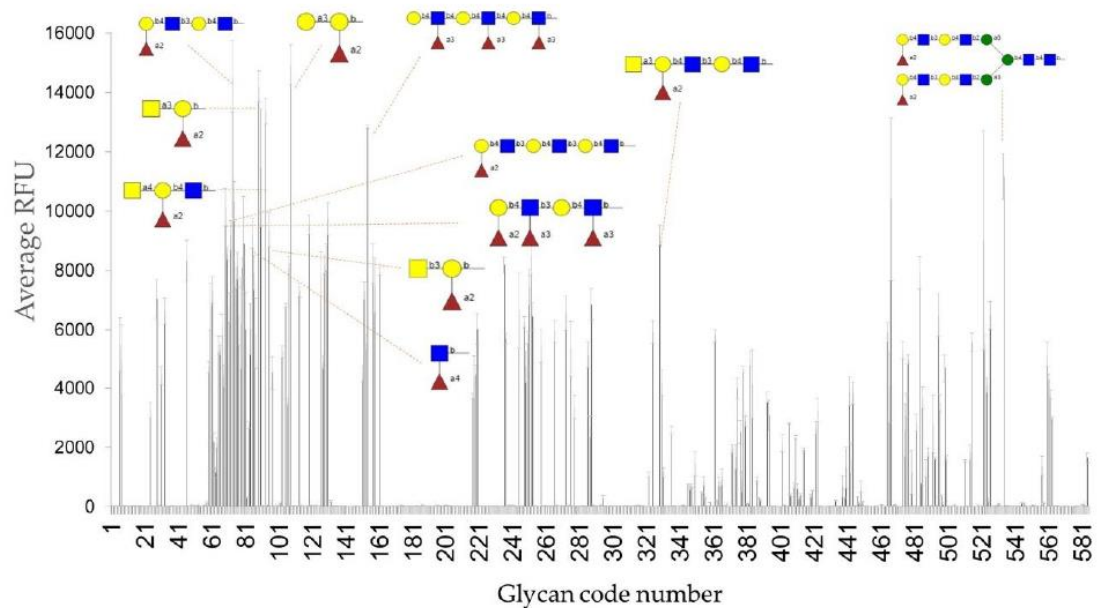


Figure S4. Glycan array signals. Relative Fluorescent Units (RFU) plot of the glycan array matrix with SapL1 at 50 $\mu\text{g}\cdot\text{ml}^{-1}$. The structures of the ten best binders are represented linked to their respective signals.

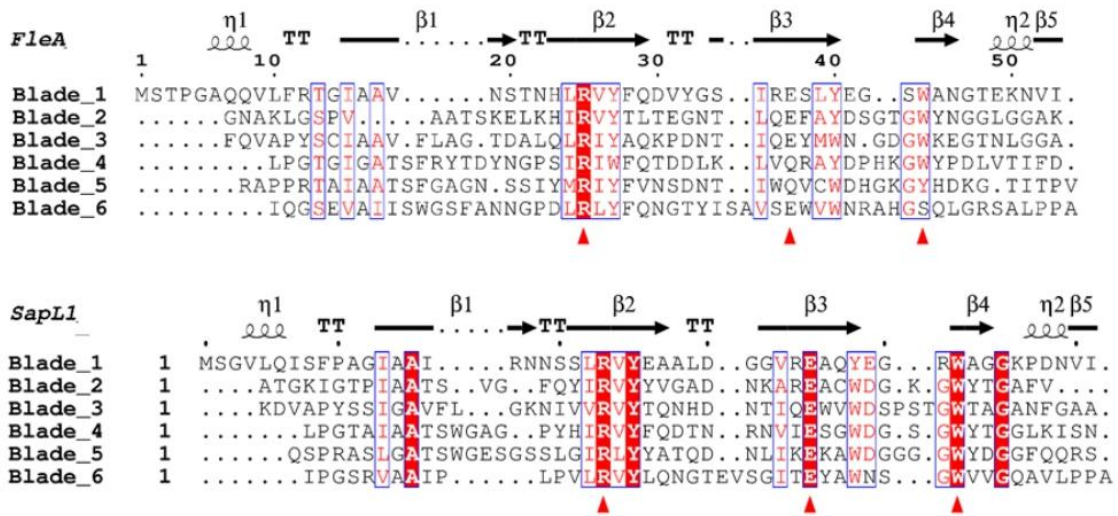


Figure S5. Conservation of blades in SapL1 and FleA structures. The conserved (SapL1) and semi conserved (FleA) triads of amino acids involved in ligand binding by hydrogen bonds are indicated with red triangles.

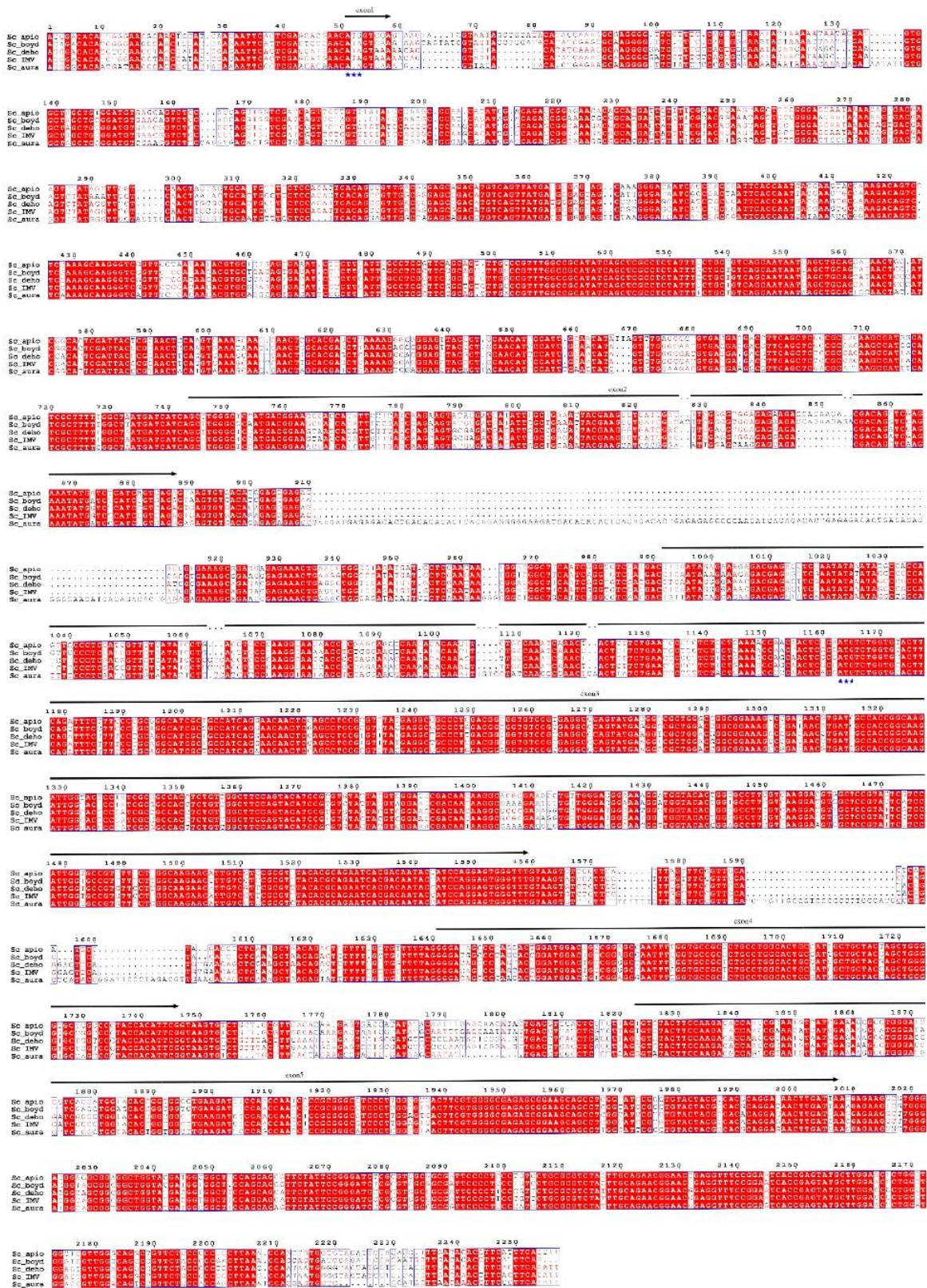


Figure S6. Alignment of the transcript containing *sapL1* from all *Scedosporium* species with sequenced genomes. Sc_apio: *S. apiospermum* SEQ_SAPIO_0132:1935229:1937485¹, Sc_boyd: *S. boydii* strain IHEM 23826 contig 288², Sc_deho: *S. dehoogii* strain 120008799-01/4 contig 122, SC_IMV, *S. sp.* IMV 00882 and Sc_aura: *S. aurantiacum* strain WM 09.24 scaffold-55³. Sequence were obtained from NCBI genome page. Prediction exons are depicted by arrows at the top and the two possible initial codons by blue stars at the bottom (positions 51 and 1163). Alignment done using Multalin⁴ and figure drawn in ESPript 3.0⁵.

References

- 1 Vandeputte, P. *et al.* Draft genome sequence of the pathogenic fungus *Scedosporium apiospermum*. *Genome Announc.* **2**, e00988–00914, doi:10.1128/genomea.00988-14 (2014).
- 2 Duvaux, L. *et al.* Draft genome sequence of the human-pathogenic fungus *Scedosporium boydii*. *Genome Announc.* **5**, e00871–00817, doi:10.1128/genomeA.00871-17 (2017).
- 3 Pérez-Bercoff, Å. *et al.* Draft genome of Australian environmental strain WM 09.24 of the opportunistic human pathogen *Scedosporium aurantiacum*. *Genome Announc.* **3**, e01526–01514, doi:10.1128/genomeA.01526-14 (2015).
- 4 Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890, doi:10.1093/nar/16.22.10881 (1988).
- 5 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324, doi:10.1093/nar/gku316 (2014).

Raw data of the glycan array assay can be consulted at https://docs.google.com/spreadsheets/d/1lyGh8bEw3kC_pgTcUaLepVxbAY7XaRRm/edit?usp=sharing&ouid=111416367677195406970&rtpof=true&sd=true

4. BIOCHEMICAL AND STRUCTURAL CHARACTERIZATION OF A CYANOVIRIN-LIKE LECTIN FROM HUMAN OPPORTUNISTIC FUNGI

4.1 Summary

In this chapter, we describe the identification and characterization of a Cyanovirin-like lectin from *S. apiospermum* (SapL6). SapL6 codifying sequence was identified by data mining using the sequence of a homologous protein from *A. fumigatus*, which has been previously studied by our research group (AFL6). Both, AFL6 and SapL6 belong to the recently discovered Cyanovirin-N homolog (CVNH) family.

Although the biological function of this family of proteins is unclear, the relevance of SapL6 fungal life is evidenced by its constitutive expression during the whole life cycle of *S. spiospermum* [87]. Hence, it represents interesting target for the development of antifungal therapies. Besides, its evolutive relationship with the powerful antiviral protein Cyanovirin-N (CN-V), brightens up the interest to evaluate its possible applications as biotechnological tool. Together, these factors evidence the relevance of approaching the study of this protein.

This work includes a detailed strategy to achieve the recombinant expression of SapL6 in *E. coli*, its biochemical characterization and its 3D-structure solved by X-ray crystallography. The analysis of its specificity was also addressed by different approaches, such as glycan array and ITC and co-crystallization with glycans. However, to date, it has not been possible to identify its ligand. A graphical abstract of this part of the project is shown in figure 4.1.

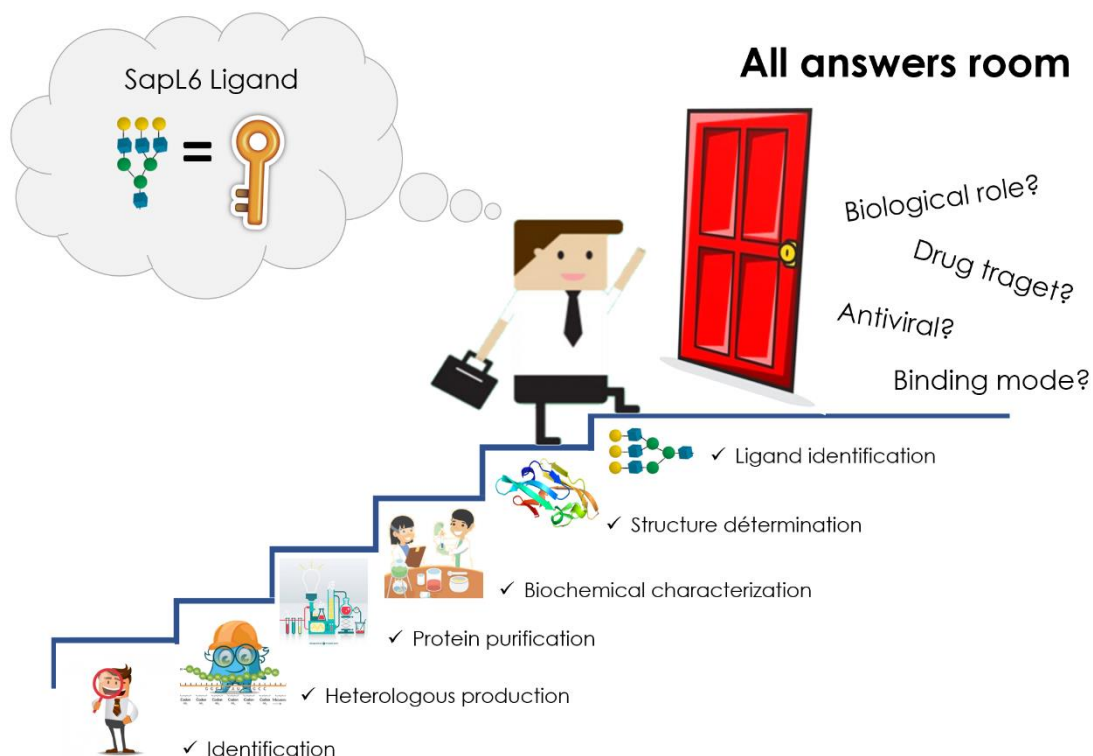


Figure 4.1: Graphical abstract of project 2. Production and characterization of SapL6. Figure created with Microsoft Office 365.

4.2 Introduction

4.2.1 Cyanovirin-N

Cyanovirin-N (CV-N) is a small protein from the cyanobacterium *Nostoc ellipsosporum* that binds mannosides and possesses strong virucidal properties against a wide range of viruses (*e.g.*, HIV, ebola, influenza, hepatitis C, herpes virus, etc.) [86,88]. Although the natural ligand and the biological role of CV-N in *N. ellipsosporum* is unknown, its potential pharmacological use has motivated a large number of studies to elucidate its antiviral mechanism of action, which has turned out to be specific to each case. In HIV, for example, inhibition is mediated by high affinity interactions with the viral envelope glycoproteins gp120 and gp41, which are highly mannosylated [65]. This interaction disables the recognition of the host cell CD4 and CCR5 receptors and prevents

conformational changes that are necessary for membrane fusion and viral entry into the target cells [88,89].

4.2.2 The CV-N fold

Cyanovirin-N is composed by two homologous repeats of ~50 amino acids that fold into a pair of two interconnected pseudosymmetric domains formed by strand exchange called domains A and B (Figure 4.2) [59,90]. Each domain comprised one helical turn a three-stranded β -sheet, a β -hairpin and an intramolecular disulfide bond. This provides the protein with great stability and forms two symmetry-related carbohydrate binding pockets located at opposite ends of the protein's long axis (BP1 and BP2) [85]. These binding sites exhibit differing affinities for the disaccharide $\text{Man}\alpha 1\text{-2Man}$. BP2 shows the highest affinity (K_a of $\sim 7.2 \times 10^{-6} \text{ M}^{-1}$), whilst the affinity for BP1 is 10 times lower ($K_a \sim 6.8 \times 10^{-5} \text{ M}^{-1}$) [90].

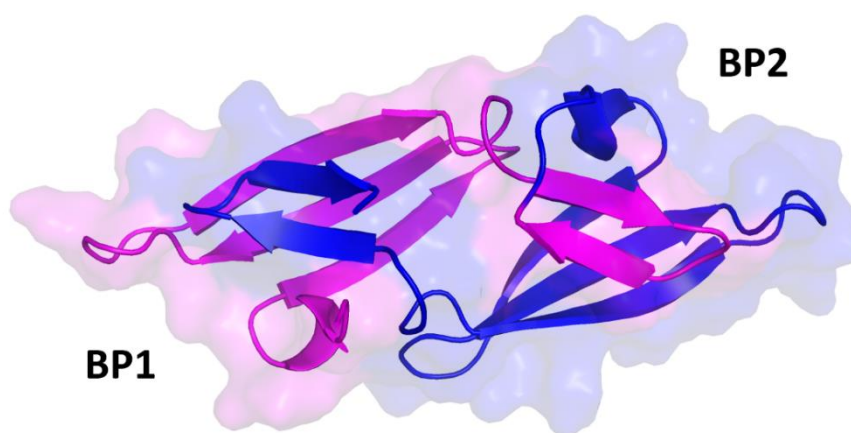


Figure 4.2. CV-N Structure. Protein model was visualized and edited with PyMol Version 1.8.4 with the PDB 1IIY. The two sugar binding pockets are indicated as BP1 and BP2.

4.2.3 Cyanovirin-N homologs (CVNHs)

Given the efficacy of CV-N as a virucide, there was great interest in the identification of similar proteins after its discovery. In 2005, an *in-silico* screening conducted by Percudani and coworkers led towards the identification of a group of CVN-related proteins that were named CyanoVirin-N Homologous

(CVNHs) [85]. Originally, all members of this protein family were identified from eukaryotic organisms such as filamentous ascomycetes and seedless plants, however, in further studies CVNHs have also been identified from prokaryotic cells [88,91]. The phylogenetics of this family has been divided into three major subgroups of CVNHs:

- **CVNH I** is represented by the fern *Ceratopteris richardii* lectin (CrCVNH). This group includes CV-N, the prokaryotic CV-like proteins MVN and CaVN, five fungal CVNHs and two fern CVNHs.
- **CVNH II** is represented by the truffle *Tuber borchii* lectin (TbCVNH). This group comprise proteins from filamentous Ascomycetes whose sequence is closely related with TbCVNH.
- **CVNH III** is represented by the mold *Neurospora crassa* lectin (NcCVNH). This group is entirely composed of CVNH polypeptides from filamentous Ascomycetes whose sequences closely related with NcCVNH.
- **Others:** CVNHs, that did not cluster with any of the three major groups.

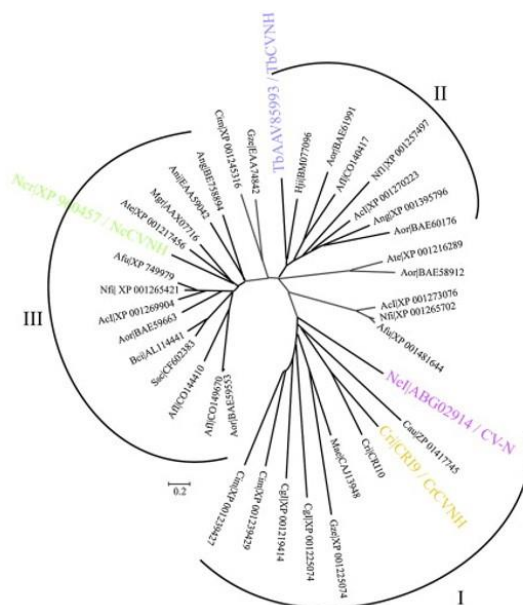


Figure 4.3. Phylogenetic tree of the CVNH Family. Cladogram of 39 predicted CVNH polypeptide sequences (labeled in color along with the three CVNH representative members and CV-N). Adapted with permission from Koharudin *et al.* 2008 [59].

In 2008, the representative members of each subgroup of CVNHs family were structurally characterized; TbCVNH (PDB: 2JZK), NcCVNH (PDB: 2JZL) and CrCVNH (PDB: 2JZJ). According to the NMR data, the overall fold of the three proteins closely reassembles the one previously described for CV-N, *e.i.* in all cases, an ellipsoid composed of two pseudo-symmetric domains containing a pair of carbohydrate-binding sites was observed. The most conserved were the residues of the hydrophobic core and those located at the interface between domains A and B. The latter contribute to orient the protein and are intimately related to the fold preservation.

On the other hand, despite their high structural similarity, noteworthy divergences are identified for some CVNH members, such as the absence of S-S bonds in the fungal members and the poor conservation of the putative sugar-binding pockets, which is common in protein families that interact with multiple ligands [59,85]. Additionally, extended loops are present in the fungal members of the family and this may stabilize the conformation of local regions.

4.2.4 CVNHs carbohydrate binding properties

Chemical shift perturbations monitored by ^1H - ^{15}N HSQC spectroscopy, showed that the three representative members of each CVNH subfamily bind to Man α 1-2Man as Cyanovirin-N. Only the fern homolog has two functional binding sites, while for the fungal proteins, only one binding site is active [59]. Functional domains were A in TcCVN and B in NcCVN (Figure 4.4 A-C). The specificity of these two proteins was further addressed by a glycan array screening, through which it was shown that NcCVNH interacts with high mannose glycoconjugates, similarly to CV-N. On the other hand, TbCVNH only recognized two of the oligomannosides present in the array but also bound mannose-lacking glycans composed exclusively by glucose or a combination of glucose, galactose and N-acetylglucosamine (Figure 4.5 D) [59].

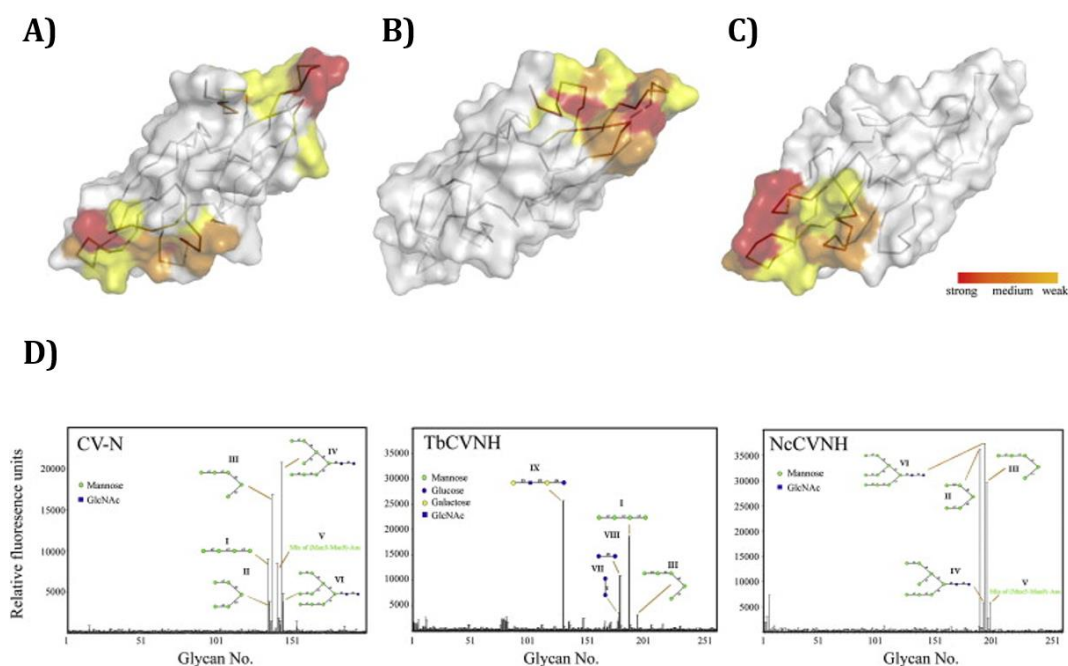


Figure 4.4. CVNHs sugar binding specificity. A–C) Structural mapping of sugar-binding sites onto space-filling representations of the 3D structures of TbCVNH, CrCVNH, and NcCVNH. Residues corresponding to strongly, moderately, and weakly affected resonances are red, orange, and yellow, respectively. D) Glycan array screening of CV-N fungal homologous TbCVN and NcCVN. Adapted with permission from Koharudin *et al.*, 2008 [59].

Additional NMR titrations performed with sugars not represented in the glycan array, resulted in the identification of sucrose binding. According to the pattern of the affected resonances, it was concluded that the sucrose binding pocket was the same as the mannose one [59]. Sucrose binding appears to be exclusive of TbCVN and it is speculated that this disaccharide could mimic the shape of an unidentified natural ligand.

4.2.5 CVNHs biological roles

It has been suggested that fungal CVNHs could be nutrient state-regulated lectins. This theory arose from the observation, by northern blot and microarray hybridization, of a severe down-regulation of TcCVN mRNA (M5G11) upon carbon or nitrogen starvation [85]. This was later supported by the reduced immunodetection of TbCVNH in mycelia grown under glucose-limiting conditions [59]. On the other hand, a recently identified *N. crassa* mRNA

encoding a protein identical in sequence to NcCVNH could be related to the metabolic/nutritional changes that accompany the circadian rhythm in this organism [92]. Taken together, these data could implicate potential physiological functions of CVNH, *e.g.*, cell-cell interaction through the recognition of specific sugars (and/or glycoproteins) bound to the surface or the detection/transduction of associated specific metabolic signals. Furthermore, nitrogen deficiencies are believed to be important environmental signals that trigger plant infection, as well as other adaptive morphogenetic transitions in fungi [93]. Hence a putative role during pathogenesis should not be ruled out.

4.2.6 CVNHs subcellular localization

Although the subcellular localization of CV-N in *N. ellipsosporum* is unknown, it is predicted to be a secreted protein since it lacks its N-terminal methionine. The secretory theory is reinforced by the presence of disulfide bonds, another signature characteristic of extracellular proteins [85]. CVNHs and in particular the fungal CNVHs are also predicted to be secreted despite the absence of signal peptide. Both TbCVNH and NcCVNH are found primarily in sedimentary subcellular fractions (membrane and organelles) of macerated fungal cells and associate with the surface of hyphae [59]. The secretion mechanism of these proteins is unknown as for many fungal lectin which are also found extracellularly without secretion signal.

4.2.7 AFL6

Previously, our research group has undertaken the study of AFL6, a hypothetical CVNH from *A. fumigatus* [94]. This protein was originally predicted by Koharudin *et al* in 2008 and it was one of the 39 sequences used to establish the phylogeny of CVNH proteins. AFL6 was produced recombinantly in *E. coli* and its structure was solved by X-ray crystallography. As expected, AFL6 shares the common CV-N fold and, in agreement with the other fungal members of the CVNH family, it does not present hemagglutinating activity.

According to the original classification of CVNHs, AFL6 belongs to the subfamily represented by the *N. crassa* lectin (NcCVNH), however a glycan array assay showed that, instead of highly mannosylated structures, AFL6 preferentially binds to glycans containing biantenary polylactosamine, such as TbCVNH from *T. borchii*. Nevertheless, no interaction with lactosamine could be confirmed by ITC and its affinity for mannose was very low.

4.3 Material and methods

4.3.1 Genetic construction

The coding sequence for SapL6 was optimized for expression in *E. coli* and ordered at Eurofins Genomics (Ebersberg, Germany). SapL6 was excised from the commercial vector where it was inserted, through standard digestion (2 h at 37° C) with the restriction enzymes NdeI and XhoI. pProNde and pET-TEV [95] vectors were digested with identical restriction patterns. pProNde is a homemade vector which results from the mutation of the NcoI to NdeI restriction site in pProEx HTb vector (EMBL, Heidelberg). All products of digestion were verified by electrophoresis in 0.8% agarose gels, from where the linearized vectors and SapL6 codifying sequence were recovered and purified by using the commercial kit of BioLabs “Monarch DNA Gel Extraction”. Then, SapL6 insert and digested plasmids were ligated using the Fermentas T4 ligase, in a ratio of 3:1. Reactions were incubated at 25°C for 1 h and were used to transform chemocompetent cells of *E. coli* TOP10 by thermic shock at 42 °C. Transformed cells were recovered for growth in LB medium plates with the corresponding antibiotic (Ampicillin 100 µg/mL for pET-TEV and Kanamycin 30 µg/mL for pProNde) and incubated at 37 °C. After 16 h, a PCR colony screening was performed using a set of primers designed to hybridize sequences flanking the gene insertion in each vector and positive hits were grown in LB media to perform plasmid extraction. Finally, gene insertion was confirmed on the

purified plasmids by digestion and the correct reading frame was validated by sequencing.

4.3.2 Production of SapL6

For heterologous production of SapL6, the recombinant plasmids (previously validated) were introduced into *E. coli* BL21(DE3) strain by thermal shock at 42°C. Transformed cells were recovered for growth in LB plates with the corresponding antibiotic (Ab) and colony PCR screening was performed to confirm the plasmid insertion. A positive hit from each vector was grown in LB+Ab media and incubated at 37°C with orbital agitation (160 rpm) until it reaches an optical density (OD₆₀₀) of 0.6. The temperature was then lowered to 16°C and when OD₆₀₀ reached 0.8, the induction was carried out overnight by the addition of isopropyl-β-D-1-thiogalactoside (IPTG) to a final concentration of 0.1 mM.

4.3.3 Purification of SapL6

Cultures were harvested by centrifugation at 5000 g for 10 min and resuspended on buffer A (50 mM Tris-HCl, 500 mM NaCl, pH 7.5) with moderate agitation for 30 minutes. Then, cells were disrupted at 1.9 kbar using one shot cell disrupter (Constant system Ltd) and cell debris was pelleted by centrifugation at 22000 g for 30 minutes. The supernatant was filtered with 0.45 μm membranes and the protein purification was carried out by immobilized metal affinity chromatography (IMAC) with Nickel-NTA His-Trap FF columns (GE Healthcare Life Sciences) on a NGC™ chromatography system (Bio-Rad) with a flow rate of 1 mL·min⁻¹. Once the supernatant was loaded, column was rinse thoroughly with buffer A until stabilization of the baseline. Bound protein was subsequently eluted through a gradient of 5–500 mM imidazole. SapL6 containing fractions were reconcentrated by ultrafiltration and sample buffer was replaced by buffer C (50 mM Tris-HCl, 100 mM NaCl, pH 7.5) using PD10 desalting column (GE

healthcare Life Sciences). Then, the protein sample was supplemented with 0.5 mM EDTA, 0.25 mM TCEP and TEV protease (1:50 ratio) for overnight His-tag cleavage at 19°C. TEV protease and uncleaved His-SapL6 were separated by IMAC and cleaved protein was collected into the flow through.

4.3.4 Size exclusion chromatography

Size exclusion chromatography (SEC) was performed using a High-Resolution ENrich™ SEC 70 column (Bio-Rad) on NGC™ chromatography system (Bio-Rad). Prior assay, protein sample was prepared at 10 mg·mL⁻¹ and centrifuged 30 min at 12000 rpm. Column was equilibrated with 50 ml of buffer D (20 mM MES, 100mM NaCl, pH 6) and 200 µl of sample were injected followed by 40 ml isocratic elution on buffer D. Fractions were monitored by absorbance at 280 nm and collected every 0.5 mL.

4.3.5 Dynamic light scattering (DLS)

DLS analyses were performed using a Zetasizer™ NanoS (Malvern Panalytical) using a 40 µl quartz cuvette. Prior to measurements, a solution of pure protein at 1 mg·mL⁻¹ on buffer C (50 mM TRIS-HCl, 100 mM NaCl, pH 7.5) was prepared and centrifuged during 30 minutes at 12,000 rpm in microcentrifuge tubes. Corresponding quartz cells were filled with sample solution and three successive DLS measurements were performed per sample.

4.3.6 Thermal Shift Assay (TSA)

The thermal stability of SapL6 was analyzed by TSA with the MiniOpticon real-time PCR system (Bio-Rad, Hercules, CA, USA). Prior assay, buffer stocks at 100 mM and a mixture containing 70 µl of SapL6 at 1 mg mL⁻¹, 7 µl of 500x Sypro Orange (Sigma-Aldrich) and 63 µl of H₂O were prepared. Then, 7.5 µl of H₂O, 12.5 µl of the corresponding buffer and 5 µl of the protein/Sypro mixture were mixed in 96-well PCR microplates. The heat exchange test was then carried out from 20°C to 100°C with a heating rate of 1°C·min⁻¹. Fluorescence intensity was

measured with Ex/Em: 490/530 nm and the data processing was performed with the CFX Manager software.

4.3.7 Isothermal Titration Calorimetry (ITC)

Experiments were performed using a Microcal ITC200 calorimeter (Malvern Panalytical, Malvern, UK) with 40 μL of ligand and 200 μM of protein according to the conditions described in Table 4.1. Both, protein, and sugars were gradually added to the sample cell by 2 μL injections in a range of 120 seconds while stirring at 1000 rpm. Experimental data were adjusted to a theoretical titration curve by the Origin ITC Analysis software.

4.3.8 Hemagglutination assay

Agglutination test was performed with fresh rabbit erythrocytes (BioMerieux, Lyon) in 96-well plates. For the test, 150 mM NaCl was used as a negative control and rPVL from *Psathyrella velutina*, as a positive control. Aliquots of each sample were prepared at 0.1 $\text{mg}\cdot\text{mL}^{-1}$ and serial double dilutions were made from the first to the 12 lines. Once the dilutions were made, 50 μL of rabbit erythrocytes 3% were placed into each well and the plate was incubated at room temperature (RT). After 2 hours the result of the experiment was evaluated and agglutination activity was calculated according to the dilution of the protein.

4.3.9 Mammalian Glycan array assay (CFG)

SapL6 was labeled with Fluorescein Isothiocyanate (FITC) (Sigma-Aldrich, St Louis, MO) according to the supplier's instructions with slight modifications. Briefly: Two milligrams of protein were dissolved in 1 mL of buffer E (100 mM Na_2CO_3 , 100 mM NaCl, pH 9), then, 40 microliters of FITC at 1 $\text{mg}\cdot\text{mL}^{-1}$, previously dissolved in DMSO, were gradually added to the protein solution and the mixture was gently stirred at room temperature overnight. Next day, the solution was supplemented with NH_4Cl to a final concentration of 50 mM and free FITC was removed using PD10 desalting column with PBS as mobile phase.

Protein concentration was determined at Abs₂₈₀ and FITC at Abs₄₉₀ using a NanoDrop™ 200 (Thermo Scientific, Hercules, CA, USA), and Fluorescein/Protein molar ratio (F/P) was estimated by the following formula:

$$\text{Molar F/P} = \frac{\text{MW}}{389} \times \frac{A_{495}/195}{[A_{280} - (0.35 \times A_{495})]/E^{0.1\%}} =$$

Where MW is the molecular weight of the protein, 389 is the molecular weight of FITC, 195 is the absorption E_{0.1%} of bound FITC at 490 nm at pH 13.0, (0.35 X A₄₉₅) is the correction factor due to the absorbance of FITC at 280 nm, and E_{0.1%} is the absorption at 280 nm of a protein at 1.0 mg mL⁻¹, Being an ideal F/P should be 0.3 > 1.

Labeled lectin was sent to the Consortium for Functional Glycomics (CFG), (Boston, MA, USA) and binding properties were assayed at 5 and 50 µg mL⁻¹ on a “Mammalian Glycan Array version 5.4” which contain 585 glycans in replicates of 6. The highest and lowest signal of each set of replicates were eliminated and the average of the remaining data were normalized to the percentages of the highest RFU value for each analysis, finally, the percentages for each glycan were averaged at different lectin concentrations.

4.3.10 Crystallization and data collection

Crystal screening was performed using the hanging-drop vapour diffusion technique by mixing equal volumes of pure protein at 5-10 mg·mL⁻¹ and precipitant solutions from commercial screenings of Molecular Dimensions (Newmarket, UK). 2 µL drops were incubated at 19°C until crystals appeared. A subsequent optimization of positive conditions for SapL6 crystallization was carried out and crystals suitable for X-ray diffraction analysis were obtained under solution containing 3 M ammonium sulphate in buffer HEPES 0.1 M pH 7. Data collection was performed on PX1 and PX2 beamlines at SOLEIL Synchrotron (Saint Aubin, FR) using a Eiger detector.

4.3.11 Structure determination

The diffraction patterns collected were processed using XDS 14 software [96] and were converted to structure factors using the CCP4 [97] program package v.6.115, with 5% of the data reserved for Rfree calculation. The structure was determined using the molecular-replacement method with Phaser MR v.2.5.16 [98], with the structure of AFL6 (unpublished data) as starting models. Models' refinement were performed using REFMAC v.5.17 [99] alternated with manual model building in Coot v.0.718. Sugar residues and other compounds that were present were placed manually using Coot.

4.4 Results and discussion

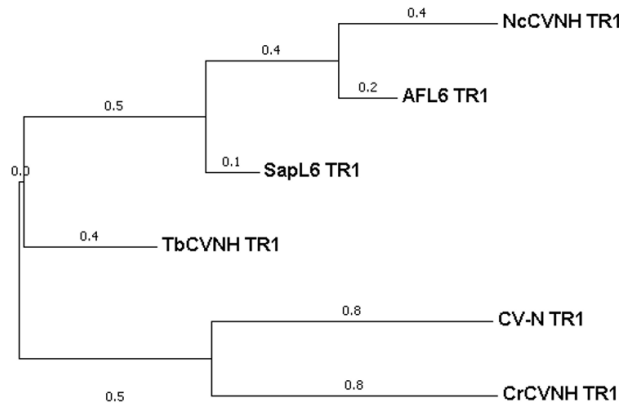
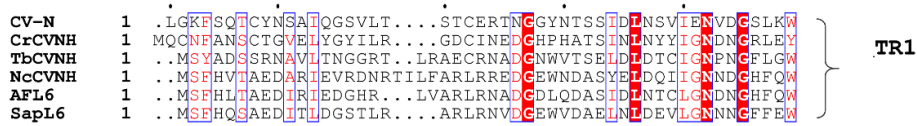
4.4.1 Identification of SapL6

The coding sequence for a hypothetical Cyanovirin-N homolog protein (EMBL accession number XM_016790607.1) was identified through data mining by using AFL6 lectin as input in the genome of *Scenedosporium apiospermum*. This sequence is produced by the transcription of the SAPIO_CDS9133 gene and it corresponds to a hypothetical lectin of ~11 kDa hereinafter referred to as SapL6.

As it is characteristic in CVNHs, the sequence of SapL6 is composed by two homologous tandem repeats (TR) that are predicted to have the same fold and display high identity between them (> 30%). The first tandem (TR1) is comprised from residues 1-49, while TR2 covers from 50-106. Surprisingly, during a multiple alignment of CVNHs sequences and phylogeny, we identified that SapL6-TR1 seems to have more similarities with TbCVNH than with NcCVNH (Figure 4.5). This was unexpected since AFL6 belong to the subfamily of CVNH-III and this family is represented by NcCVNH. Hence, we perform a new phylogenetic analysis with 25 of the original 39 sequences used by Koharudin *et al* in 2008 (Figures 4.6 and 4.18) [59]. The analysis suggests that, although SapL6 also belongs to the CVNH-III subfamily, its evolutionary distance with TbCVNH is

shorter than in the case of AFL6 and NcCVNH. Therefore, their biological role and specificity could also be closely related.

A)



B)

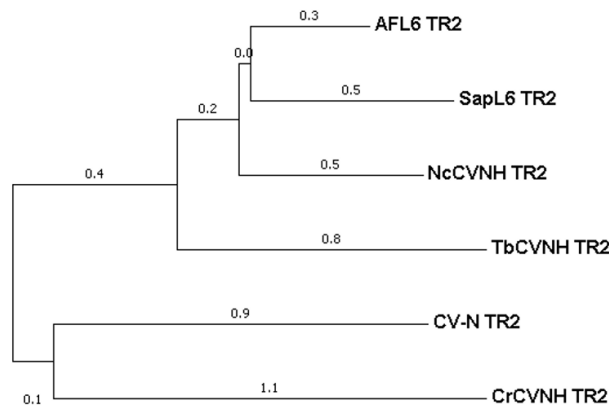
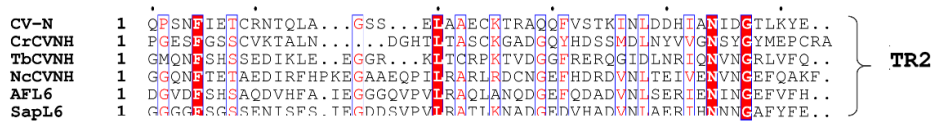


Figure 4.5: Comparison of CVNH tandem repeats. A) Alignment of the sequences of the first tandem repeat of the most representative members of the CVNHs and their phylogenetic tree. B) Alignment of the TR2 sequences of the first tandem repeat of the most representative members and their phylogenetic tree. Sequence's alignments were performed using Clustal W [100] and visualized with ESPrnt3 [101]. Evolutionary histories were inferred using the maximum likelihood method based on the Jones *et al.* algorithm, 1992 [102]. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches). Evolutionary analyzes were performed in MEGA7 [103].

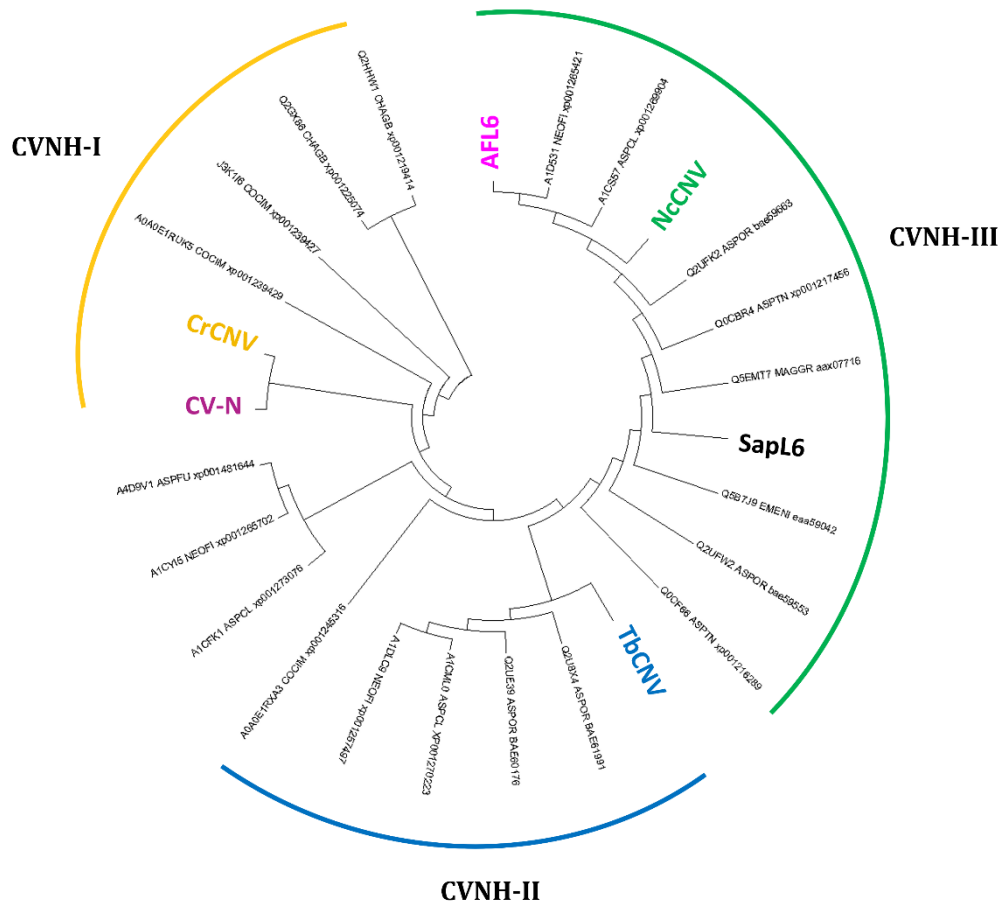


Figure 4.6: Phylogenetic tree of CVNHs. Phylogeny was inferred using 29 of the original 39 sequences used to define the original tree of CVNHs in 2008. Sequence's alignments was performed using Clustal W [100] and the evolutionary history was inferred using the maximum likelihood method based on the Jones *et al.* algorithm, 1992 [102]. Evolutionary analyzes were performed in MEGA7 [103].

4.4.2 Production and purification of SapL6

The coding sequence for SapL6 was fused to an N-terminal His-tag cleavable by Tobacco Etch Virus protease (TEV) by cloning it into two different expression vectors (pET-TEV and pProNde), which were chosen to evaluate the best production performance (Figure 4.7 A and D). Then, expression was induced and purification was carried by Nickel affinity chromatography in two steps before and after TEV cleavage of the His-tag. The cleavage of this flexible fusion allow us to obtain pure protein in the flow-through of the second IMAC chromatography, would prevent possible problems during crystallization and

avoids possible interferences with SapL6 folding and activity. The final yield was $15 \text{ mg}\cdot\text{L}^{-1}$ and $9.31 \text{ mg}\cdot\text{L}^{-1}$ for pProNde and pET-TEV vectors, respectively.

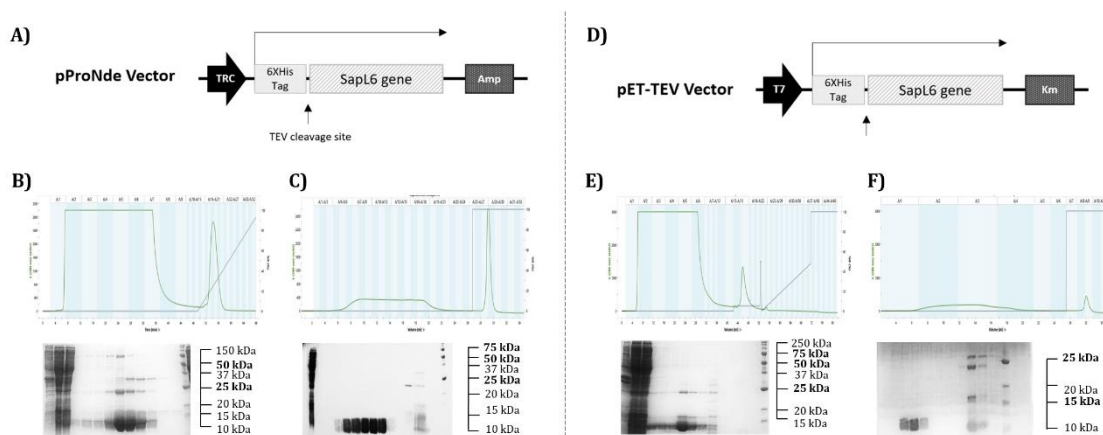


Figure 4.7. SapL6 expression and purification. A and D) Schematic representation of the genetic constructs used for the expression of SapL6 in pProNde and pET-TEV vectors. **B and E)** Representative chromatograms of SapL6 purification before TEV cleavage with their respective profiles on 15% SDS-polyacrylamide gels (insoluble fraction, soluble fraction, flow through, washing, elution and molecular weight marker, from left to right respectively). Green lines represent Absorbance and dark lines represent the elution gradient. **C and F)** Representative chromatograms of SapL6 purifications after TEV cleavage, with their respective profiles on 15% SDS-polyacrylamide gels.

4.4.3 Molecular size and oligomeric stage of SapL6

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. Originally, this was considered a crystallization artifact, however, nowadays it has been demonstrated that CV-N can exist both in monomeric and in domain-swapped dimeric form also in solution [94]. Conversely, no exchange or dimerization has been observed in the CVNH characterized so far.

The molecular size and oligomerization state of SapL6 were evaluated through SEC and DLS analyses. According with the data obtained from those experiments, SapL6 is a monomeric protein of $\sim 10.5\text{-}12.8 \text{ kDa}$. No evidence of dimerization was found in any case (Figure 4.8).

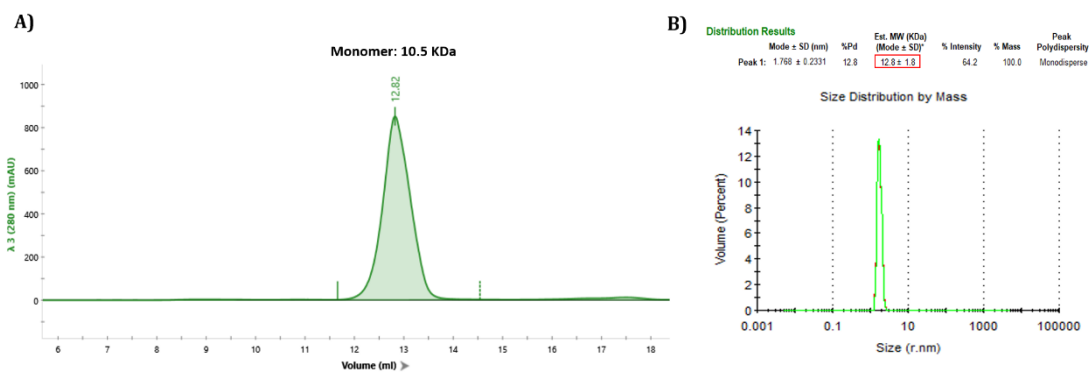


Figure 4.8. SapL6 oligomerization stage. A) Size exclusion chromatogram of SapL6. B) Size Distribution by Mass of SapL6 obtained by dynamic light scattering.

4.4.4 Thermal stability of SapL6

The stability of the protein was assayed in 26 different buffers in a pH range from 5 to 10 by thermal shift assay (TSA). No significant differences were found between the tested conditions, which indicates that the pH or buffer type don't have a big influence on the stability of SapL6 (at least not on the gradient studied). Addressing the small differences in detail, it can be generalized that the most stable condition ranged between pH 6.5 and 9 with a T_m between 45-48 °C where the denaturation occurred in a single event. On the other hand, multiple T_m were predominantly observed at low pH, but it is pertinent to highlight that the additional signals were originated from small peaks that could have been generated from automatic baseline adjustment (Figure 4.9). When stored in MES buffer pH=6 or PBS pH=7.4, SapL6 is stable up to 12 months at 4°C and no evidence of degradation or misfolding were identified even after 24 months storage at -20 °C. The protein is therefore quite stable despite a low T_m .

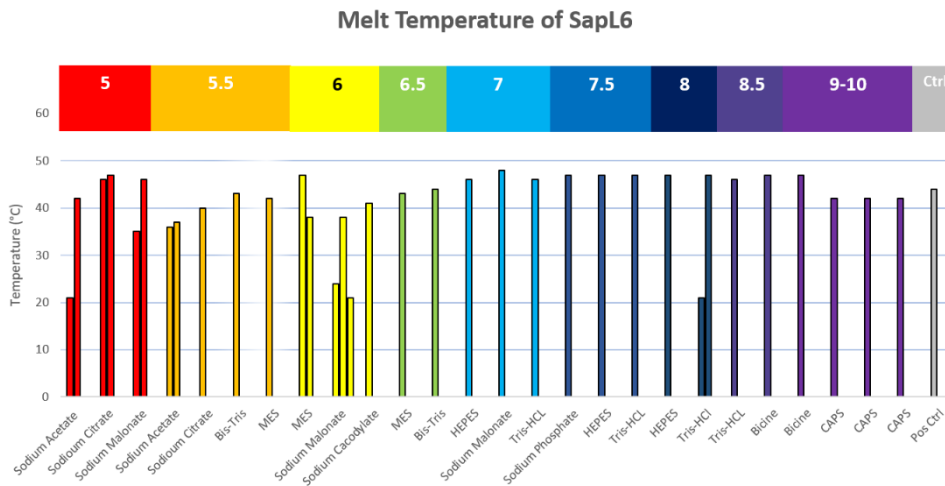


Figure 4.9. SapL6 thermal Stability. Melt temperatures of SapL6 obtained through the Thermal Shift Assay (TSA). A temperature gradient from 20 to 100°C in was applied under 26 different buffer conditions with a pH range from 5 to 10.

4.4.5 Hemagglutination activity of SapL6

Since it is well known that many lectins have the capacity to agglutinate blood cells, we have performed an hemagglutination assay on rabbit erythrocytes. However, no hemagglutination activity was detected even at high protein concentrations (Figure 4.10). This result is consistent with what has been previously reported for AFL6 and other homologous proteins from the CNV-homologs in fungi. The absence of hemagglutination can be explained by the fact that SapL6 as other fungal CVNHs would be monovalent with only one the two potential binding sites functional. However, the possibility that the determinants recognized by this lectin are absent in the analyzed cells, cannot be discarded.

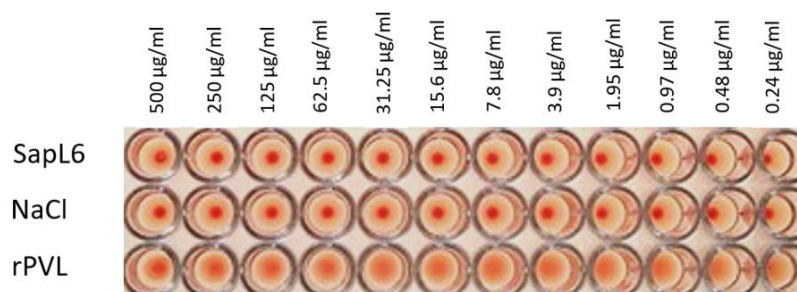


Figure 4.10. Hemagglutination activity of SapL6. assay performed on fresh rabbit erythrocytes. Negative and positive controls consist of 150 mM NaCl and the rPVL lectin from *P. velutina*, respectively.

4.4.6 SapL6 carbohydrate binding properties ITC

The affinity of SapL6 for Man α 1-2Man α was addressed by ITC, however, it was not possible to identify a clear pattern of interaction. Since it is known that TbCVNH also can bind glycans composed of glucose, galactose and N-acetylgalactosamine, we also titrated SapL6 using those sugars and a large set of other glycans available in our laboratory, but no binding was detected. In some cases, titrations with high concentrations of ligands (*i.e.*, mannose, N-acetylgalactosamine, sucrose, etc.) display patterns that resemble a very weak interaction, however, this was not confirmed when the molar ratio protein: ligand was increased. Therefore, this phenomenon was attributed to the energy realized by the sugar dilution in the sample cell. A table summarizing ITC experiments is presented in Table 1.

Table 4.1. ITC experimental parameters (SapL6)

Protein concentration (mM)	Ligand	Ligand concentration (mM)	Molar ratio
0.1	Man α 1-2Man	50	100
0.05		100	400
0.5	D-Mannose	250	100
0.2		100	100
0.06		150	500
0.06		30	100
0.06	Methyl α -D-mannopyranoside	30	100
		150	500
	Glucose	30	100
		150	500
	GlcNac	30	100
		150	500
	Galactose	30	100
		150	500
	Lactose	30	100
		150	500
	L-Rhamnose	30	100
		150	500
	Sucralose	30	100
		150	500
Lactulose	30	100	
	150	500	
L-Xylose	30	100	
D-Xylose	30	100	
Arabinose	30	100	

	Mannolactone	30	100
	D-Gluconolactone	30	100
0.1	Sucrose	50	100
0.06		30	100
0.06		150	500

4.4.7 SapL6 carbohydrate binding properties; glycan array

To expand our search, we used the mammalian glycan array version 5.4 of the Consortium for Functional Glycomics (USA) consisting of 585 mammalian glycans. For this, SapL6 was labeled with the FITC fluorophore. However, the low proportion of primary amines in its sequence (1%, 1 residue) entailed low labeling (molar ratio ~ 0.04) and weak fluorescence signals with elevated standard deviations were obtained from the glycan array screening (Figure 4.11).

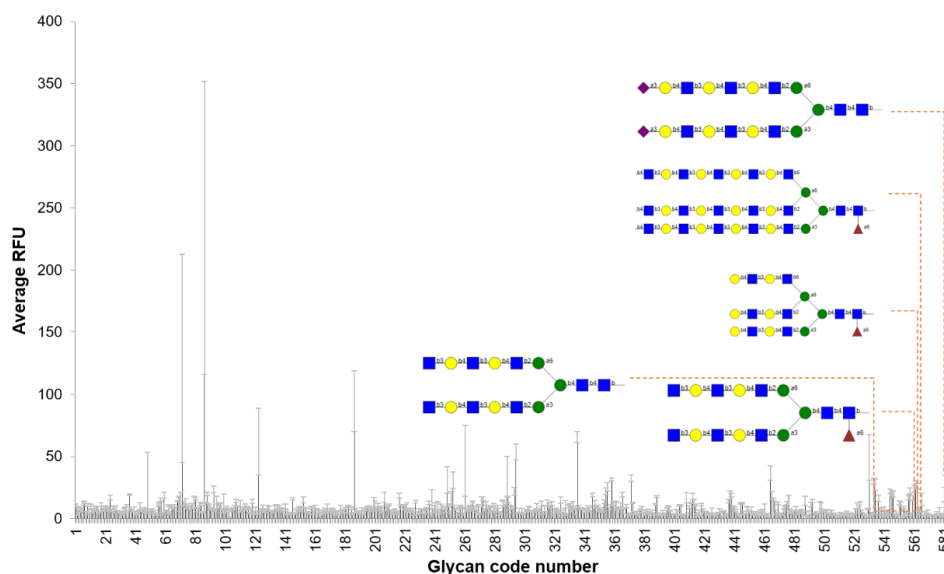


Figure 4.11. SapL6 interaction with the mammalian glycan array of CFG. Relative Fluorescent Units (RFU) plot of the glycan array matrix with SapL6 at $50 \mu\text{g}\cdot\text{mL}^{-1}$. The experiment was carried out in duplicate at 50 and $5 \mu\text{g}\cdot\text{mL}^{-1}$ of protein. Only the plot obtained at $50 \mu\text{g}\cdot\text{mL}^{-1}$ is included in the figure. The structures of the 5 best binders averaging both experiments are shown linked to their respective signals. Raw data can be consulted at <https://docs.google.com/spreadsheets/d/1U8oGfSXKJ90RqfzHz5gISeIS4zvQO2dL/edit?usp=sharing&oid=111416367677195406970&rtpof=true&sd=true>

Although the signals are relatively poor, it is interesting and highly noteworthy the fact that these results, resemble the previous data obtained in 2015 for AFL6. where it was possible to detect only some low interaction with the polylectosamine containing glycans. It also agrees with the data reported by Koharudin *et al* 2008, where TbCVN was bound to mannose-lacking glycans composed exclusively by glucose or a combination of glucose, galactose, and N-acetylglucosamine [59]. This leads us to raise three possible scenarios:

- i) SapL6 and AFL6 are involved in host recognition process but their target was not present in the glycan array.
- ii) SapL6 and AFL6 are not involved in host recognition and therefore no interaction with mammal's glycans should be expected.
- iii) SapL6 and AFL6 are host binding lectins that should display affinity for the carbohydrates of the glycan arrays respectively tested, but they are not active.

Assuming the first and/or second scenarios are correct, it was determined to expand the search with other chips that contain glycans from different sources, not just mammals. However, due to the previously observed inconveniences during the labeling of SapL6, on this occasion, it was decided to proceed with its detection by using anti-his antibodies (anti-His Ab) fluorescent-labeled. For this, a new batch of SapL6 was produced; on this occasion, the histidine tail was not cleaved the elution parameters were modified during the first affinity chromatography to improve the sample purity (Figure 4.12). Then, the sample was additionally purified by size exclusion chromatography and were tested on two new glycan matrix platforms; 1) the microbe-based glycan array matrix [104], developed by the groups of Professor Peter H. Seeberger, at Max Planck Institute (Germany and 2) the broad-spectrum microbial glycan array of the Imperial College of London (UK). According to our contacts in these platforms, it was not

possible to identify any interaction during the screening, although the formal reports are still pending.

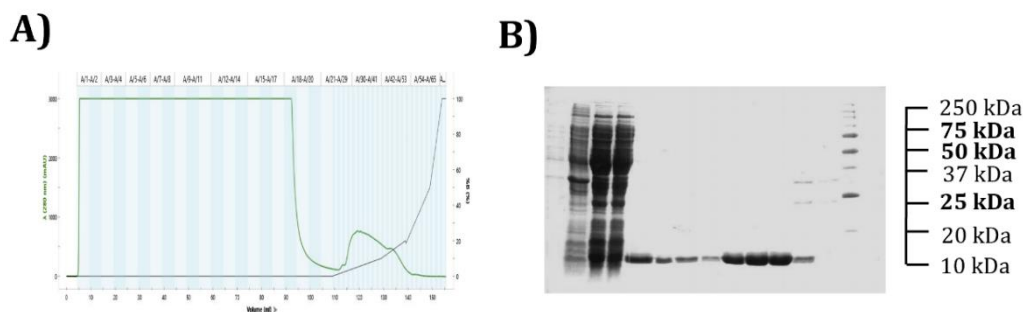


Figure 4.12. optimization of SapL6 purification. The figure shows a representative chromatograms of affinity chromatography purification of SapL6 after the optimization of elution parameters. Green lines represent Absorbance and dark lines represent the elution gradient. The electrophoretic profile of the purification 15% SDS-polyacrylamide gels is also shown (insoluble fraction, soluble fraction, flow through, washing, elution and molecular weight marker, from left to right respectively).

4.4.8 SapL6 crystallization

After thousands of tested conditions, small SapL6 crystals were obtained in a solution containing 3 M $(\text{NH}_4)_2\text{SO}_4$ and 100 mM HEPES buffer pH = 7. Crystals diffracted at 2 Å resolution and the protein structure was solved by molecular replacement. However, we encountered problems during refinement leading to poor statistics and we decided to optimize nucleation. For this, we studied the influence of small variations of each of the components of the, already simple, mother liquor. Through this screening, we identify that SapL6 crystallizes in a very narrow window of conditions, mainly influenced by the concentration of ammonium sulfate.

From all conditions tested, large and well-defined symmetric crystals were obtained only with 2.8 M ammonium sulfate. Interestingly, even when the rest of the components remained identical, the protein precipitated at slightly higher concentrations (≥ 2.9 M) and the droplets remained clear at ≤ 2.7 M (Figure 4.13). A great influence was also observed for buffers systems (both, in the protein sample and mother liquor). On the other hand, protein concentration does not

seem to be as strict a parameter such as the other components, since it was possible to obtain crystals in protein concentrations ranging from 7 to 10 mg mL⁻¹ with high reproducibility.

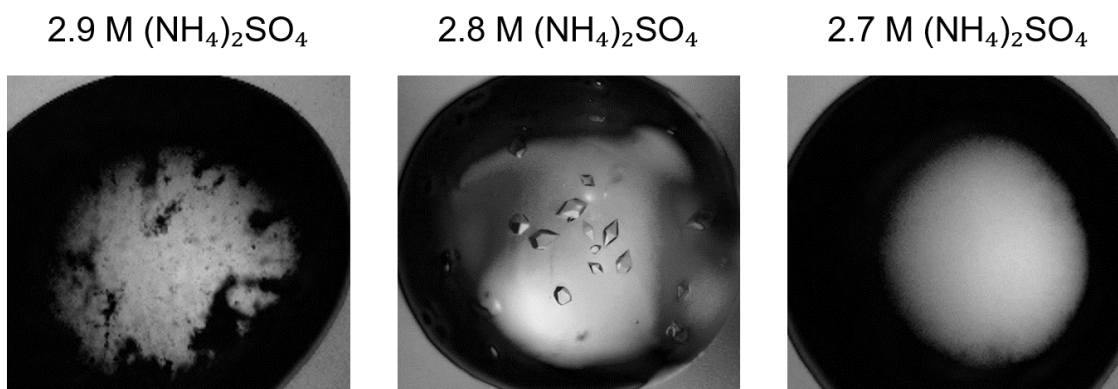


Figure 4.13. SapL6 crystallization. Three pictures of representative conditions tested for SapL6 crystallization are shown in the figure. Drops were made by mixing equal volumes of SapL6 (9.5 mg mL⁻¹ in PBS buffer, pH=7.4) and mother liquor (100 mM HEPES buffer pH=7, with the respective concentration of ammonium sulphate).

The new crystals were obtained for the apo protein or during cocrystallization assay with different sugars. They diffracted up to 1.35 Å, several high-quality datasets were obtained, the best one was chosen but its refinement is still ongoing. Hence, the original structure solved at 2 Å is reported in this draft of the thesis.

4.4.9 SapL6 structure

SapL6 structure was solved by molecular replacement at 2 Å with AFL6 structure as searching model. The crystal space group was found to be P4₁2₁2 with one molecule per asymmetric unit and 30.69% solvent (Table 1). The overall structure of the model closely resembles the CV-N fold; where the two tandem repeats (TR1 and TR2) have an identical topological organization and form two domains interconnected by chain exchange (Figure 4.14). In SapL6, domain A is composed of residues 1-37 and 95-106, while domain B is positioned between amino acids 38-94. Each domain is composed of two helical turns and 5 β-strands divided by a 2-fold symmetry axis.

4.4.10 Comparison of SapL6 with other CVNHs

The structural comparison of SapL6 and the most representative members of the CNVH family shows that although the sequences of these proteins display <50% of identity, their 3D-fold is extremely conserved (Figure 4.15 A). SapL6 is an elongated protein presenting largely beta-sheets and displays internal two-fold pseudosymmetry.

Consistently with our previous data, the topological distribution of SapL6_TR1 reassembles TbCVNH_TR1 better than NcCVNH_TR1, mainly due to the absence of the first helical turn of NcCVNH_TR1 (Figure 4.15 C and E, rmsd 1.55 and 1.43 respectively). This is particularly relevant since the inactivity of the BP1 of NcCVNH_TR1 has been attributed to the substitution of this $\alpha 1$ by a flexible tail, resulting in the loss of the concave pocket required for sugar binding [59]. Hence, the functionality of this binding site in AFL6 and SapL6 seems plausible.

Surprisingly, in addition to AFL6 (rmsd 0.522), the CV-N structure is the one with the most similarities to SapL6 in both domains (rmsd 1.26). Except for disulfide bridges, the most important difference is the presence of 5 additional residues in SapL6_TR2 (55-DSVPE-66). This small sequence leads to the elongation of the loop connecting the two strands $\beta 6$ and $\beta 7$ as well as the strands themselves in SapL6 compared to all other CNVHs and CV-N (Figure 4.15 B-F).

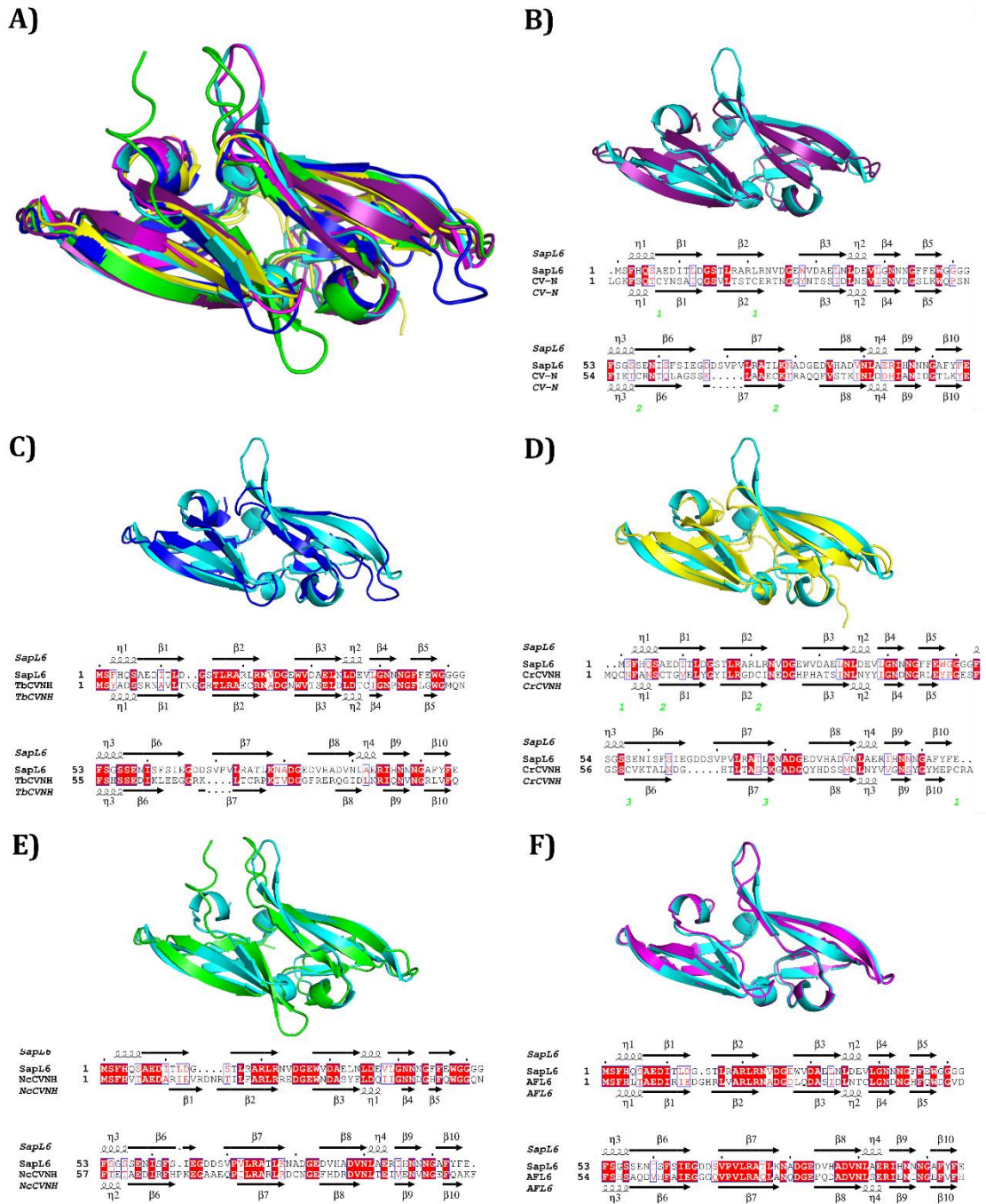


Figure 4.15. Comparison of SapL6 with other CVNHs. A) Structural alignment of SapL6, AFL6, CV-N and the most representative members of the CVNH family (TbCVNH, CrCVNH and NcCVNH). B) Structural comparison of SapL6 and CV-N (PDB: 1IIY). Green numbers indicate bondable cysteine pairs C) Structural comparison of SapL6 and TbCVNH (PDB: 2JZK). D) Structural comparison of SapL6 and CrCVNH (PDB: 2JZJ). Green numbers indicate bondable cysteine pairs E) Structural comparison of SapL6 and NcCVNH (PDB: 2JZL). F) Structural comparison of SapL6 and AFL6 [94]. Structural alignments and protein models were visualized and edited with PyMol Version 1.8.4. Sequence alignments were performed using Clustal W [100] and visualized with ESPrnt3 [101].

4.4.11 Structural insights of putative carbohydrate-binding pockets in SapL6

SapL6 has two shallow grooves at both ends of the ellipse that perfectly overlap the carbohydrate-binding pockets of CV-N. Although the topology of these regions is very similar, the residues directly involved in the binding of Man α 1-2Man in CV-N greatly differ from those found in SapL6. In the cyanobacteria members, the binding pocket of the domain A (BP1), the disaccharide interaction is mediated by hydrogen bonds with two aspartic acids that are replaced by arginine and asparagine residues, respectively (D95 / N99 and D23 / R22) in SapL6 (Figure 4.16 A). While in the binding pocket located in domain B, residues N53 and Q78 are replaced by glycines, only asparagine (N42) remains conserved. In fact, this residue is conserved in all members of the CV-N family (Figure 2.16 B). Interestingly, while the composition of the binding pockets of SapL6 differs greatly from that of CV-N, these residues are identical among all the fungal members of this family showing divergence of the fungal members from the bacterial one.

In 2009, a chimeric protein was synthetically engineered by fusing the A domain of TbCVNH with the B domain of NcCVNH. This chimera maintained its mannose-binding properties but also exhibited sucrose-binding activity and its structure was solved with this sugar attached to both pockets [105]. We have evaluated the similarities between the residues directly involved in such interactions and we did not find significant differences compared to previous comparison with CV-N. In domain B, the two asparagine residues remain conserved while Arg81 and Asp45 are substituted by very similar residues leading to the conservation of the hydrogen binding network (Figure 4.17 A). In domain A, the only observed change consists in the substitution of glutamine Q98 by histidine H96. Surprisingly, a greater variability is observed among the residues involved in the interaction with sucrose among the fungal members of the family (Figure 4.17 B).

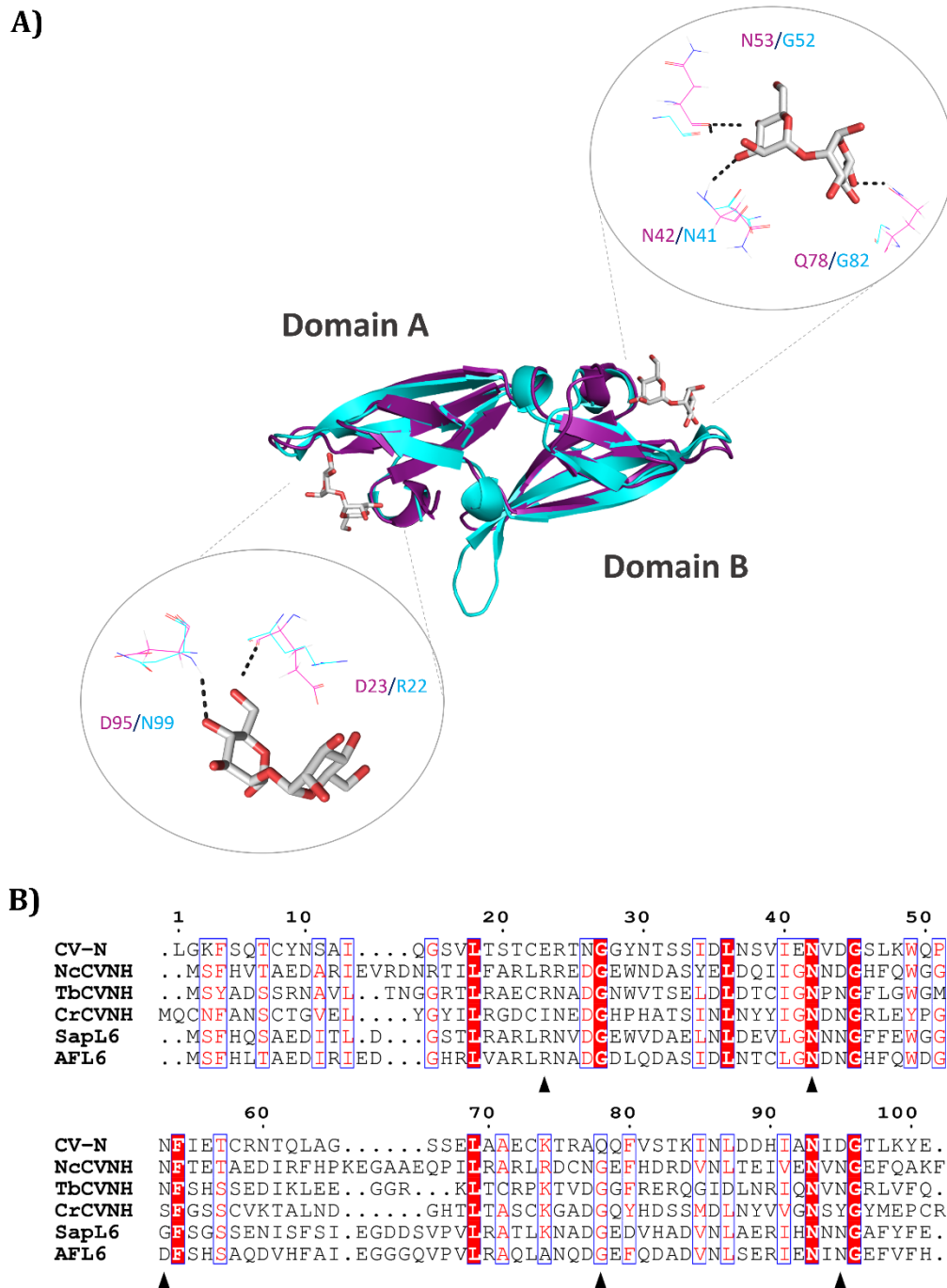


Figure 4.16. Structural comparison of CV-N and SapL6 carbohydrate binding pockets. A) Alignment of CV-N (purple) and SapL6 (cyan) structures. **B)** Sequence alignment of SapL6 and the most representative members of the CVNH family. Black triangles indicate the residues responsible for hydrogen bonding with Man α 1-2Man in CV-N (PDB: 1IIY). Sequences alignment was performed using Clustal W [100] and visualized with ESPrIPT3 [101]. Protein models were visualized and edited with PyMol Version 1.8.4.

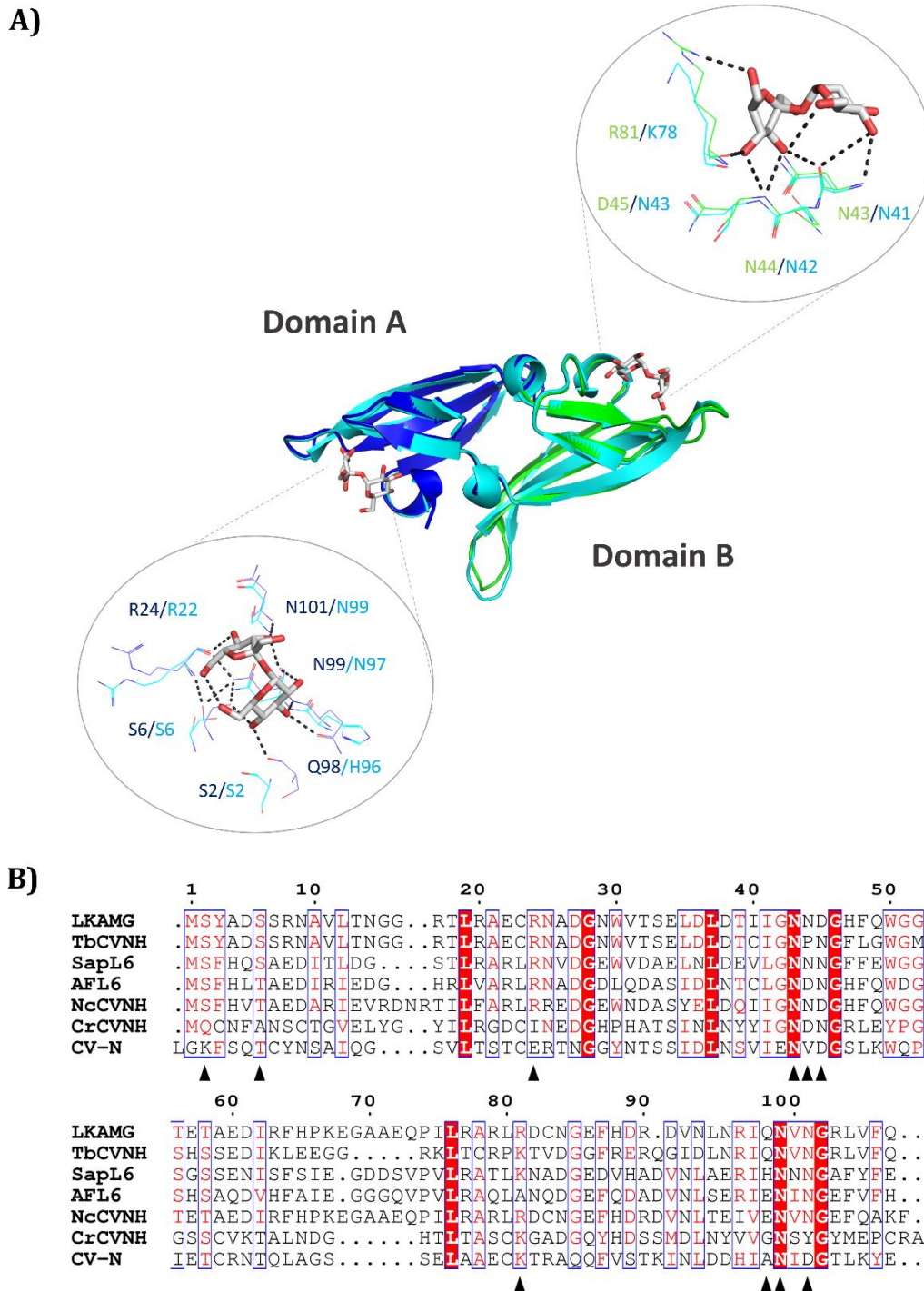


Figure 4.17. Structural comparison of the chimera-LKAMG and SapL6 carbohydrate binding pockets. A) Alignment of the structures of SapL6 (cyan) and the chimera-LKAMG (domain A in blue, domain B in green, PDB: 3HP8). B) Sequence alignment of SapL6 and the most representative members of the CVNH family. Black triangles indicate the residues responsible for hydrogen bonding with Man α 1-2Man in CV-N (PDB: 1IIY). Sequences alignment was performed using Clustal W [100] and visualized with ESPript3 [101]. Protein models were visualized and edited with PyMol Version 1.8.4.

4.5 Conclusions

In this study, we have identified and recombinantly produced a new cyanovirin-like lectin from the human opportunistic microfungi *S. apiospermum*. Although SapL6 was successfully produced and characterized, it has not been possible to identify any carbohydrate-binding activity. Due to its low affinity for mannosylated glycans and monovalency, SapL6 is not expected to present CV-N activity against viruses such as HIV. However, its applications as a pharmacological target are still valid.

Cyanovirin-like lectins are a great enigma of nature. Although it is clear that its biological role must be very important for living, the functions of the already identified members of this family remain unclear and their characterization has only aroused questions such as:

- ❖ Which evolutive force leads to the preservation of such structure among so divergent organisms even when the sequences are not conserved?
- ❖ What is their natural ligand?
- ❖ Do all these proteins share the same function in their organisms of origin?
- ❖ Which function is this?
- ❖ Why are they only present in cyanobacteria, fungi and ferns?
- ❖ What is the process of evolutionary divergence that led to the transmission of CV-N genes only in these species and the differences between bacterial and fungal members?

We believe that the identification and characterization of new CVNHs could provide important information to contribute to the general understanding of such proteins and answer these questions. For this reason, despite we were aware of the enormous challenge that this objective entails, we have approached the study of SapL6, not entirely unsuccessfully. We believe that the structural information gathered here will be useful to address possible interactions with

glycans through *in silico* analysis and might guide towards the discovery of its natural ligand. Which, we consider, is the ultimate clue to finally understanding the biological role of this family of proteins and to evaluate their potential as drug targets.

For our part, we continue to search for strategies that allow us to overcome the current obstacle in which we are stopped, and we are testing the co-crystallization of the protein with diverse glycans, but so far, only the apo structure was obtained.

4.6 Contributions

Dania Martínez Alarcón (DMA) performed cloning, protein expression, protein purification, protein labeling, hemagglutination, isothermal microcalorimetry measurements. Glycan array experiments were requested as service to the Consortium of Functional Glycomics (CFG) and the analysis of data was performed by DMA. Protein crystallization and structure determination of SapL6 were performed by DMA and Annabelle Varrot (AV). DMA prepared all the figures and discussed all results. A.V. and Roland J. Pieters (RJP) administered the project, conceived the design of the study, obtained funding, contributed to data analysis and supervised DMA.

4.7 Supplementary information

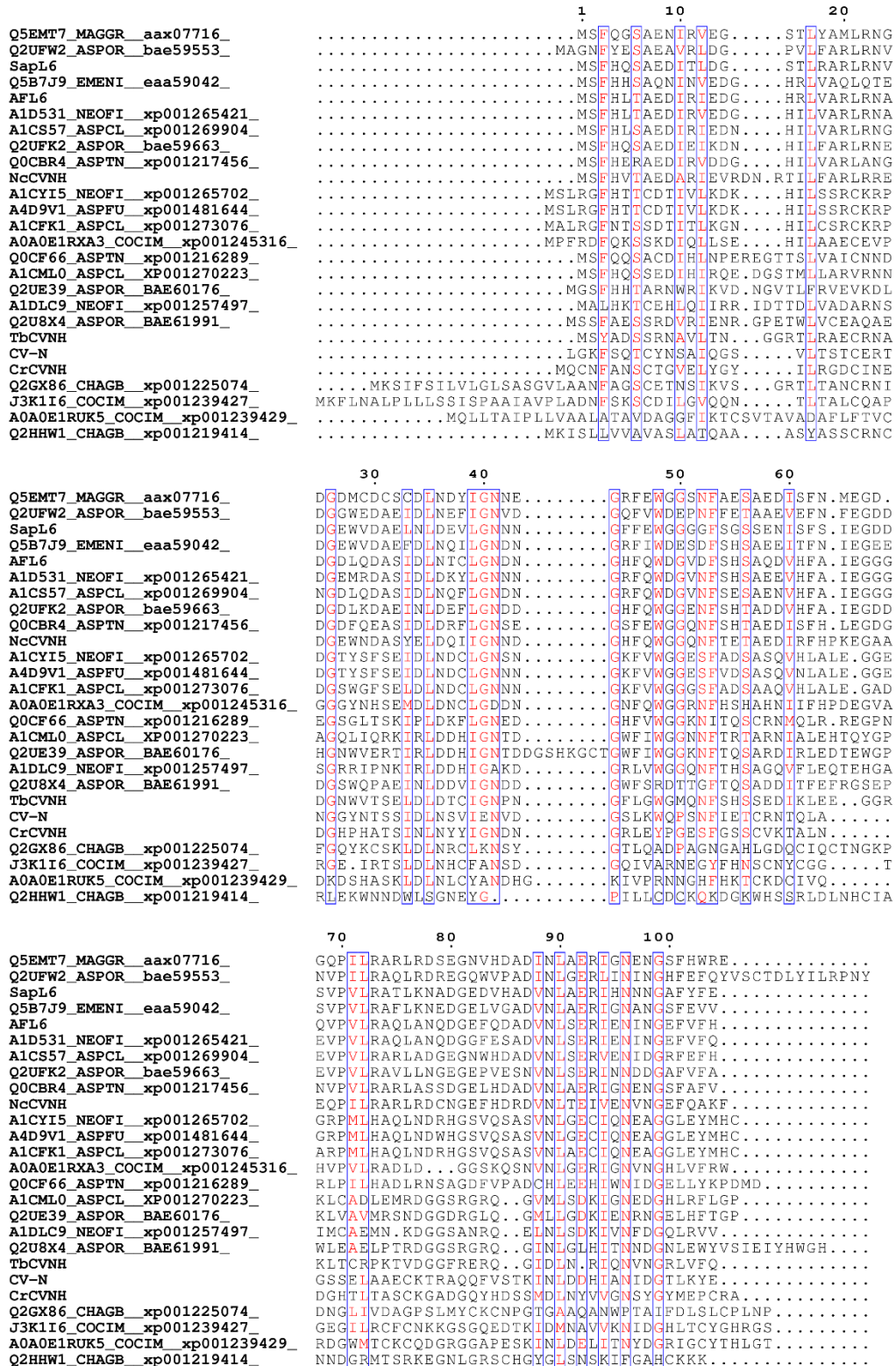


Figure 4.18: Multiple alignment of sequences of the CVNH used to build the family phylogeny. Sequence's alignment was performed using Clustal W [100] and visualized with ESPrnt3 [101].

5. IDENTIFICATION OF NEW NATIVE LECTINS FROM PROTEIN EXTRACTS OF MEDICAL RELEVANT FUNGAL STRAINS

5.1 Summary

Fungi are an invaluable source of novel lectins that are key elements for the development of glycobiology and biomedical research, either as tools or as drug targets. So, there is great interest in their identification and characterization. Nevertheless, fungal lectins have been poorly studied and their reports are scarce compared to those from other organisms. This represents a problem in the case of life-threatening species for humans, such as *Aspergillus fumigatus* and *Scedosporium apiospermum*. Therefore, the main aim of this part of the project was to identify new wild-type lectins from those pathogens. For this, we have analyzed their intracellular proteins and secretomes through sugar-affinity chromatography, SDS-page and mass spectrometry. A total of 54 bands (presumably corresponding to carbohydrate-binding proteins) were recovered from gels and 22 of them have been sequenced to date. The analysis of the peptides obtained by mass spectrometry, allowed us to identify 7 hypothetical lectins from those samples, four of them from *S. apiospermum* (PVL-like lectin, SapCP, CcMBL-like lectin and SapFKBP12). The preliminary analysis of those proteins revealed that some might be attractive pharmacological targets and others could find application as biotechnological tools. In this chapter, we present a detailed analysis of their sequences, structure and putative role in the fungal life, as well as their phylogenetic distribution and potential applications.

5.2 Introduction

Due to their biological relevance and possible applications, fungal lectins have arisen the attention of the scientific community, either to explore their implementation as tools or to propose them as drug targets. This interest, has led to a growing rate of new fungal lectins discovery in the recent years, however, there is still a considerable gap compared to lectins from other organisms, such as plants, animals or bacteria. The lack of information is particularly alarming in the case of microfungi, whose represent only 15% of the total lectins identified from the complete kingdom, while the remaining 85% come from mushrooms and yeasts [52,56,57].

S. apiospermum and *A. fumigatus* are opportunistic microfungal pathogens that affect mainly immunocompromised patients. The rate of infections caused by those microorganisms has increased in the recent decades [4] and nowadays they are most common filamentous fungus to colonize the lungs of patients with cystic fibrosis [12,13,40]. Furthermore, they have a high rate of recidivism and display resistance towards several antifungals [5,106]. Since lectins are involved in several mechanisms elemental for the fungal development and host-adhesion process, there is an emerging interest in the identification and characterization of lectins from those two microfungal pathogens. Nevertheless, their identification reports are limited to date, particularly in the case of *S. apiospermum*.

There are two main strategies for new lectins identification; i) the *in silico* approach, which use data mining to predict putative proteins encoded on the DNA of species whose genome has been sequenced and, ii) the traditional approach, which takes advantage of the carbohydrate-binding properties of those proteins and uses affinity chromatography for their purification. Each one of those strategies have their advantages/disadvantages.

On one hand, the *in silico* prediction requires *a priori* knowledge of similar lectins from other species and the identification must be followed by the recombinant expression of the predicted proteins and their subsequent validation. This implies a slow process that limits the study of a large number of proteins simultaneously. In addition, this strategy is only available for organisms whose genomes have been sequenced and incomplete or partially assembled genomes could drive to false negatives during the identification. Finally, it is important to consider that genomic sequences can vary according to the database or even between strains of the same species.

On the other hand, the major disadvantages of the traditional approach for the identification of new lectins are that some of these proteins are expressed differentially during different stages of fungal development or are produced only under certain conditions, such as abiotic stress, colonization, etc. Furthermore, some of those lectins are present in very low quantities, hindering their detection. Even if the experimental approach manages to evade those inconveniences, it is important to consider that some lectins need a specific arrangement of carbohydrate linkages to achieve binding and they would not be retained by conventional carbohydrate-based matrixes usually constituted of monosaccharides. Additionally, when using peptide-based mass spectrometry to sequence the identified proteins, sequence databases are also needed. This brings back the problems associated with incomplete or partially assembled genomes and sequence variations between databases or strains.

The *in silico* approach has been used during this work to predict SapL1 and SapL6 lectins from *S. apiospermum*. Whose production and characterization are described in chapters III and IV. Now, the main aim at this stage of the project was to identify native lectins of medical relevance from mycelium cultures of *A. fumigatus* and *S. apiospermum*. For this, we have performed an extraction of total

proteins from mycelia of both microfungi and we also have analyzed the fraction of proteins secreted to the culture media. All fractions were purified by carbohydrate-affinity chromatography and proteins recovered from SDS-PAGE gels were proteolyzed and the peptide identified by mass spectrometry. A general overview of the project is shown in figure 5.1.

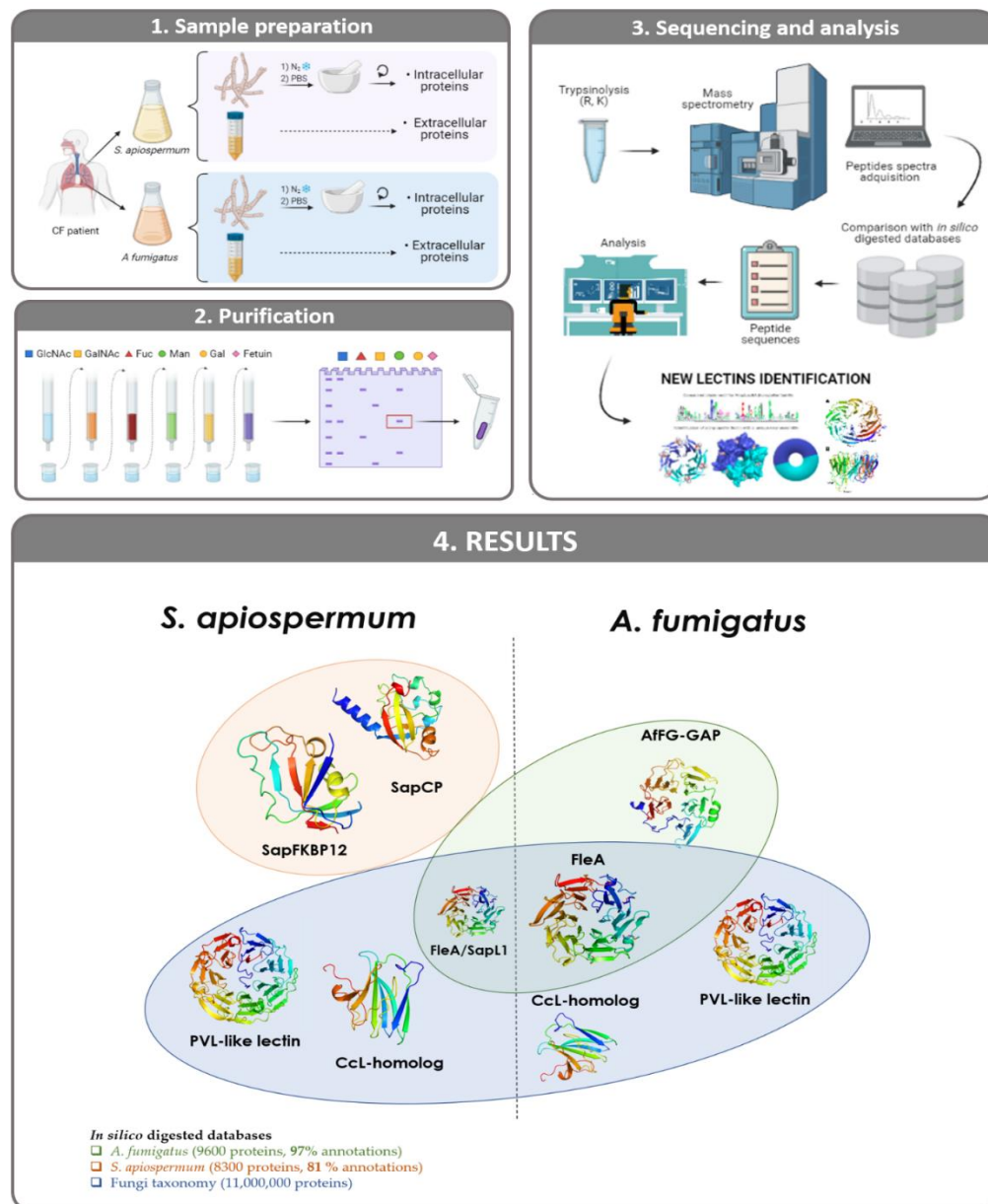


Figure 5.1: Graphical abstract of project 3. 1) protein extraction and sample preparation. 2) schematic representation of the purification train assembled for native lectins purification. N-acetylglucosamina, N-acetylgalactosamina, fucose, mannose, glucose and fetuin columns are shown from right to left respectively. 3) Schematic representation of data collection, analysis and identification of new lectins. 4) Graphical summary of identified proteins. Figure generated using Biorender.com and Microsoft Office 365.

5.3 Material and methods

5.3.1 Fungal strains

The identification of native lectins from medically relevant fungal strains was carried out from protein extracts of *S. apiospermum* IHEM 14462 and *A. fumigatus* IHEM 6951, kindly provided by Prof. Muriel Cornet, TIMC, CHU Grenoble.

5.3.2 Sample preparation

Three weeks after inoculation, culture media was removed by filtering with 0.45 μm membranes. After filtration, 50 mL of each culture media were collected for further analysis and the biomass was frozen in liquid nitrogen, grinded and suspended in Phosphate buffer saline (PBS) buffer pH 7.4. Finally, samples were centrifuged at 24000 g for 30 minutes to remove cellular debris and soluble fraction was filtered again with 0.45 μm filters.

5.3.3 Lectins purification

Six different matrixes for carbohydrate affinity chromatography were assembled in a consecutive purification train. Then, columns were equilibrated with 10 column-volumes (CV) of PBS buffer pH 7.4 and each sample was sequentially eluted through all of them as described in the figure 5.1.2. Immediately after the sample was eluted, resins were washed with 10 CV of PBS buffer pH 7.4 supplemented with 1 M of NaCl to disrupt unspecific interactions. Finally, bounded proteins were recovered from each column by eluting 4 CV of PBS buffer pH 7.4 supplemented with 250 mM of the corresponding sugar.

5.3.4 SDS-PAGE

Fractions eluted from each column were concentrated to 200 μl via ultrafiltration with Vivaspin® filters of 10 kDa cutoff (Sartorius). Then, samples were boiled 10 minutes in presence of LDS sample buffer 4X (Expedeom Ltd), centrifuged and ran in a 4–20% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad Ltd) at 200 mV during 40 minutes. Gels were stained with InstantBlue Coomassie

staining solution (Abcam) overnight. In the following day, the visible bands were manually sectioned and stored in 10% EtOH at 4°C for further analysis.

5.3.5 Proteins sequencing

Samples were sent to EDyP-service platform, CEA Grenoble, where mass spectrometry analysis were performed. For this, samples were digested using trypsin, which is an endopeptidase cutting after arginines and lysines. Then, the peptides generated from the digestion of each sample were separated and their MS spectrum experimentally obtained were compared to a database of proteins digested *in silico*. Three different databases were used for the analysis of the peptides: *N. Fumigata* database, which contains 9600 proteins (96.9% annotation); *S. apiospermum* database, which contains 8300 proteins (81,1% annotation) and an extended database containing all proteins of the Fungi taxonomy (11,000,000 proteins).

5.4 Results and discussion

5.4.1 Purification of native lectins

3-week-old fungal mycelia of *A. fumigatus* and *S. apiospermum* were used for the analysis. It is important to highlight that, even when both cultures were grown under identical conditions, they presented considerable morphological differences at harvesting time. *Scedosporium* mycelia showed highly organized growth and did form robust spherical granules (hyphal pellets). The coloration of this culture was sparkly-deep brown and the weight of wet biomass recovered by filtration was 25 gr per liter of culture. While, *Aspergillus* culture was a turbid homogeneous solution with an opaque-light brown color and the mass of the wet pellet was only 10 gr per liter of culture. The carbohydrate binding proteins of each sample were purified and concentrated as described in materials and methods. Their electrophoretic profile is shown in figures 5.2 and 5.3, where it is possible to observe at least four clear bands in the intracellular fractions and

multiple bands of lower concentrations in the extracellular fractions of both fungi.

The molecular size of proteins identified from the intracellular fractions are slightly different between *S. apiospermum* and *A. fumigatus*, however they all follow the same pattern. On the other hand, it is clear that the size patterns of proteins secreted to the media are pretty different between the two fungal strains (Figures 5.2 and 5.3).

In a second attempt to identify very diluted proteins, samples were concentrated a 2nd time through ultrafiltration until achieve the half of their original volume (Figure 5.4). This strategy allowed us to identify 23 additional proteins but five of the original bands disappeared possibly due to degradation. The missing bands in the second experiment are indicated in the figures 5.2 and 5.3 as green labels with the range numeration from 56-59. A total of 54 proteins bands with presumptive carbohydrate-binding activity were obtained from this second attempt of identification.

All the bands were recovered from the gels and stored in EtOH 10% and the 22 stronger bands were selected for mass spectrometry analysis while the rest remained at 4°C until further analysis.

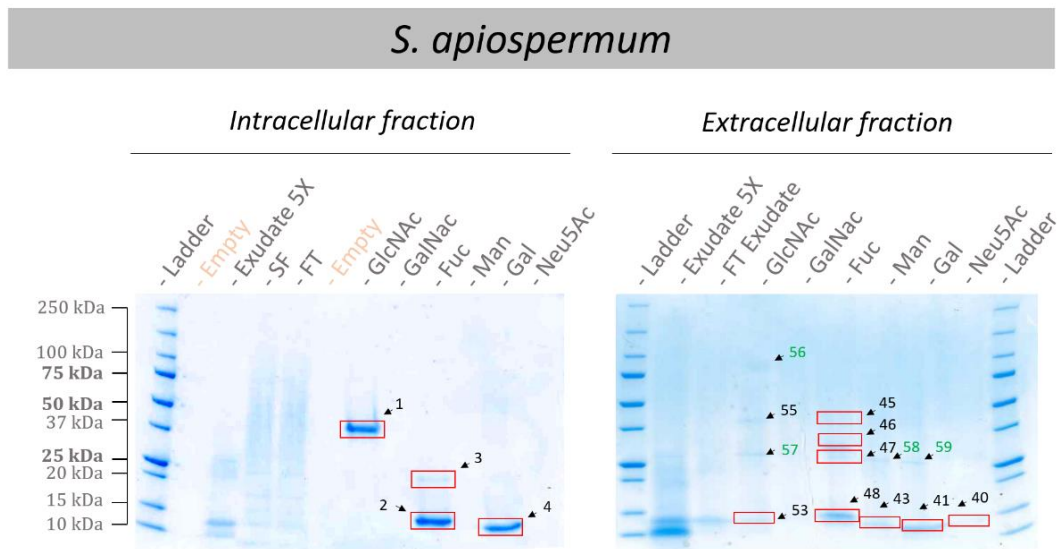


Figure 5.2: Native lectins identified from *S. apiospermum* mycelium cultures. Left) SDS-PAGE purification of intracellular proteins. Lane 1: ladder; lane 3: 20 μ L of culture media (concentrated 5X); Lane 4: 5 μ L of total protein extract (soluble fraction); Lane 5: flow-through; Lines 7-12, 20 μ L of samples obtained after elution from GlcNAc, GalNAc, fucose, mannose, galactose and fetuin resins, from left to right respectively. Right) SDS-PAGE of purification of proteins secreted to the media. Samples as disposed as previously described for left panel.

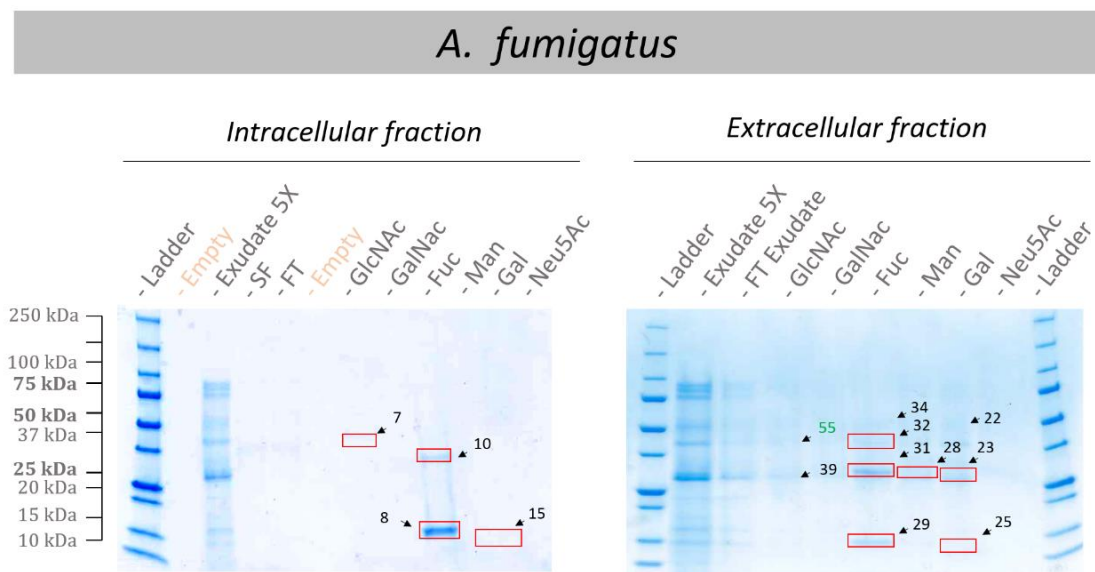


Figure 5.3: Native lectins identified from *A. fumigatus* mycelium cultures. Left) SDS-PAGE purification of intracellular proteins. Lane 1: ladder; lane 3: 20 μ L of culture media (concentrated 5X); Lane 4: 5 μ L of total protein extract (soluble fraction); Lane 5: flow-through; Lines 7-12, 20 μ L of samples obtained after elution from GlcNAc, GalNAc, fucose, mannose, galactose and fetuin resins, from left to right respectively. Right) SDS-PAGE of purification of proteins secreted to the media. Samples as disposed as previously described for left panel.

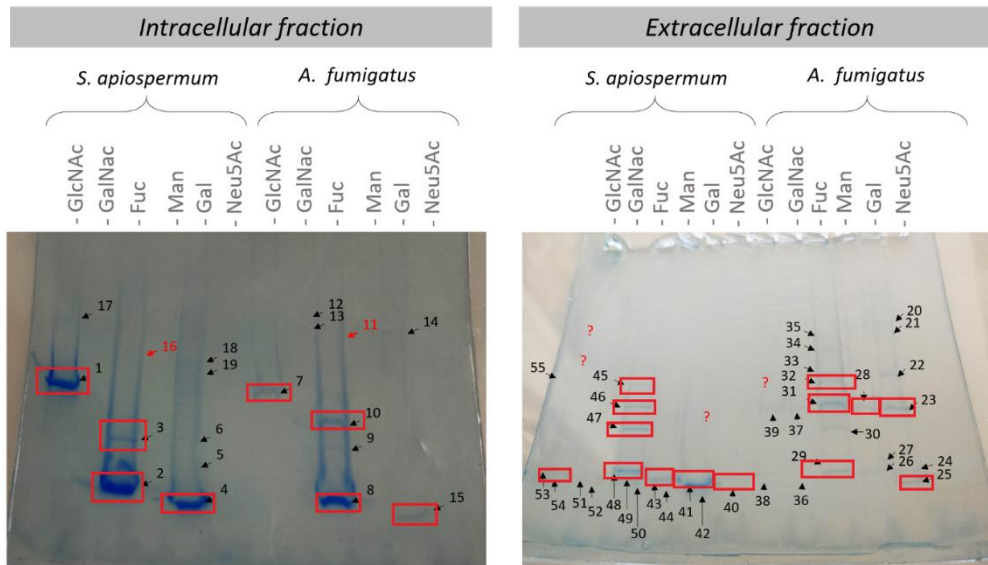


Figure 5.4: Native lectins identified from fungal cultures 20X concentrated. Left) SDS-PAGE purification of proteins from the intracellular fraction of both mycelia. Lines 1-6: 20 μ L of samples obtained from the purification of total protein extract from *Aspergillus* on GlcNAc, GalNAc, Fucose, Mannose, Galactose and fetuin resins, from left to right respectively; Lines -12: 20 μ L of samples obtained from purification total protein extract from *Scedosporium*. **Right)** SDS-PAGE of purification of proteins secreted to the media by *Aspergillus* and *Scedosporium*. Samples as disposed as previously described for left panel.

5.4.2 Sequencing of native lectins

Several peptides were successfully identified from each one of the bands that were analysed. However, when they were compared with the *in silico* digested databases of *Neosartorya Fumigata* (9600 proteins, 96.9% annotated) and *S. apiospermum* (8300 proteins, 81.1% annotated), some of the bands with a lot of material (e.g., bands 1, 2 and 4) appeared to have a low signal (few identified peptides). This result was unexpected since the concentration of a protein sample can be roughly estimated using the "spectral count" corresponding to the number of mass spectra that allow the identification of a peptide in a sample. Therefore, abundant proteins will generally be identified with high spectral count and these results might imply that those proteins were not identified because their sequence was absent from the databases used. For this reason, we performed again the analysis using an extended database that contains all the proteins from

the Fungi taxonomy (11,000,000 proteins) and, through this strategy, the problematic samples were successfully identified with high spectral counts. These results can mean either that the identified proteins are a contamination of another species of fungi or that the *N. Fumigata* and *S. apiospermum* databases are not sufficiently annotated leading to the absence of those proteins in them whilst they display homology with proteins from another fungal species. This latest explanation seems to be more plausible since all the proteins identified in another species of fungi are lectins and their carbohydrate binding specificity is congruent with the one of the lectins identified as homologous. The possibility of cross contamination can, however, not be ruled out. The individual cases will be discussed in the following sections, where we present a detailed analysis the 22 protein bands analyzed, their evolutive relationship across the fungal kingdom and their putative role in the life cycle of *S. apiospermum* and *A. fumigatus*.

All details regarding the samples and their analysis are listed in Table 1. The list of peptides identified for each protein can be found in the supplementary information of the chapter.

Table 5.1. Summary of mass spectrometry hits.

Sample			Major protein identified												
Organism	ID	Sugar	Gene	Description	Cov. (%)	MW	# of pept	Spectral count	Data base	Conf. score					
INTRACELLULAR	Sa	B1	GlcNAc	pvl	Q309D1_9AGAR Lectin PVL (Fragment) OS=Lacrymaria velutina OX=71681 GN=pvl PE=1 SV=1	74.18	42420	39	393	Fungi	1				
		B2	Fuc	CC1G_10558	A8NDX2_COPC7 Uncharacterized protein OS=Coprinopsis cinerea (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2	100	13733	17	517		1				
		B3				95.28		11	208		1				
		B4	Gal	SAPIO_CDS7060	AOA084G113_PSEDA Peptidylprolyl isomerase OS=Pseudallescheria apiosperma OX=563466 GN=SAPIO_CDS7060 PE=4 SV=1	99.21	11835	13	90		2				
		B7				92.79		14	28						
	EXTRACELLULAR	Af	B7	GlcNAc	pvl	Q309D1_9AGAR Lectin PVL (Fragment) OS=Lacrymaria velutina OX=71681 GN=pvl PE=1 SV=1	57.22	42420	30		175	Fungi	1		
			B8	Fuc	CC1G_10558	A8NDX2_COPC7 Uncharacterized protein OS=Coprinopsis cinerea (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2	82.68	13733	14		50		1		
			B10				fleA		Q4WW81 LECF_ASPFU Fucose-specific lectin OS=Neosartorya fumigata (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=fleA PE=1 SV=1		82.54			34661	33
			B15	Gal	CC1G_10558	A8NDX2_COPC7 Uncharacterized protein OS=Coprinopsis cinerea (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2	55.12	55.12	13733		6		3		
			B23				AFUA_1G04130		Q4WK08_ASPFU FG-GAP repeat protein, putative OS=Neosartorya fumigata (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=AFUA_1G04130 PE=4 SV=2		89.9			33760	31
B25		40.39	13	25	Af	3									
B28		Man	89.9	30	139	1									
B29		Fuc	CC1G_10558	A8NDX2_COPC7 Uncharacterized protein OS=Coprinopsis cinerea (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2	78.74	13733	10	58	Fungi	1					
B_31					fleA		Q4WW81 LECF_ASPFU Fucose-specific lectin OS=Neosartorya fumigata (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=fleA PE=1 SV=1	76.51		34661	30		317	Af	1
B_32								73.97			29		159		1
Sa	B_40	Neu5Ac	SAPIO_CDS8583	AOA084FZZ4_PSEDA Uncharacterized protein OS=Pseudallescheria apiosperma OX=563466 GN=SAPIO_CDS8583 PE=3 SV=1	40.74	13912	7	44	Sa	1					
	B_41	Gal			82.22		9	25		1					
	B_43	Man			82.19		8	42		1					
	B_45	Fuc	fleA	Q8NJT4 LECF_ASPFM Fucose-specific lectin OS=Neosartorya fumigata OX=746128 GN=fleA PE=3 SV=1	83.76	34520	30	92	Fungi	1					
	B_46				81.53		30	189		1					
	B_47	CC1G_10558	A8NDX2_COPC7 Uncharacterized protein OS=Coprinopsis cinerea (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2	78.74	13733	9	62	1							
	B_48			95.28		11	104		1						
	B_53	GlcNAc	SAPIO_CDS8583	AOA084FZZ4_PSEDA Uncharacterized protein OS=Pseudallescheria apiosperma OX=563466 GN=SAPIO_CDS8583 PE=3 SV=1	82.22	13912	8	31	Sa	1					

Af: *Aspergillus fumigatus*, Sa: *Scedosporium apiospermum*, cov.: coverage

1	Very confident
2	Quite confident
3	Low confidence

5.4.3 Identification of SapCP from *S. apiospermum*

Four small proteins of ~13 kDa were recovered from the extracellular fraction of *S. apiospermum* (B40, B41, B43 and B53). Initially, we consider that those bands corresponded to different proteins since they were recovered using distinct affinity resin (Fetuin, Gal, Fuc and GlcNAc). Yet, according to the peptides identified by mass spectrometry, they all correspond to a hypothetical protein (accession number A0A084FZZ4) from the genome of *S. apiospermum*. This uncharacterized protein contains a secretory signal peptide of 18 residues and seems to belong to the Cerato-platanin proteins family (CPPs). This is a group of small secreted proteins with oligosaccharide-binding activity that are involved in fungus-host interactions. We called it SapCP for *Scedosporium apiospermum* cerato-platanin. The alignment of the peptides identified for each one of those samples and the SapCP, as well as the *in silico* modeling of SapCP structure, are shown in figure 5.5.

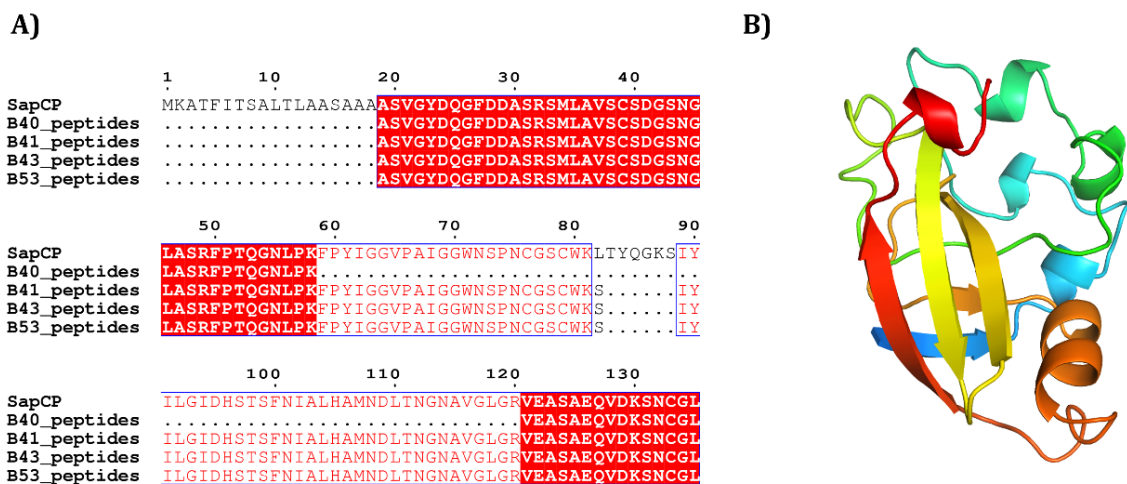


Figure 5.5: SapCP. **A)** Alignment of the sequences of SapCP (accession number A0A084FZZ4) and the peptides identified for samples B40, B41, B43 and B53. Sequence's alignment was performed using Clustal W [100] and visualized with ESPrpt3 [101]. **B)** Modeling the structure of SapCP. Model was generated by the RaptorX webserver [107]. The order of the polypeptide sequence is indicated by a gradual change of colour from blue (N-terminal end) to red (C-terminal end).

5.4.3.2 Cerato-platanin proteins (CPPs) –generalities and biological roles

The family of CPPs is a group of small fungal proteins rich in cysteines that are relatively new to science [108]. This family was created in 2004 and owes its name to the phytotoxin ceratoplatanin (CP) produced by the ascomycete *Ceratocystis platani*, which is a 120 amino acids protein secreted extracellularly. In general, all the member of the CPPs family are highly secreted to the medium, so they are abundantly found in culture filtrates [109,110]. However, some studies have also reported their localization on fungal cell walls [111,112].

CP is considered as microbe-associated a molecular pattern (MAMP) due its ability to induces host defense responses [109,113]. CPPs have been shown to play a dual role in fungi where, on one hand, they are elementary for the life cycle of those organisms and display multifunctional roles during the different stages of their development, such as the growth of hyphae, remodeling of fungal cell wall and the formation of chlamydospores [114,115]. On the other hand, they are also related to fungal colonization through various processes that can be beneficial or lethal to the host [116]. Some of those proteins, for example, are elicitors of defense responses in plants through diverse mechanisms that include activation of reactive oxygen species, production of nitric oxide, MAPK activation, ethylene and jasmonic acid signaling, etc.[109]. In contrast to their role in plant resistance, some CPPs can act as phytotoxins and are key elements during various stages of pathogenicity [110,111,115,117]. For example, it has been shown that CPPs gene expression increases when *Trichoderma virens* and *Trichoderma harzianum* are grown in the presence of tomato or cotton roots [118,119], while the CPP knockout mutants of *B. cinerea* and *M. grisea* display reduced virulence on their host plants [120,121]. It is speculated that their carbohydrate binding activity might play an essential role during fungus-host interactions through the recognition of plant polysaccharides during early stages

of infections [113,122]. Glycans recognized by CPPs that have been reported to date include GlcNAc [123], chitin [122,123] cellulose and β 3-glucan [124,125].

To date, the study of CPPs has been exclusively devoted to fungi interacting with plants and there is extensive literature on them, whereas only two CPPs have been identified from human pathogenic fungi: CS-Ag from *Coccidioides immitis* which is a specific antigen of *Coccidioides* [126] and rAsp f13 of *Aspergillus fumigatus* is an allergen for humans [127].

5.4.3.3 CPPs distribution

CPPs appear to exist only in fungi, bacteria, oomycota, plants but animals do not possess CP homologs. Genes encoding for CPPs have been found, so far, in more than 100 fungal genomes, including species from the ascomycota and basidiomycota branches of dikarya [109]. It is remarkable that, while basidiomycota possess multiple CPP encoding genes (>12), the ascomycota only possesses one homolog with few exceptions that might have up to 2 or 4 [108]. CPPs genes seem to be lost in early branches of jelly fungi as well as in some groups of yeast or yeast- like forms in their life cycle [108].

5.4.3.4 CPPs biochemical properties

Almost all CP homologues contain a secretory signal peptide, which explain their extracellular prevalence. The length of mature CPPs ranges from 105-134 amino acids per monomer and they can exist as dimers, trimers, tetramers or a mixture of those forms [112,128]. An N-glycosylation site is also present in most CPPs sequences, although no evidence of glycosylation has been found in CP *in vivo* [110]. Interestingly, during our analysis, the complete sequence of mature SapCP was identified in three out of four samples with exception of the peptide LTYQGKS (residues 82 -88), which was consistently missing in all of them. The absence of this peptide could be explained by the presence of glycosylations, which has been previously reported for some others CPPs, such as CS-Ag.

However, given that this peptide does not contain the signal sequence for N-glycosylation (Asn-X-Ser/Thr), we consider that the lack of its spectra must be due to an O-glycosylation at the residue of Threonine-83, Serine-88 or both.

5.4.3.5 CPPs structure

Although only five CPPs structures have been solved to date, the overall topology is highly conserved (Figure 5.6). CP presents a globular fold that contains two α -helices and six β -strands that form a double $\psi\beta$ -barrel fold [113]. In CP and MpCP3-5, the external face of barrel reveals a shallow crevice rich in polar and aromatic residues, which is responsible for the interaction with N-acetylglucosamine [123]. This is also widely distributed among other families of proteins, e.g. expansins, fungal endoglucanases, and barwins plant defense proteins, where are involved in the recognition/modification of polysaccharides [113].

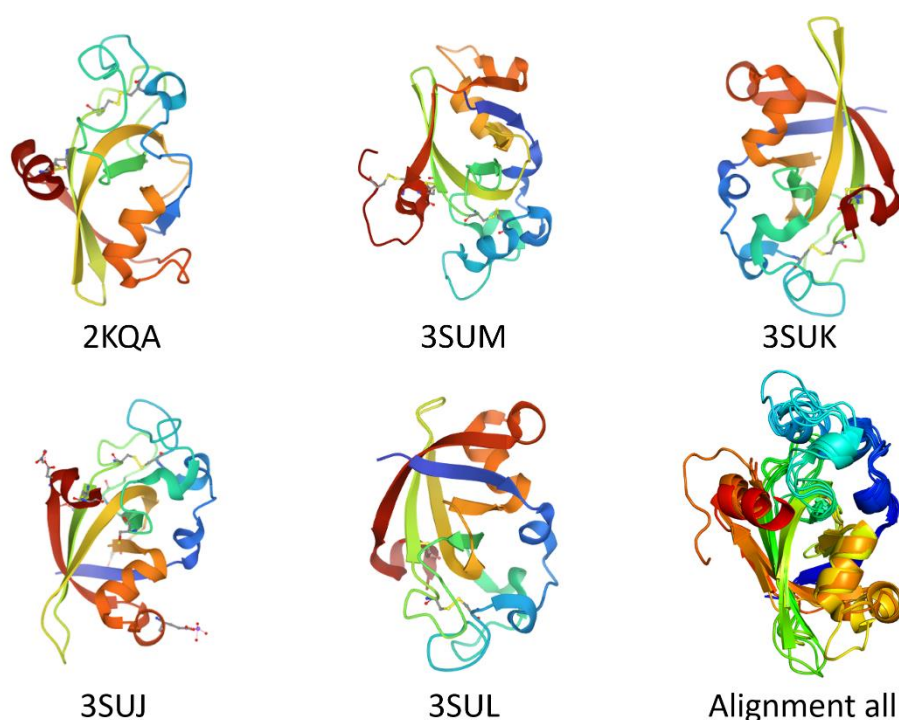


Figure 5.6: CPPs structural analysis. 3D structures of CP from *Ceratocystis platani* (PDB: 2KQA) [113]; MpCP1 (PDB: 3SUM), MpCP2 (PDB: 3SUK), MpCP3 (PDB: 3SUJ), and MpCP5 (PDB: 3SUL) from *Moniliophthora perniciosa* [123]. The bottom right corner of the panel presents structural alignment of those five CPPs and the structure modelled for SapCP.

SapCP contains all the conserved elements of the CPPs proteins family for its fold and for the recognition of polysaccharides. Its recovery from distinct affinity columns during our study can be explained by the fact that CPPs have been reported to bind different carbohydrate structures.

5.4.3.6 Potential applications of SapCP

Given the relevance of CPP in fungal development and its possible role during host recognition and colonization processes, there is an immeasurable number of possible uses for SapCP. Therefore, we consider that would be very interesting to achieve its full characterization.

SapCP could be used:

- As specific antigen for the detection of *Scedosporium* as in the case of CS-Ag from *C. immitis*.
- As a sensitization agent of the immune system for immunosuppressed patients.
- To elucidate its role during host-recognition in order to evaluate its potential as target for the development of anti-adhesive molecules.
- To evaluate if it is involved in remodelling and growth of the fungal cell wall. In such a case, it could become a target for the development of novel antifungal therapies.
- To elucidate the first crystallographic structure of a CP from human pathogens that could be used for drug design.

5.4.4 Identification of PVL-like lectins from *S. apiospermum* and *A. fumigatus*

A GlcNAc binding protein was recovered from the intracellular fraction of proteins from *S. apiospermum* and *A. fumigatus* (Figure 5.4, B1 and B7). As observed in electrophoretic profiles, the protein B1 corresponds to one of the most abundant proteins in *Scedosporium* sample while the apparent abundance

of B7 in *Aspergillus* is considerably lower (Figures 5.2 and 5.3). It has been possible to identify 34 different peptides for sample B1 and 25 for B7. The overall analysis of those peptides did not display any match with proteins reported in the databases of *Scedosporium* nor *Aspergillus* but, interestingly, they both displayed high identity scores with the PVL lectin from the mushroom *Psathyrella velutina* (accession number Q309D1_9AGAR). The coverage percent were 69.11% and 50.63 %, for B1 and B7, respectively (Figures 5.7 and 5.8).



Figure 5.7: Alignment of the peptides identified from B1 and the PVL sequence. Sequence's alignment was performed using Clustal W [100] and visualized with ESPrpt3 [101].

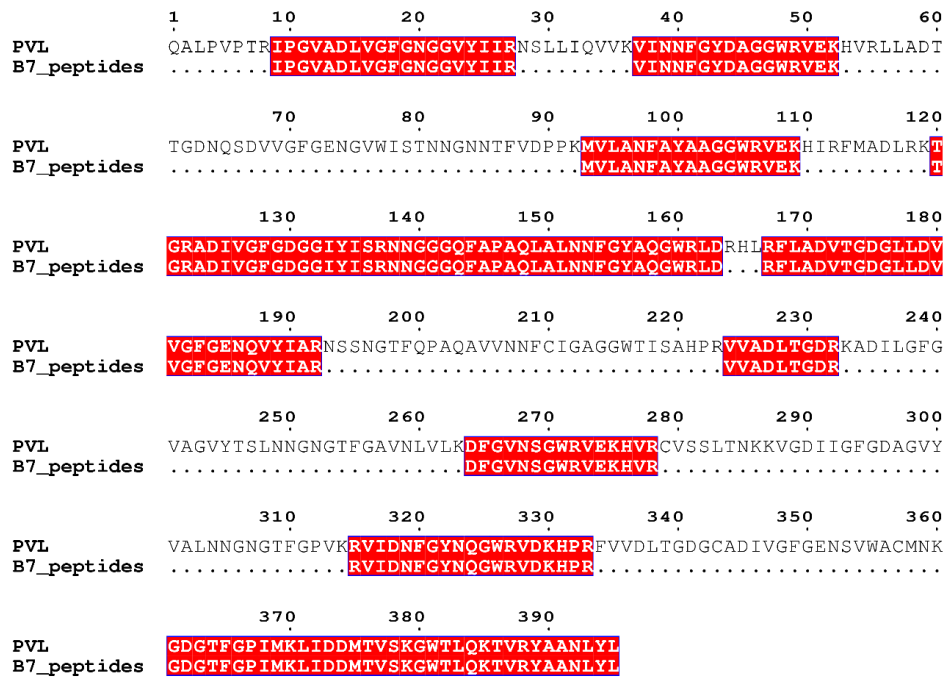


Figure 5.8. Alignment of the peptides identified from B7 and the PVL sequence. Sequence's alignment was performed using Clustal W [100] and visualized with ESPript3 [101].

5.4.4.2 *Psathyrella velutina* Lectin (PVL)

PVL is a 7 beta-propeller lectin of ~42 kDa with affinity for terminal GlcNac that has been purified from the fruiting bodies of the mushroom recently renamed *Lacrymaria lacrymabunda* (Figure 5.9) [53].

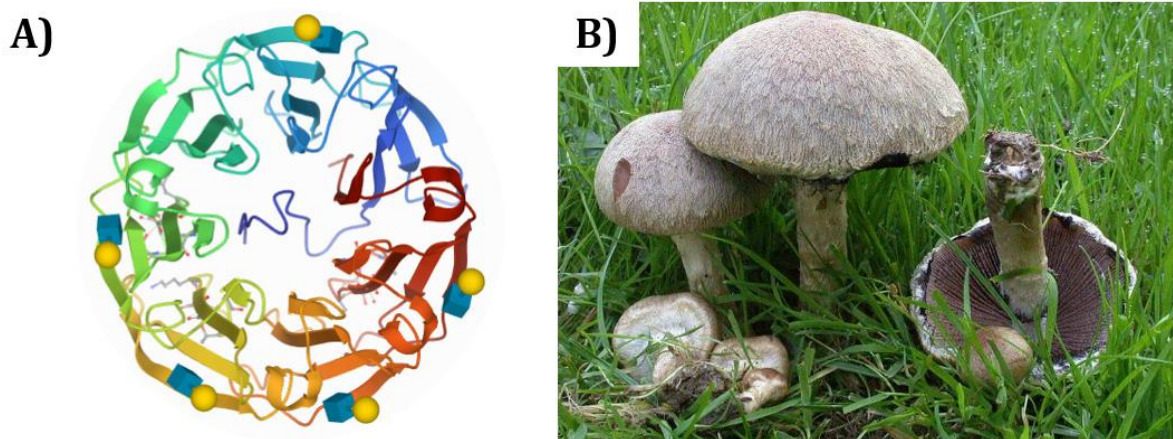


Figure 5.9: *Psathyrella velutina* Lectin (PVL). A) crystal structure recombinant PVL in complex with GlcNAc-D-1,3Galactoside (Reprinted with permission from Audfray *et al.*, 2014, PDB:4UP4). B) Photo of *P. velutina* mushroom (taken from the French mycology database, mycobd.fr).

PVL has been shown to have different binding specificity compared to classical GlcNAc-specific lectins, as it preferentially binds to non-reducing GlcNAc residues such as GlcNAc- β -1,2Man [129]. This is particularly interesting in the context of chronic inflammatory diseases of an autoimmune nature such as rheumatoid arthritis [130,131], as well as in microbial and viral infections, where due to the absence of terminal galactose, there is a prevalence of aberrant N-glycans with terminal GlcNAc on immunoglobulins (IgG) [132].

It has been shown that PVL also can bind to N-acetylneuraminic acid (Neu5Ac) and negatively charged polysaccharides such as heparin and pectin fragments [133]. However, while GlcNAc and Neu5Ac share the same binding pocket, polysaccharides appear to be recognized by an independent site [134]. This characteristic, together with its multivalency, suggests that PVL might play an important role during host-binding and the saprophytic colonization of soils. The most plausible mechanism establishes that the mycelial colonization is guided to areas rich in organic materials through the crosslinking of plant cell wall components, such as pectins or polygalacturonic acids, with cell surface glycoconjugates containing GlcNAc or Neu5Ac [53,134]

5.4.4.3 PVL distribution in fungi kingdom and phylogeny

The presence of PVL-like sequences has been found in the genomes of other mushrooms and of several bacteria [53] but its presence has not been reported in any microfungi to date. To better understand the implications of this distribution, first it is necessary to briefly review the fungi evolution.

The fungal kingdom is divided in “Early diverging fungi” and the “Dykaria subkingdom” (Figure 5.10). In turn, dikarya is composed by ascomycota and basidiomycota phylums, to which belong the vast majority of fungal species known to date. Ascomycetes are fungi with septate mycelium that produce endogenous ascospores. It is the largest taxonomic group and includes a great

diversity of types of fungi, such as false mushrooms, truffles, terrestrial languages, most lichens, some molds and yeasts. On the other hand, basidiomycetes are fungi that contain a microscopic structure called basidia, which produces haploid spores by meiosis. This division includes macroscopic fungi such as the classic mushrooms, represented in edible fungi, toxic fungi, hallucinogenic fungi and phytopathogenic fungi. Due to the early divergence of these two phylums, there are considerable differences between their life cycle, reproduction and morphology.

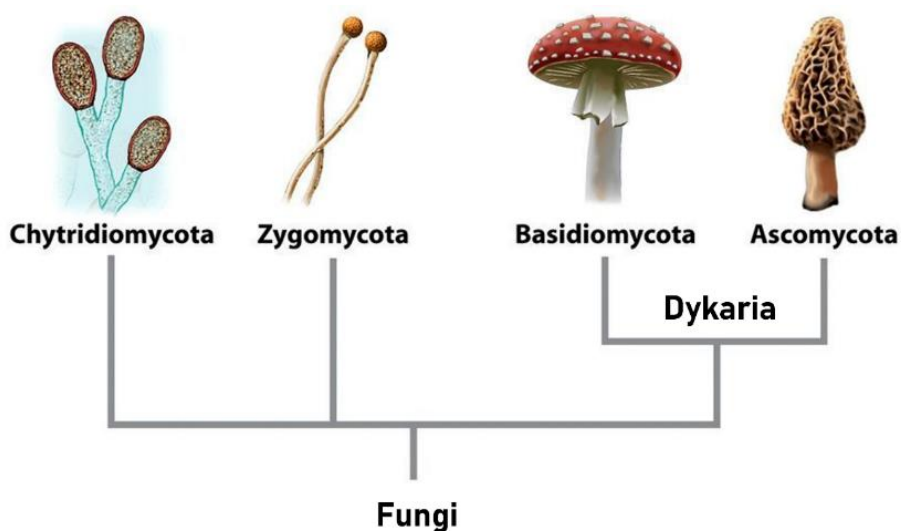


Figure 5.10: Evolutionary tree of the major groups of fungi. Modified from Biology: Life on Earth. 2008, Pearson Prentice Hall, Inc.

To date, the main fungal specimens where PVL-homologous lectins have been identified belong to the basidiomycota phylum, while *S. apiospermum* and *A. fumigatus* belong to the ascomycota one. Therefore, excited by the possibility of having discovered for the first time a PVL-like lectin in ascomycetes, we look all the proteins registered in the Swiss database UniProt that display percentages of identity superior to > 50 % with PVL and as expected, all these proteins were found only in basidiomycetes. Surprisingly, the phylogenetic analysis of those sequences placed B1 and B7 as the proteins closer related to PVL (Figure 5.11), that is possible, but not plausible due to the evolutionary distance between these

fungal species. Therefore, and considering the fact that we have previously worked in our laboratory with a recombinant version of PVL, the suspicion of a possible sample contamination aroused. To evaluate this, we compared the peptides identified for B1 and B7 with PVL-like proteins and we found that most of them correspond to highly conserved regions, which, in conjunction with the size discrepancy between B1 and B7, seems to detract from the possibility of cross-contamination. However, further studies are required to confirm the presence of those PVL-like lectins in *S. apiospermum* and *A. fumigatus*.

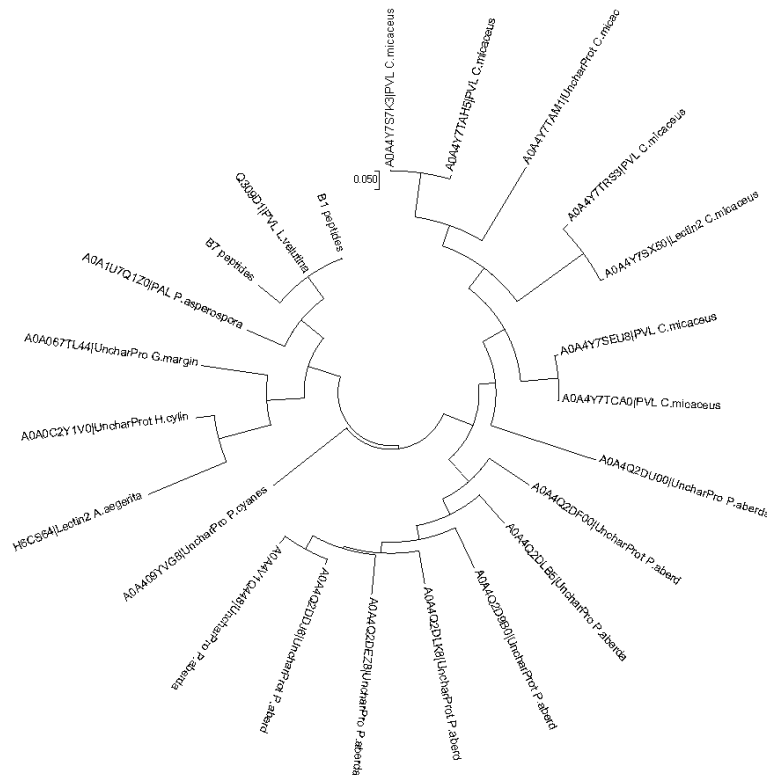


Figure 5.11: Phylogenetic analysis of PVL-like lectins. The evolutionary history was inferred using the maximum likelihood method based on Jones *et al.* algorithm 1992 [102]. The tree with the highest probability is displayed and drawn to scale. Evolutionary analyzes were performed in MEGA7 [103].

5.4.5 Identification of the fucose-binding lectins *FleA* and *SapL1*

This section describes the identification of the fucose-binding lectins from *A. fumigatus* (*FleA*) and *S. apiospermum* (*SapL1*), which have been widely described

in other sections of the manuscript. Hence, generalities of these lectins are omitted here.

5.4.5.2 Identification of FleA from *A. fumigatus*

The fucose binding lectin FleA (accession numbers Q4WW81 and Q8NJT4) from *A. fumigatus* was unambiguously identified in both, the intracellular and extracellular fraction of proteins of *Aspergillus*. In the intracellular fraction it corresponded to a protein of ~32 kDa (Figures 5.3 and 5.4, B10) that was recovered from the elution of the fucose-affinity chromatography and display 82.54% of coverage. A total of 33 different peptides were identified from this sample with high spectral count (231). In the extracellular fraction FleA was found in two different bands (B31 and B32) with a respective spectral count of 318 and 159. This is consistent with the amount of material observable in each band and serves as an indirect indication that most of the protein was secreted to the media. A total of 30 and 31 different peptides were identified from those samples, however, in both cases the percentage of coverage was the identical (76.51%). An alignment of identified peptides for B10, B31, B32 and FleA sequence is shown in figure 5.12.

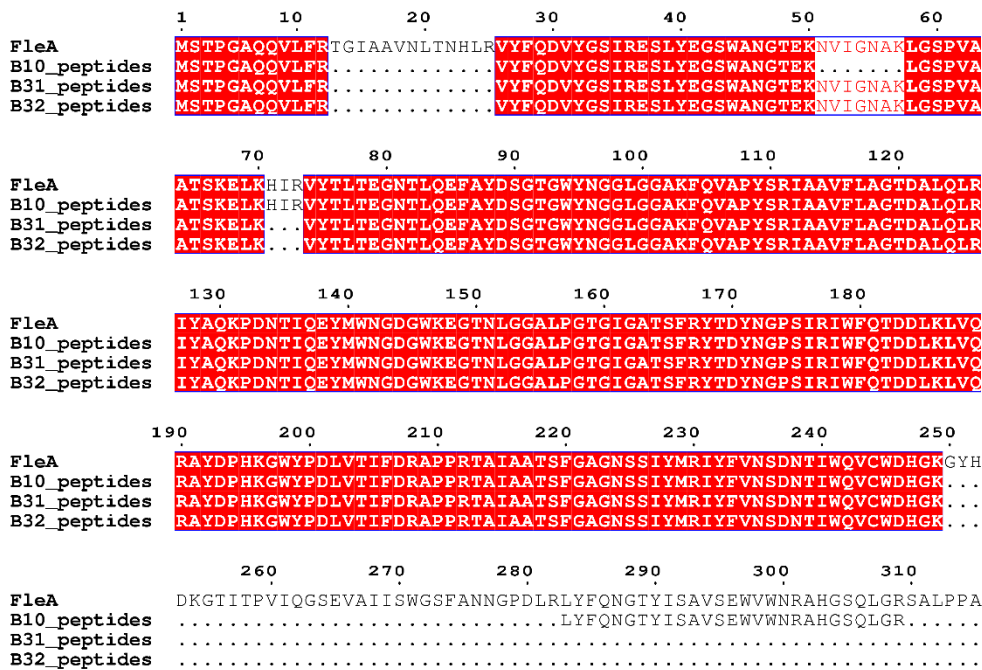


Figure 5.12. Alignment of the peptides identified from B10, B31, B32 and the FleA sequence. Sequence's alignment was performed using Clustal W [100] and visualized with ESPrnt3 [101].

5.4.5.3 Identification of FleA and SapL1 from *S. apiospermum*

Two proteins of ~35 and 50 kDa (B45 and B46, respectively) were recovered from the extracellular fraction of proteins of *S. apiospermum* after fucose-affinity chromatography. Initially we considered that one of those samples might correspond to SapL1, since the electrophoretic profile was consistent with the molecular size of this protein. However, only few SapL1 peptides were identified from band 45. Furthermore, the spectral account listed for those peptides was low, indicating that SapL1 is not the main protein in the sample. Instead, FleA was identified as the major protein in both bands.

There are two different sequences of FleA in the UniProt database, FleA_Q4WW81 from the database of *A. fumigatus* and FleA_Q8NJT4 from the database of fungi. Those sequences present two mismatches: in position 183, there is one additional residue of aspartic acid in the sequence of FleA_Q4WW81 that is not present in FleA_Q8NJT4 and at position 20, the serine residue is replaced by a lysine in FleA from the database of *A. fumigatus*. Interestingly, the

peptides identified from samples B45 and B46 contain a serine at position 20 as in FleA_Q8NJT4 and the extra aspartic acid as in FleA_Q4WW81 (Figure 5.13).

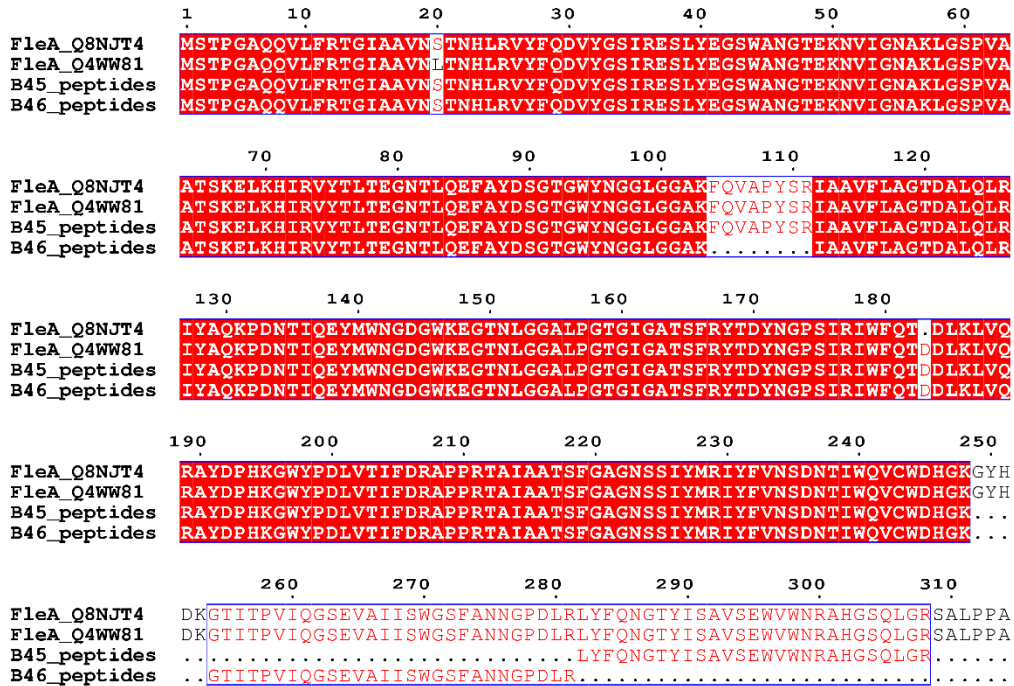


Figure 5.13: Alignment of the peptides identified from B45 and B46 and FleA sequences. Sequence's alignment was performed using Clustal W [100] and visualized with ESPrnt3 [101].

Even though, some SapL1 peptides were identified in B45, their spectral account was too low to be considered as evidence the identification of this lectin. This might be explained by the stage of development of the cultures, since it is known that this lectin is mainly expressed during the conidial stage. A different approach is being carried out to confirm the presence of SapL1 transcripts in mRNA from *S. apiospermum* conidia as they could not be isolated in mRNAs from mycelium.

5.4.6 Identification of AfFG-GAP repeat protein from *A. fumigatus*

A protein of approximately 33 kDa from the extracellular fraction of proteins of *A. fumigatus* was recovered from mannose and galactose affinity chromatography (B23 and B28, respectively). The sequence analysis of those

samples led towards the identification of >30 different peptides that display 89.9% of identity with a sequence of an hypothetical FG-GAP repeat protein deduced from the genome of *A. fumigatus* (accession number Q4WK08_ASPFU), further referred as AffG-GAP. It presents a 25 residues long signal peptide inferring secretion that is consistent with the extracellular identification of B23/B28. The peptide signal is excised since we observed 100% of coverage with the sequence of the mature protein and 89.9% with the immature one. AffG-GAP contains two repeats of the FG-GAP domain corresponding to residues 108-174 and 220-286 for the first and second repeat, respectively. They are indicated with dash lines in figure 5.14.

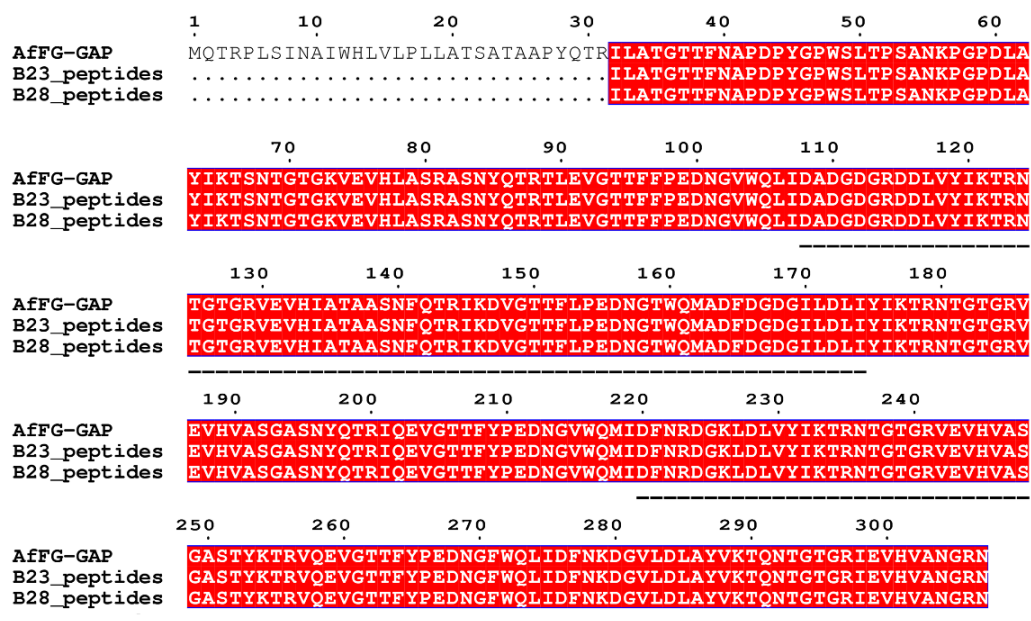


Figure 5.14: Alignment of the peptides identified from B23/B28 and AffG-GAP sequence. The dotted line indicates FG-GAP repeats. Sequence's alignment was performed using Clustal W [100] and visualized with ESPrpt3 [101].

Some peptides matching with the same sequence were also found from a band of ~10 kDa with galactose binding properties (sample B25), however, the coverage of sequences and the spectral count were considerably lower than those found for B23 and B28. Therefore, we consider that the peptides identified in this sample

could be the product of degradation of B23. The comparison of the AffG-GAP sequence and the peptides obtained for sample 25 is shown in figure 5.15.

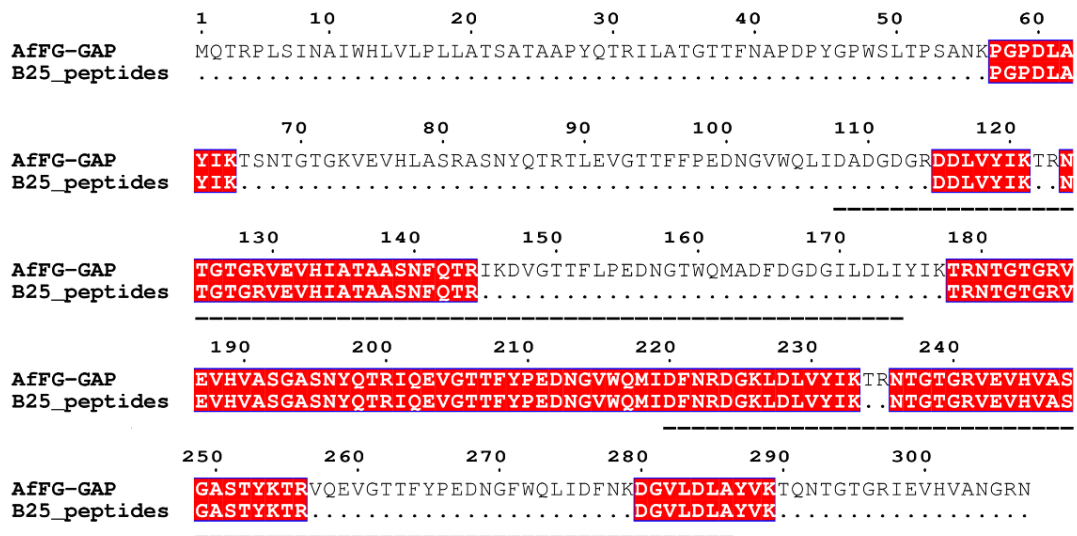


Figure 5.15: Alignment of the peptides identified from B25 and AffG-GAP sequence. The dotted line indicates FG-GAP repeats. Sequence's alignment was performed using Clustal W [100] and visualized with ESPript3 [101].

5.4.6.2 The FG-GAP repeat family

Members of the FG-GAP repeat protein family present a weak sequence homology with a domain mostly represented by the repeat FG (phenylalanyl-glycyl) and GAP (glycyl-alanyl-prolyl). This domain is commonly found in the N-terminal region of the α -chains of integrins, a region shown to be important for ligand binding and protein-protein interactions with some components of the extracellular matrix [135]. Furthermore, it has been mapped that in certain β -integrins the ligand binding is influenced by the FG-GAP repeats of the α -subunits [136].

5.4.6.3 FG-GAPs structure

In integrins, FG-GAPs are typically present in up to seven repeats and are predicted to fold into a β -propeller with putative calcium-binding motifs whilst in other FG-GAP proteins, they may have as few as two conserved motifs, like here for AffG-GAP [136]. With the sequence obtained from samples B23 and B28,

we performed an *in silico* structural model for AfFG-GAP revealing the formation of a 5 bladed β -propeller (Figure 5.16 A). This predicted model resembles the structure of the N-terminal domain of the α subunit of the integrins *i.e.*, a β -propeller made by up to 7 copies of FG-GAPs repeats.

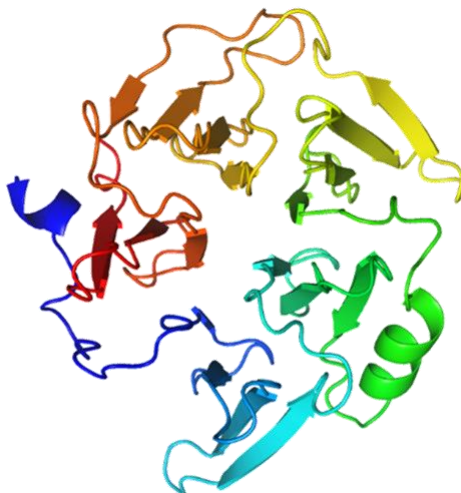


Figure 5.16: Structural model of AfFG-GAP. Model was generated using the RaptorX webserver [107] with the mature sequence of the protein identified from B23 and B28. The order of the polypeptide sequence is indicated by a gradual change of colour from blue to red (N-terminal to C-terminal ends).

5.4.6.4 AfFG-GAPs putative role

In 2011, four protein samples covering 9-43% of the sequence of AfFG-GAP were also identified in a broad proteomic approach aimed to identify *A. fumigatus* antigens. The authors proposed that, due to its possible interaction with the β -chain of human integrins, microbial secreted FG-GAP repeat proteins may be responsible for integrin-mediated cell adhesion and signal transduction during pathogenesis [136,137]. However, additional studies need to be conducted to validate AfFG-GAP biological roles and evaluate its potential for biotechnological applications.

5.4.7 Identification of CcMBL-homologs in *A. fumigatus* and *S. apiospermum*

Six of the samples sequenced display high identity with a hypothetical protein of ~13,7 kDa from the mushroom *Coprinopsis cinerea* (Figure 5.17). Although, this sequence corresponds to a hypothetical protein, it has been annotated in the Swiss database UniProt as a mucin binding lectin (accession numbers A8NDX2 and B3VS76). Therefore, we have called this protein *Coprinopsis cinerea* mucin-binding lectin (CcMBL). CcMBL-homologs were identified from both microfungi. Most of those samples were recovered from fucose-affinity chromatography but two were identified from galactose-affinity chromatography.

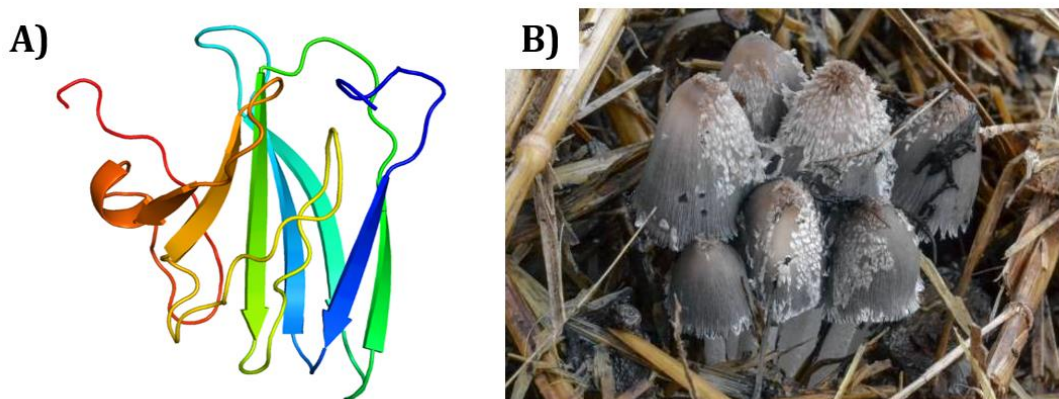


Figure 5.17: CcMBL. A) Modeling the structure of CcMBL sequence. Model was generated using the RaptorX webserver [107]. B) Photo of *Coprinopsis cinerea* mushroom (taken from the French mycology database, mycodb.fr).

5.4.7.2 CcMBL-like proteins from *S. apiospermum*

In the case of *Scedosporium*, two fucose-binding lectins were identified from both intracellular and extracellular fractions. They have a molecular weight of ~13 kDa (B2 and B47) and ~20 kDa (B3 and B48). According to peptides recovered from mass spectrometry and the analysis performed with the three *in silico* digested databases, those four proteins are similar to CcMBL with B1 covering 100% of the sequence, B3 and B48 covering both 95.28% and B47 covering 78.74% (Table 5.1

and Figure 5.20). Initially, we considered that the size discrepancies between the pairs of proteins B2/B3 and B47/B48 could be due the presence of glycosylation in the biggest proteins (B3 and B47), since glycosidic moieties increase the size of peptides hindering their identification in the expected region for their mass spectra, which result in the lack of the corresponding sequence. So, following this idea, we proposed the residue of serine-80 as the most plausible candidate to be glycosylated since the only difference between the reconstructed sequences for B3 and B2 is the SPVGR peptide, which is missing only in B3. However, when bands of B47 and B48 were latter analyzed, SPVGR peptide could not be identified in any of them. So, there are two possible explanations for it:

- a) serine-80 is glycosylated in B3/B47 but not in B2/B48 and SPVGR peptide could not be identified in B48 due the low quantity of material in this sample.
- b) this peptide was only identified in B2 because small peptides are difficult to detect in samples with low quantity of material. Hence, none of the proteins is glycosylated and that B3 and B47 correspond to thermostable protein dimers, that were not disrupted during the denaturation.

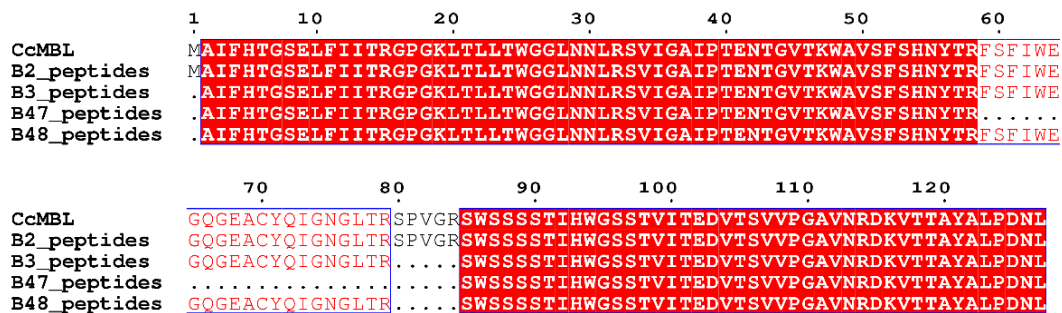


Figure 5.18 Alignment of the peptides identified from B2, B3, B47 and B48 with the sequence of CcMBL. Sequence's alignment was performed using Clustal W [100] and visualized with ESPript3 [101].

5.4.7.3 CcMBL-like proteins from *A. fumigatus*

During our second purification attempt, four fucose-binding lectins were recovered from the extract of *A. fumigatus* intracellular proteins (Figure 5.4). Only

two were selected for identification: FleA, previously discussed and a small protein of ~13 kDa (B8) which corresponded to the most abundant protein whose peptides covered 83% of the CcMBL sequence. On the other hand, a CcMBL-like protein (B29) was also identified from the extracellular fraction of *Aspergillus* proteins but its peptides only gave a 78.74% sequence coverage. The only difference between those two samples was, again, the absence of SPVGR peptide in B29 (Figure 5.21).

As in *Scedosporium* samples, fucose-binding proteins of ~20 kDa were identified in both intracellular and extracellular fractions but those bands have not been selected for identification in that round (B9 and B30, Figure 5.4). The possibility of finding similar proteins to CcMBL on those samples is not ruled out.

	1	10	20	30	40	50	60	
CcMBL	M	AIFHTGSELFII	TRGPGKLL	LLTWGGLN	LNRSVIGA	IP	TENTGVTKWAVSF	SHNYTRFSFIWE
B8_peptides	.	AIFHTGSELFII	TRGPGKLL	LLTWGGLN	LNRSVIGA	IP	TENTGVTKWAVSF	SHNYT.....
B29_peptides	.	AIFHTGSELFII	TRGPGKLL	LLTWGGLN	LNRSVIGA	IP	TENTGVTKWAVSF	SHNYT.....
	70	80	90	100	110	120		
CcMBL	GQGEACYQ	IIGNLTRSPVGR	RSWSSSSTI	HWGSSTVI	TEDVTSVVP	GAVNRDKVT	TAYALPDNL	
B8_peptides	RSPVGR	RSWSSSSTI	HWGSSTVI	TEDVTSVVP	GAVNRDKVT	TAYALPDNL	
B29_peptides	RSWSSSSTI	HWGSSTVI	TEDVTSVVP	GAVNRDKVT	TAYALPDNL	

Figure 5.19: Alignment of the peptides identified from B8 and B29 with the sequence of CcMBL.

5.4.7.4 CcMBL-like galactose-binding proteins

CcMBL similar proteins were also detected in the intracellular fraction of both fungi, *S. spiospermum* (B4) and *A. fumigatus* (B15) after galactose chromatography.

The analysis of mass spectrometry reveals that B4 actually contains a mixture of two different proteins. From one of them, 13 different peptides covering 99.21% of the sequence of CcMBL were identified, while the second one corresponds to a peptidylprolyl isomerase from *S. spiospermum* that will be latter discussed. The presence of this second protein explains the drastic difference on the amount of material between B4 and B15.

B15 was originally listed as CcMBL related protein with a high confidence score (level 1), however in a second analysis this score was degraded to level 3 since the spectral count of this sample was very low (14). This is common in samples with low concentration. Despite this fact, the 6 different peptides identified from this sample actually cover 55.12% of the CcMBL sequence and its presence is consistent with the finding of similar peptides in B4 from *Scedosporium*. The comparison of the CcMBL sequence and the peptides obtained for samples B4 and B15 is shown in figures 5.22.

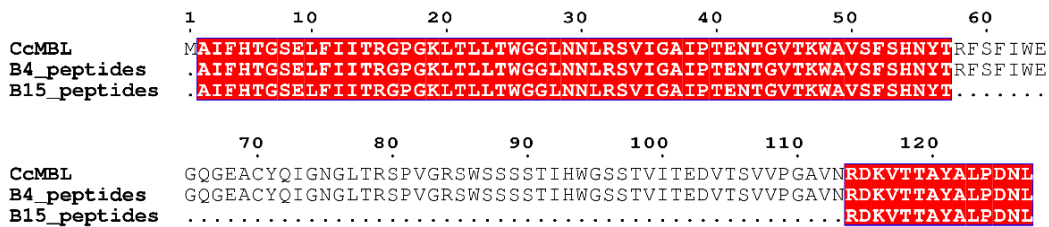


Figure 5.20: Alignment of the peptides identified from B15 and B4 with the sequence of CcMBL. Sequence's alignment was performed using Clustal W [100] and visualized with ESPrpt3 [101].

5.4.7.5 Discussion

CcMBL-like proteins were identified from six of the selected bands from both fucose and galactose affinity chromatography. Their sequence identity with CcMBL seems to suggest an evolutionary relationship between them. Nevertheless, it is not clear why these proteins were not annotated as hypothetical protein in both genomes. So, in order to inquire about the function of this protein, we performed a blast search for similar sequences in other organisms. Surprisingly, no hits were found in UniProt and no protein family was related to this sequence. This could be explained by the fact that *C. cinerea* paralog multi-copy genes are found in highly recombinant regions but this contrasts with our findings of highly similar sequences in both microfungi.

The affinity of those proteins is particularly interesting since it has been shown that mucins of cystic fibrosis lungs (CF) have an increased content of both fucose

and galactose. The increment of those sugars, derives from the abundance of the blood antigen group Lewis-X at the non-reducing end of the mucins glycosylations [22]. Therefore, the sugar-binding specificity of these proteins, would support the mucin-binding activity predicted for CcMBL. In conjunction, this information and the well-known role of mucins during fungal pathogenesis leads to the conclusion that the CcMBL-like proteins identified in this study might represent interesting drug targets. So, further studies need to be performed to confirm the presence of those proteins in the genomes of those microfungi and on their characterization.

5.4.8 Identification of SapFKBP12 from *S. apiospermum*

A strong band of ~12 kDa was recovered from galactose purification in the intracellular fraction of *S. apiospermum* (B4). This band contains CcMBL as previously discussed and a hypothetical peptidyl-prolyl isomerase (accession number A0A084G113), further referred as SapFKBP12. We don't know why this protein display affinity by galactose, but it is plausible that the overall fold of this monosaccharide resembles the structure of the SapFKBP12 substrate (peptidylproline). Surprisingly, SapFKBP12 could not be found when the peptide spectra were compared to the *Scedosporium in silico* digested database, but it was possible to identify it when the files were re-interrogated against an expanded database containing all the proteins of the fungal taxonomy. Paradoxically, it is listed as a *S. apiospermum* protein in the fungal database, which seems to be a good example of how incomplete annotation of databases and strains sequences variations might greatly influence the identification of proteins by mass spectrometry. The peptides identified cover 92.79% of the sequence of SapFKBP12, whose prevalence in B4 seems to correspond only approximately ~25% of the total amount of protein in the band according to the spectral counts (Figure 5.23).

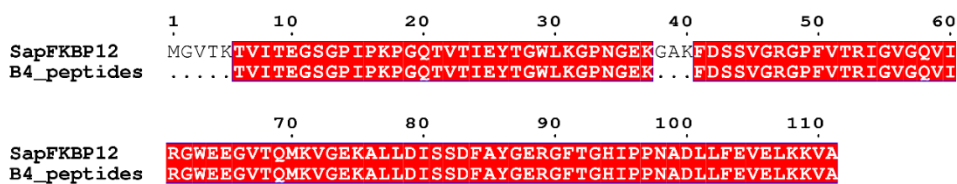


Figure 5.21: Alignment of the peptides identified from B4 and the SapFKBP12 sequence (accession number A0A084G113). Sequence's alignment was performed using Clustal W [100] and visualized with ESPript3 [101].

5.4.8.2 SapFKBP12 and immunophilins

SapFKBP12 belongs to the family of immunophilins that is a highly conserved cluster of intracellular peptidyl-prolyl isomerases (PPIases) catalyzing the interconversion between the *cis* and *trans* isomers of peptide bonds containing proline [138]. These proteins are important pharmacological targets against fungal infections and act as intracellular receptors for immunosuppressive drugs such as rapamycin, cyclosporine and tacrolimus (originally called FK506). Nevertheless, their isomerase activity does not appear to be related to their binding to these compounds [139].

5.4.8.3 Immunophilins-functions and classification

Immunophilins act as proteins chaperons and are fundamental for the proper folding of diverse proteins. Some of them are also involved in specific regulatory process including cellular signaling, apoptosis and transcription mechanisms [140,141]. Immunophilins are traditionally classified into two subfamilies: "cyclosporin-binding cyclophilins (CyPs)" and "FK506-binding proteins (FKBPs)". Besides, they can be grouped into a subclassification according to their cellular distribution [142]. According to our analysis, the ~12 kDa protein that we have identified (A0A084G113) belongs to the FKBP family and more specifically to the subgroup of cytoplasmic immunophilins FKBP12 so, we named it SapFKBP12.

FKBP12 subfamily owes its name to the first member discovered, which is a small protein constitutively expressed in all human tissues (hFKBP12). hFKBP12

homologs have been identified in a very wide range of eukaryotes [143]. The sequences of all the member of this subfamily are highly conserved and, in all cases, they correspond to small cytosolic proteins of ~12 kDa with at least one peptidyl-prolyl isomerases (PPIase) domain. As all the other subfamilies of FKBP proteins, FKBP12 members display affinity by the immunosuppressive drug FK506 (tacrolimus).

5.4.8.4 Calcineurin inhibition as antifungal therapy

In the presence of tacrolimus (FK506), FKBP12 proteins form a stable intracellular complex (FKBP12-FK506) that inhibits the phosphatase calcineurin (CaN) and prevent the dephosphorylation of its downstream targets [144]. This inhibition plays an ambivalent role in the treatment of fungal infections because fungal-CaN is fundamental for pathogenesis but, on the other hand, the inhibition of human-CaN leads towards immunosuppression (Figure 5.24).

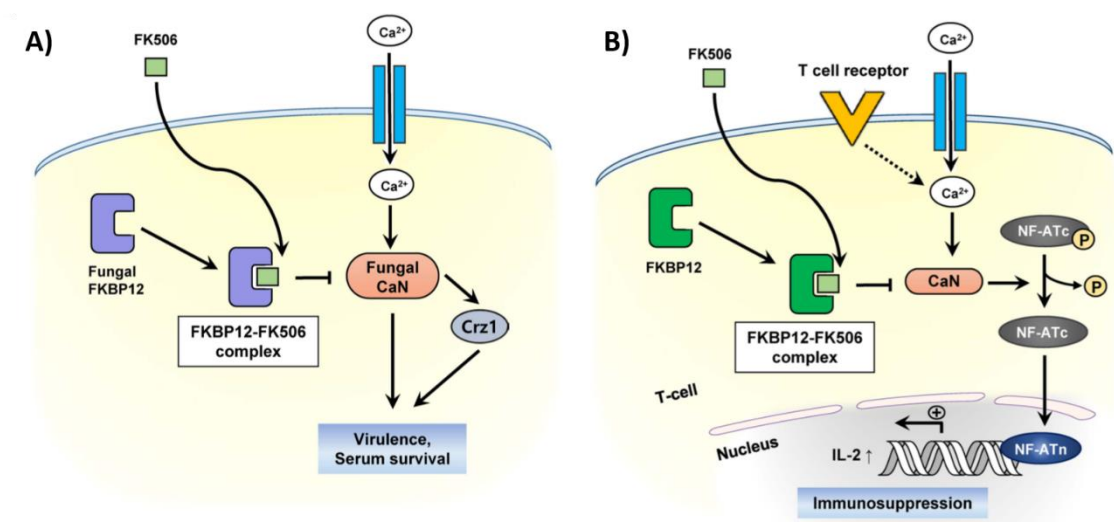


Figure 5.22: Ambivalent effect of calcineurin inhibition in fungi and human cells. A) Mechanism of the immunosuppressive action of FK506. **B)** Antifungal action of FK506 against fungal pathogens. Crz1: Calcineurin-responsive zinc finger 1. Adapted with permission from Jung *et al.*, 2020 [146].

- In fungi, CaN activates the transcription factor CRZ1 (calcineurin-responsive zinc finger transcription factor) which induces the expression of virulence genes on *Cryptococcus neoformans*, *C. albicans*, *C. glabrata*, and *A. fumigatus*,

among other deadly fungal pathogens [145]. Furthermore, CaN pathway plays an essential role during growth and hypha morphogenesis, so, its inhibition is an efficient strategy against fungal infections [146]. Consequently, calcineurin is considered as one of the most promising targets for antifungal drug development.

- In human cells, the function of CaN is to dephosphorylate the nuclear factor of activated T cells (NFAT). This phosphorylation allows it to be translocated to the nucleus where it cooperates with other nuclear transcription factors to induce the synthesis of interleukin-2 (IL-2) and initiate the transcription program for the activation of T cells. Therefore, CaN inhibition leads towards suppression of the immune system in humans.

5.4.8.5 FKBP12 and the panfungal strategy

Due to its central role in fungal pathogenesis, calcineurin is an attractive pharmacological target. Nevertheless, the development of specific antifungals targeting CaN inhibition is complex to address due the great similarity between human and fungal calcineurins whose catalytic (CnA) and a regulatory (CnB) subunits are nearly identical [143,147,148].

In this context, FKBP12s play an essential role since it has been shown that fungal-FKBP12s share only 40-50% of identity with their human counterpart [143]. Consequently, the development of novel drugs targeting FKBP12s is emerging as an alternative for the upstream inhibition of CaN in a highly specific manner. An additional advantage of this approach, lies in the fact that FKBP12s are very well distributed among the fungal kingdom and their sequences are highly conserved between human pathogens. Hence, fungal FKBP12s have emerged a prime candidate for the development of broad-spectrum antifungals and their discovery from emerging fungal pathogens will significantly aid in the development of a panfungal strategy (Figure 5.25).

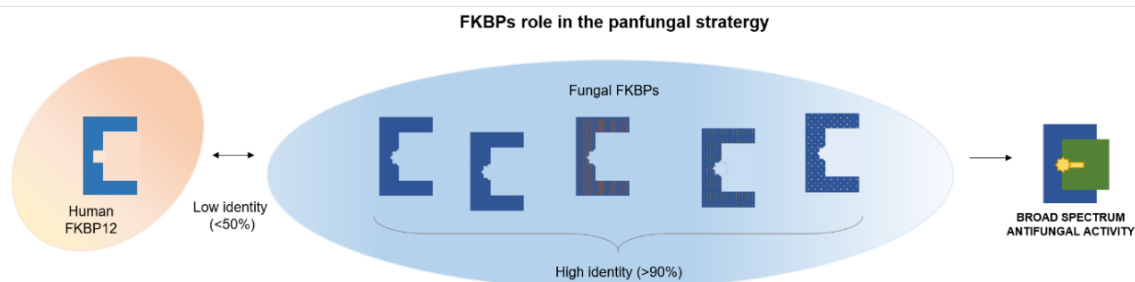


Figure 5.23: The role of FKBP's in the panfungal strategy. Schematic representation of the FKBP's conservation among humans and fungal kingdom. Figure created with Microsoft Office 365.

5.4.8.6 Structural differences of the FK506-FKBP12-CaN complex in humans and fungi

In 2019, Juvvadi1 and coworkers published the crystal structures of four FK506-mediated ternary complexes (FKBP12-FK506-CaN) from highly clinically relevant human fungal pathogens: *A. fumigatus* (PDB ID: 6TZ7), *C. albicans* (PDB ID: 6TZ6), *C. neoformans* (PDB ID: 6TZ8) and *Coccidioides immitis* (PDB ID: 5B8I) [148]. In all the fungal complexes, the overall fold of calcineurin was very similar to its human homologous (PDB ID: 4OR9) previously published by Griffith *et al.* 1995 [149]. While, the conformation of fungal and human FKBP12 (hFKBP12) revealed relevant differences, especially in the loops 40s and 80s, where most of the residues involved in FK506 are located (Figure 5.26). Most of those residues are also conserved in SapFKBP12 (Figure 5.26B, black triangles). They have also identified a phenylalanine (F88) residue that is critical for binding and is conserved among the fungal FKBP12 whilst it is replaced by a histidine in hFKBP12 (Figure 5.26B, yellow star). Hence, this residue has been proposed as an important guide in the development of FK506 analogs specific for the inhibition of fungal CaN.

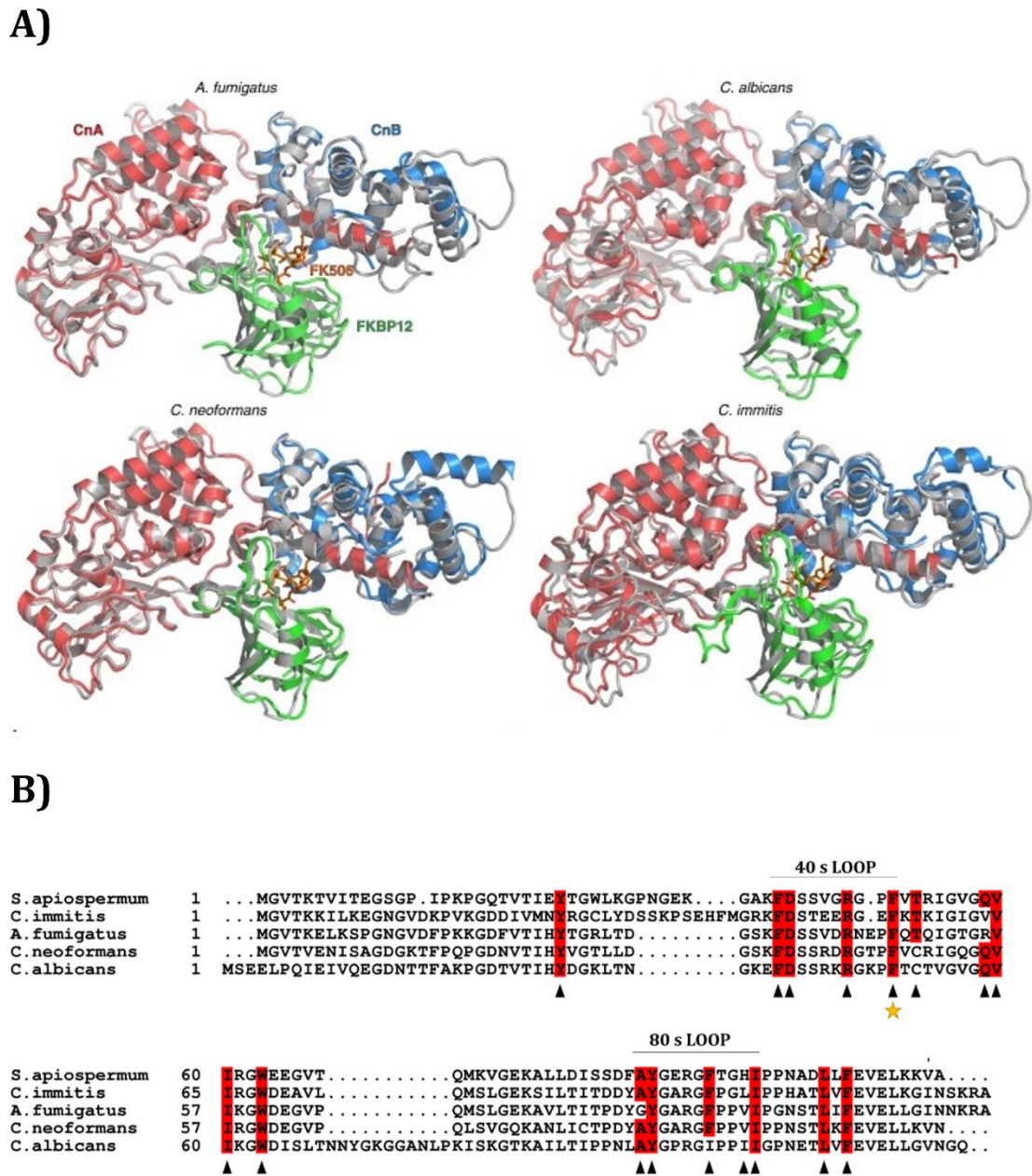


Figure 5.24: FKBP12 from pathogenic fungi: alignment and structure of complexes. A) Overlay of fungal CnAs with the bovine complex (bovine: PDB 1TCO gray; fungal: CnA red, CnB blue, FKBP12 green and FK506 in orange). Adapted with permission from Juvvadi *et al.*, 2019 [148]. UniProt accession number are given in parenthesis. B) Sequences alignment of fungal FKBP12s. *A. fumigatus* FKBP12 (6TZ7), *C. albicans* FKBP12 (P28870), *C. neoformans* FKBP12 (O94746), *C. immitis* (J3K5Z5), *S. apiospermum* FKBP12 (A0A084G113). Sequence's alignment was performed using Clustal W [100] and visualized with ESPript3 [101]. Residues within 5 Å of FK506 are indicated by black triangles. From those, conserved residues in fungi are highlighted red. The yellow star indicates the residue of phenylalanine 88 critical for FK506 binding.

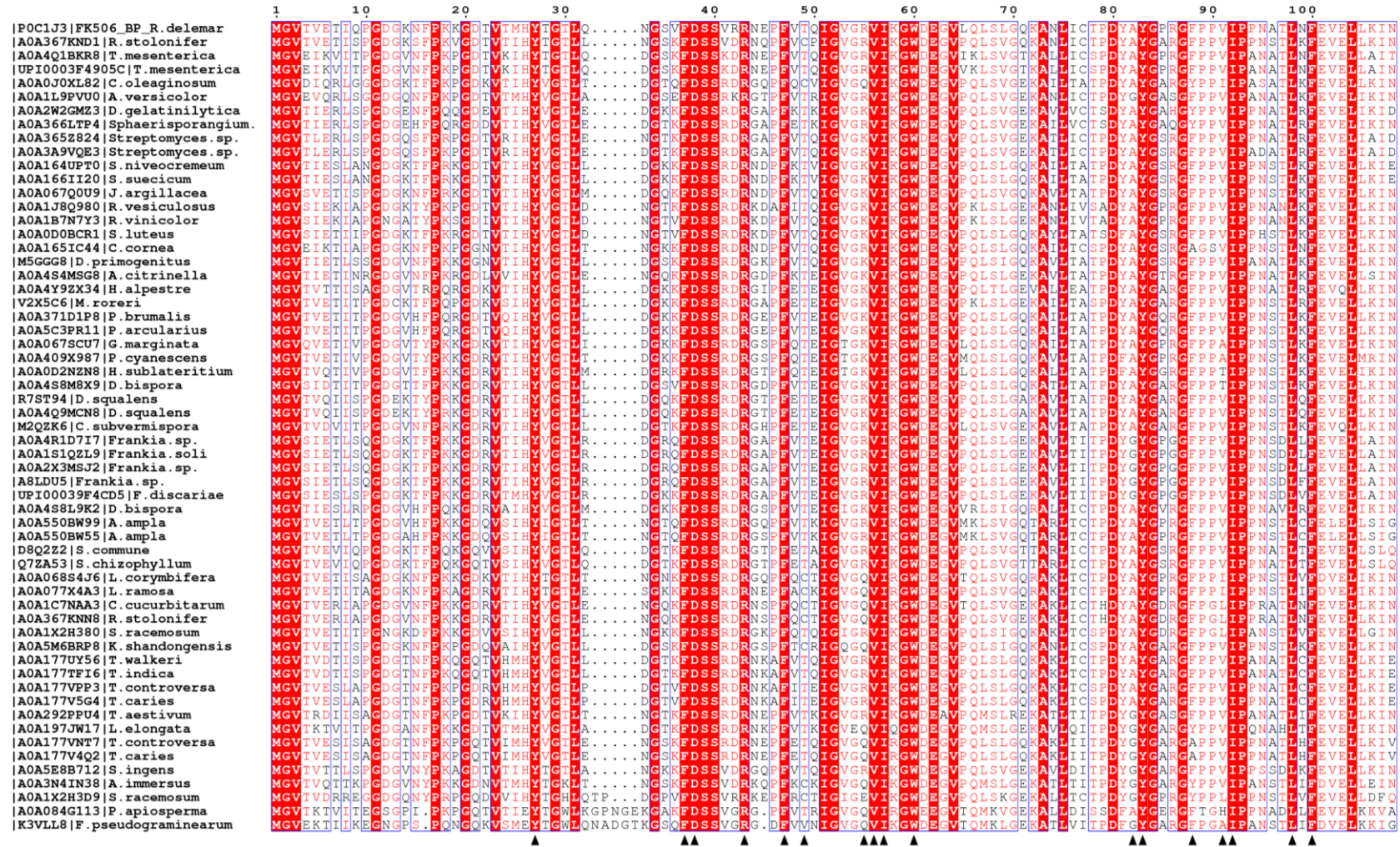


Figure 5.25: Alignment of sequences of highly conserved fungal FKBP12. The figure shows the alignment of all fungal FKBP12 that display >90% identity with SapFKBP12 from the UniProt database. Black triangles indicate the residues within 5 Å of FK506 in the crystal structures of the complexes with FKBP12 from *A. fumigatus* FKBP12 (6TZ7), *C. albicans* FKBP12 (P28870), *C. neoformans* FKBP12 (O94746), *C. immitis* (J3K5Z5). Sequence's alignment was performed using Clustal W [100] and visualized with ESPrpt3 [101].

5.4.8.7 SapFKBP12 potential applications

We have shown that the residues critical for FK506 binding are well conserved among SapFKBP12 and the FKBP12s from other fungal pathogens that display susceptibility towards tracolimus (FK506) treatment, such as the case of *A. fumigatus*, *C. albicans*, *C. neoformans* and *C. immitis*. This information is of utmost importance because it suggests the clinical use of tracolimus as a potential treatment in non-systemic scedosporiosis such as topical infections. This is the most common affection caused by this pathogen in immunocompetent patients and whose treatment is very challenging. Furthermore, the identification of SapFKBP12 will also contribute to the development of new treatments not only against *Scedosporium* infections but also in the development of a broad-spectrum antifungal therapy.

5.5 Conclusions

The analysis of peptides obtained by mass spectrometry led to the identification of 8 different proteins from the 22 bands that were analysed by mass spectrometry. Five of them were found in the protein extracts of *S. apiospermum* and 4 were identified from *A. fumigatus* extracts. The preliminary analysis of those proteins revealed that some might be attractive pharmacological targets, and others could find application as biotechnological tools. We consider that the proteins that deserve special attention are SapFKBP12 and CcMBL as pharmacological targets and SapCP as allergen or as antigens for the specific detection of *Scedosporium* in humans. The two PVL-like lectins could also be of medical relevance; however, their identification needs to be confirmed. On the other hand, we don't discard the possibility that in the 32 samples not sequenced there are also lectins that remain unidentified. Hence, the study of these samples should be addressed in the future.

The information provided here represents a general strategy for approaching the study of the lectinome of those two species of these fungal pathogens and, we hope that our findings will encourage the study and characterization of those new proteins, in the search for biomedical applications and new drug targets.

5.6 Contributions

Fungal cultures were provided by Prof. Muriel Cornet (MC). Valerie Chazalet (VC) performed extraction of total proteins. Dania Martínez Alarcón (DMA) performed protein purification. Mass spectrometry analysis was requested to the EDyP-service platform and were performed by Julia Novion (JN). The analysis of data was performed by DMA. DMA prepared all the figures and discussed results. Annabelle Varrot (AV) and Roland J. Pieters (RJP) administered the project, conceived the design of the study, obtained funding, contributed to data analysis, and supervised DMA.

5.7 Supplementary information

Table 5.2. List of peptides identified by mass spectrometry for sample B1

sequence	modifications	psm score	mass	accession number and description
QALPVPTR		45.1	880.513046	Q309D1_9AGAR Lectin PVL (Fragment) OS=Lacrymaria velutina OX=71681 GN=pvl
VVADLTGDR		59.89	944.492722	
NSLLIQVVK		52.65	1012.628067	
LIDDMTVSK		58.88	1020.516144	
GDGTFGPIMK		58.57	1021.49028	
DFGVNSGWR		53.38	1036.472656	
LIDDMTVSK	Oxidation (M5)	61.27	1036.511063	
GDGTFGPIMK	Oxidation (M9)	44.18	1037.485199	
GWTLQKTVR		42.67	1087.613831	
DFGVNSGWRVEK		67.71	1392.678619	
VIDNFGYNQGWR		83.89	1467.689514	
VINNFYDAGGWR		92.85	1467.689514	
MVLANFAYAAGGWR		98.99	1525.749985	
ADIVGFGDGGIYISR		137.47	1538.772919	
MVLANFAYAAGGWR	Oxidation (M1)	87.7	1541.744904	
RVIDNFGYNQGWR		86.06	1623.790619	
LIDDMTVSKGWTLQK		91.29	1733.902206	
LIDDMTVSKGWTLQK	Oxidation (M5)	112.4	1749.897125	
DFGVNSGWRVEKHVR		43.93	1784.907059	
VIDNFGYNQGWRVDK		88.25	1809.879837	
VINNFYDAGGWRVEK		77.7	1823.895477	
TGRADIVGFGDGGIYISR		121.84	1852.943176	
MVLANFAYAAGGWRVEK		52.26	1881.955948	
MVLANFAYAAGGWRVEK	Oxidation (M1)	92.58	1897.950867	
IPGVADLVGFGNGGGVYIIR		105.82	1916.052002	
GDGTFGPIMKLIDDMTVSK	Oxidation (M9, M15)	53.04	2055.985703	
VIDNFGYNQGWRVDKHPR		56.86	2200.092621	
NSLLIQVVKVINNFYDAGGWR		82.8	2462.307022	
NNGGGQFAPAQLALNNFGYAQQWR		154.34	2550.21524	
VGDIIGFGDAGVYVALNNGNGTFGPVK		102.86	2650.339157	
FLADVTGDGLLDVVGFGENQVYIAR		220.69	2667.354462	
KVGDIIIGFGDAGVYVALNNGNGTFGPVK		119.83	2778.434113	
VGDIIGFGDAGVYVALNNGNGTFGPVKR		120.72	2806.440262	
IPGVADLVGFGNGGGVYIIRNSLLIQVVK		35.51	2910.66951	
KVGDIIIGFGDAGVYVALNNGNGTFGPVKR		129.06	2934.535217	
FVVDLTGDGCADIVGFGENSVWACMNK	Carbamidomethyl (C10, C24)	83.05	2960.314346	
FVVDLTGDGCADIVGFGENSVWACMNK	Carbamidomethyl (C10, C24); Oxidation (M25)	93.52	2976.309265	
ADILGFGVAGVYTSLNNGNGTFGAVNLVLK		119.05	2980.565842	
KADILGFGVAGVYTSLNNGNGTFGAVNLVLK		193.53	3108.660797	

Table 5.3. List of peptides identified by mass spectrometry for sample B2.

sequence	modifications	psm score	mass	accession number and description
WAVSFHNYTR		84.35	1366.64183	A8NDX2_COPC7 Uncharacterized protein OS= <i>Coprinopsis cinerea</i> (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2
DKVTTAYALPDNL		64.29	1419.724548	
LTLTGWGGLNNLR		100.96	1469.835434	
SVIGAIPTENTGVTK		72.15	1485.803879	
AIFHTGSELFITR		86.79	1603.872223	
AIFHTGSELFITR	Acetyl (Protein N-term)	88.41	1645.882782	
MAIFHTGSELFITR		79.17	1734.912704	
MAIFHTGSELFITR	Oxidation (M1)	103.79	1750.907623	
GPGKLTLLTWGGLNNLR		94.31	1809.026093	
AIFHTGSELFITRGPVK		101.33	1943.062881	
FSFIWEGQGEACYQIGNGLTR	Carbamidomethyl (C12)	158.21	2432.121933	
SVIGAIPTENTGVTKWAVSFHNYTR		100.98	2834.43515	
SWSSSSTIHWGSSTVITEDVTSVVP GAVNR		110.29	3145.531662	
SWSSSSTIHWGSSTVITEDVTSVVP GAVNRDK		88.37	3388.653564	
AIFHTGSELFITRGPVKLTLTGWGGLNNLR		54.23	3394.887756	
SPVGRSWSSSSTIHWGSSTVITEDVTSVVP GAVNR		125.09	3641.807449	
SWSSSSTIHWGSSTVITEDVTSVVP GAVNRDKVTTAYALPDNL		32.37	4547.245651	

Table 5.4. List of peptides identified by mass spectrometry for sample B3.

sequence	modifications	psm score	mass	accession number and description
WAVSFHNYTR		84.67	1366.64183	A8NDX2_COPC7 Uncharacterized protein OS= <i>Coprinopsis cinerea</i> (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2
DKVTTAYALPDNL		64.2	1419.724548	
LTLTGWGGLNNLR		81.3	1469.835434	
SVIGAIPTENTGVTK		75.67	1485.803879	
AIFHTGSELFITR		82.9	1603.872223	
AIFHTGSELFITR	Acetyl (Protein N-term)	78.06	1645.882782	
GPGKLTLLTWGGLNNLR		92.54	1809.026093	
AIFHTGSELFITRGPVK		77.94	1943.062881	
FSFIWEGQGEACYQIGNGLTR	Carbamidomethyl (C12)	51.68	2432.121933	
SVIGAIPTENTGVTKWAVSFHNYTR		116.74	2834.43515	
SWSSSSTIHWGSSTVITEDVTSVVP GAVNR		93.2	3145.531662	

Table 5.5. List of peptides identified by mass spectrometry for sample B4.

sequence	modifications	pms score	mass	accession number and description	
WAVSFHNYTR		68.58	1366.64183	A8NDX2_COPC7 Uncharacterized protein OS= <i>Coprinopsis cinerea</i> (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2	
DKVTTAYALPDNL		32.49	1419.724548		
LTLTGWGGLNNLR		83.49	1469.835434		
SVIGAIPTENTGVTK		43.71	1485.803879		
AIFHTGSELFITR		64.95	1603.872223		
AIFHTGSELFITR	Acetyl (N-term)	67.77	1645.882782		
GPGKLLTLLTWGGLNNLR		50.07	1809.026093		
FSFIWEGQGEACYQIGNGLTR	CarbaMet(C12)	121.35	2432.121933		
SVIGAIPTENTGVTKWAVSFHNYTR		75.99	2834.43515		
SWSSSSTIHWGSSTVITEDVTSVVP GAVNR		98.04	3145.531662		
SWSSSSTIHWGSSTVITEDVTSVVP GAVNRDK		40.52	3388.653564		
AIFHTGSELFITRGPGLTLLTWGGLNNLR		60.34	3394.887756		
SPVGRSWSSSSTIHWGSSTVITEDVTSVVP GAVNR		69.13	3641.807449		
IGVGQVIR		29.82	840.518143		A0A084G113_PSEDA SEDA Peptidylprolyl isomerase OS= <i>Pseudallescheria apiosperma</i> OX=563466 GN=SAPIO_CDS7060 PE=4 SV=1
GWEEGVTQMK		44.68	1163.528107		
GWEEGVTQMK	Oxidation (M9)	42.57	1179.523026		
TVITEGSGPIPK		56.04	1197.660507		
FDSSVGRGPFVTR		41.86	1423.72084		
ALLDISSDFAYGER		83.18	1555.751816		
PGQTVTIEYTGWLK		99.48	1591.824615		
VGEKALLDISSDFAYGER		47.69	1968.979248		
IGVGQVIRGWEEGVTQMK		26.12	1986.03569		
GFTGHIPPADLLFEVELK		69.14	2096.094238		
PGQTVTIEYTGWLKGPNGEK		74.56	2174.100784		
GFTGHIPPADLLFEVELKK		75.08	2224.189194		
GFTGHIPPADLLFEVELKKVA		34.52	2394.294724		
TVITEGSGPIPKPGQTVTIEYTGWLK		111.24	2771.474564		

Table 5.6. List of peptides identified by mass spectrometry for sample B7.

sequence	modifications	psm score	mass	accession number and description
QALPVPTR		51.28	880.513046	Q309D1_9AGAR Lectin PVL (Fragment) OS= <i>Lacrymaria velutina</i> OX=71681 GN=pvl PE=1 SV=1
CVSSLTNK	CarbaMet (C1)	46.64	907.443314	
VVADLTGDR		62.16	944.492722	
NSLLIQVVK		53.52	1012.628067	
LIDDMTVSK		56.9	1020.516144	
GDGTFGPIMK		58.52	1021.49028	
DFGVNSGWR		39.71	1036.472656	
LIDDMTVSK	Oxidation (M5)	61.32	1036.511063	
GDGTFGPIMK	Oxidation (M9)	44.11	1037.485199	
VVADLTGDRK		30.46	1072.587677	
TVRYAANLYL		39.67	1182.639679	
DFGVNSGWRVEK		76.03	1392.678619	
VIDNFGYNQGWR		92.25	1467.689514	
VINNFYDAGGWR		84.93	1467.689514	
MVLANFAYAAGGWR		82.81	1525.749985	
ADIVFGDGGIYISR		121.89	1538.772919	
MVLANFAYAAGGWR	Oxidation (M1)	97.35	1541.744904	
RVIDNFGYNQGWR		86.09	1623.790619	
LIDDMTVSKGWTLQK		104.13	1733.902206	
LIDDMTVSKGWTLQK	Oxidation (M5)	107.91	1749.897125	
DFGVNSGWRVEKHVR		39.66	1784.907059	
VIDNFGYNQGWRVDK		88.16	1809.879837	
VINNFYDAGGWRVEK		77.02	1823.895477	
TGRADIVFGDGGIYISR		76.45	1852.943176	
MVLANFAYAAGGWRVEK	Oxidation (M1)	67.72	1897.950867	
IPGVADLVGFGNGGVYIIR		108.44	1916.052002	
VIDNFGYNQGWRVDKHPR		67.54	2200.092621	
NNGGQFAPLAQLALNFGYAQGWR		134.48	2550.21524	

FLADVTGDLDDVVGFGENQVVIAR		127.29	2667.354462
NNGGGQFAPAQLALNNGFYAQGWRLDR		100.53	2934.427353

Table 5.7. List of peptides identified by mass spectrometry for sample B8.

sequence	modifications	psm score	mass	accession number and description
WAVSFHNYTR		74.46	1366.64183	A8NDX2_COPC7 Uncharacterized protein OS= <i>Coprinopsis cinerea</i> (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2
DKVTTAYALPDNL		64.16	1419.724548	
LTLTGWGLNLR		88.63	1469.835434	
SVIGAIPENTGVTK		70.06	1485.803879	
AIFHTGSELIITR		79.43	1603.872223	
AIFHTGSELIITR	Acetyl (N-term)	61.96	1645.882782	
GPGKLTLLTWGGLNLR		85.27	1809.026093	
AIFHTGSELIITRGPVK		60.87	1943.062881	
SVIGAIPENTGVTKWAVSFHNYTR		109.56	2834.43515	
SWSSSTIHWGSSTVITEDVTSVVPNAVNR		100.3	3145.531662	
GPGKLTLLTWGGLNLRVIGAIPENTGVTK		46.83	3276.819412	
AIFHTGSELIITRGPVKLTLLTWGGLNLR		75.86	3394.887756	
SPVGRSWSSSTIHWGSSTVITEDVTSVVPNAVNR		115.62	3641.807449	
SWSSSTIHWGSSTVITEDVTSVVPNAVNRDKV TTAYALPDNL		35.5	4547.245651	

Table 5.8. List of peptides identified by mass spectrometry for sample B10.

sequence	modifications	psm score	mass	accession number and description
AHGSQLGR		27.18	824.425293	LECF_ASPFU Fucose-specific lectin OS= <i>Neosartorya</i> <i>fumigata</i> (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=fleA PE=1 SV=1
LGSPVAATSK		69.77	929.518188	
FQVAPYSR		51.6	966.49231	
IWFQTDLLK		48.44	1164.581528	
YTDYNGPSIR		70.75	1184.546188	
STPGAQQVLF		72.31	1202.640778	
STPGAQQVLF	Acetyl (N-term)	83.39	1244.651337	
LGSPVAATSKELK		83.61	1299.739792	
MSTPGAQQVLF		77.47	1333.681259	
VYFQDVYGSIR		81.25	1345.666656	
GWYPLDVTIFDR		80.31	1480.735077	
IAAVFLAGTDALQLR		93.57	1557.887863	
ESLYEGSWANGTEK		71.35	1569.694687	
IWFQTDLLKLVQR		104.96	1660.893692	
TAIAATSFAGNSSIYMR		147.55	1816.877762	
TAIAATSFAGNSSIYMR	Oxidation (M17)	112.63	1832.872681	
EGTNLGGALPGTGIGATSF		100.55	1874.948654	
GWYPLDVTIFDRAPPR		49.55	1901.978821	
AYDPHKGWYPLDVTIFDR		92.93	2192.069092	
APPRTAIAATSFAGNSSIYMR	Oxidation (M21)	52.26	2254.116425	
ESLYEGSWANGTEKVNIGNAK		147.98	2266.086548	
YTDYNGPSIRIWFQTDLLK		53.41	2331.117157	
LYFQNGTYISAVSEWVWNR		57.47	2332.127655	
IYFVNSDNTIWQVCWDHGK	Carbamidomethyl	72.26	2381.08992	
IYAQKPDNTIQEYMWNGDGWK		59.47	2556.174332	
IYAQKPDNTIQEYMWNGDGWK	Oxidation (M14)	65.13	2572.16925	
AYDPHKGWYPLDVTIFDRAPPR		61.19	2613.312836	
YTDYNGPSIRIWFQTDLLKLVQR		64.77	2827.429321	
VYFQDVYGSIRVYEGSWANGTEK		95.13	2897.350784	
EGTNLGGALPGTGIGATSFRTYDNGPSIR		83.16	3041.484283	
VYTLTEGNTLQEFAYDSGTGWYNGGLGAK		123.78	3168.467636	
IYAQKPDNTIQEYMWNGDGWKEGTNLGGALPGTGIGATSF		97.91	4413.112427	
IYAQKPDNTIQEYMWNGDGWKEGTNLGGALPGTGIGATSF	Oxidation (M14)	79	4429.107346	

Table 5.9. List of peptides identified by mass spectrometry for sample B15

sequence	modifications	psm score	mass	accession number and description
WAVFSHNYTR		46.93	1366.64183	A8NDX2_COPC7 Uncharacterized protein OS= <i>Coprinopsis cinerea</i> (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2
DKVTTAYALPDNL		34.53	1419.724548	
LTLTGWGGLNNLR		77.45	1469.835434	
SVIGAIPTENTGVTK		68.59	1485.803879	
AIFHTGSELFITR		76.72	1603.872223	
GPGKLLTWGGLNNLR		50.68	1809.026093	

Table 5.10. List of peptides identified by mass spectrometry for sample B23.

sequence	modifications	psm score	mass	accession number and description
ASNYQTR		48.84	838.393311	Q4WK08_ASPFU FG-GAP repeat protein, putative OS= <i>Neosartorya fumigata</i> (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=AFUA_1G04130 PE=4 SV=2
LDLVYIK		49.9	862.516388	
DDLVIYK		49.02	864.459274	
PGPDLAYIK		78.64	972.528015	
DGVLDLAYVK		63.76	1091.586273	
LDLVYIKTR		77.36	1119.665176	
DGKLDLVYIK		63.32	1162.65976	
VEVHVASGASTYK		81.23	1346.683029	
DGKLDLVYIKTR		66.78	1419.808548	
VEVHVASGASTYKTR		86.54	1603.831818	
VEVHVASGASNYQTR		93.71	1616.79068	
VEVHIATAASNQTR		99.86	1642.842712	
TSNTGTGKVEVHLASR		72.8	1655.8591	
DGVLDLAYVKTQNTGTGR		73.37	1906.974869	
TQNTGTGRIEVHVANGRN		48.15	1922.967087	
NTGTGRVEVHVASGASTYK		53.17	1932.965363	
TRNTGTGRVEVHVASGASTYK		43.02	2190.114151	
NTGTGRVEVHVASGASNYQTR		124.13	2203.073013	
NTGTGRVEVHIATAASNQTR		68.41	2229.125046	
TRNTGTGRVEVHVASGASNYQTR		64.86	2460.221802	
TRNTGTGRVEVHIATAASNQTR		51.26	2486.273834	
VQEVGTTFFPEDNGFWQLIDFNK		46.6	2746.291519	
IQEVGTTFFPEDNGVWQMIDFNR		85.4	2758.26973	
IQEVGTTFFPEDNGVWQMIDFNR	Oxidation (M18)	80.66	2774.264648	
TLEVGTTFPEDNGVWQLIDADGDGR		61.1	2851.330109	
DVGTTFPEDNGTWQMADFDDGILD LIYIK	Oxidation (M16)	85.9	3474.617752	
ILATGTTFNAPDPYGPWSLTPSANKPGPD LAYIK		73.44	3572.819122	
TLEVGTTFPEDNGVWQLIDADGDGRDD LVYIK		86.7	3697.778824	
IKDVGTTFPEDNGTWQMADFDDGILD LIYIK	Oxidation (M18)	94.82	3715.796768	
VQEVGTTFFPEDNGFWQLIDFNKDVLD LAYVK		53.79	3819.867233	
TLEVGTTFPEDNGVWQLIDADGDGRDD LVYIKTR		83.53	3954.927612	

Table 5.11. List of peptides identified by mass spectrometry for sample B25

sequence	modifications	psm score	mass	accession number and description
LDLVYIK		44.8	862.516388	Q4WK08_ASPFU FG-GAP repeat protein, putative OS= <i>Neosartorya fumigata</i> (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=AFUA_1G04130 PE=4 SV=2
DDLVIK		43.17	864.459274	
PGPDLAYIK		78.64	972.528015	
DGVLDLAYVK		47.15	1091.586273	
DGKLDLVYIK		51.72	1162.65976	
VEVHVASGASNYQTR		95.99	1616.79068	
VEVHIATAASNFQTR		83.15	1642.842712	
NTGTGRVEVHVASGASTYK		45.51	1932.965363	
NTGTGRVEVHVASGASTYKTR		52.6	2190.114151	
NTGTGRVEVHVASGASNYQTR		70.35	2203.073013	
NTGTGRVEVHIATAASNFQTR		65.74	2229.125046	
TRNTGTGRVEVHVASGASNYQTR		44.74	2460.221802	
IQEVGTTFFPEDNGVWQMIDFNR	Oxidation (M18)	27.1	2774.264648	

Table 5.12. List of peptides identified by mass spectrometry for sample B28

sequence	modifications	psm score	mass	Accession number and description
ASNYQTR		48.76	838.393311	Q4WK08_ASPFU FG-GAP repeat protein, putative OS= <i>Neosartorya fumigata</i> (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=AFUA_1G04130 PE=4 SV=2
LDLVYIK		52.77	862.516388	
DDLVIK		43.72	864.459274	
PGPDLAYIK		78.65	972.528015	
DGVLDLAYVK		64.57	1091.586273	
DGKLDLVYIK		70.01	1162.65976	
VEVHVASGASTYK		71.38	1346.683029	
DGKLDLVYIKTR		66.79	1419.808548	
VEVHVASGASNYQTR		94.8	1616.79068	
VEVHIATAASNFQTR		99.74	1642.842712	
TSNTGTGKVEVHLASR		74.9	1655.8591	
DGVLDLAYVKTQNTGTGR		65.55	1906.974869	
TQNTGTGRIEVHVANGRN		47.79	1922.967087	
NTGTGRVEVHVASGASTYK		87.16	1932.965363	
NTGTGRVEVHVASGASTYKTR		81.21	2190.114151	
NTGTGRVEVHVASGASNYQTR		122.09	2203.073013	
NTGTGRVEVHIATAASNFQTR		76.84	2229.125046	
TRNTGTGRVEVHVASGASNYQTR		69.42	2460.221802	
ILATGTTFNAPDPYGPWSLTSPANK		68.75	2618.301666	
VQEVGTTFFPEDNGFWQLIDFNK		40.89	2746.291519	
IQEVGTTFFPEDNGVWQMIDFNR		97.1	2758.26973	
IQEVGTTFFPEDNGVWQMIDFNR	Oxidation (M18)	67.97	2774.264648	
TLEVGTTFPPEDNGVWQLIDADGDGR		72.54	2851.330109	
IQEVGTTFFPEDNGVWQMIDFNRDGK	Oxidation (M18)	43.41	3074.40802	
DVGTTFLPEDNGTWQMADFDGDLIYIK	Oxidation (M16)	75.59	3474.617752	
ILATGTTFNAPDPYGPWSLTSPANKPGPDLAYIK		86.01	3572.819122	
TLEVGTTFPPEDNGVWQLIDADGDGRDDLVIK		73.21	3697.778824	
IKDVGTTFLPEDNGTWQMADFDGDLIYIK	Oxidation (M18)	110.86	3715.796768	
VQEVGTTFFPEDNGFWQLIDFNKGVLDLAYVK		46.81	3819.867233	
TLEVGTTFPPEDNGVWQLIDADGDGRDDLVIKTR		86.53	3954.927612	

Table 5.13. List of peptides identified by mass spectrometry for sample B29

sequence	modifications	psm score	mass	accession number and description
WAVFSHNYTR		80.95	1366.64183	A8NDX2_COPC7 Uncharacterized protein OS= <i>Coprinopsis cinerea</i> (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2
DKVTAYALPDNL		74.36	1419.724548	
LTLTGWGGLNNLR		101.05	1469.835434	
SVIGAIPTENTGVTK		74.03	1485.803879	
AIFHTGSELIITR		79.99	1603.872223	
AIFHTGSELIITR	Acetyl (N-term)	61.83	1645.882782	
GPGKLTLLTWGGLNNLR		85.85	1809.026093	
AIFHTGSELIITRGPVK		72.09	1943.062881	
SVIGAIPTENTGVTKWAVFSHNYTR		118.52	2834.43515	
SWSSSSTIHWGSSTVITEDVTSVVPQAVNR		102.51	3145.531662	

Table 5.14. List of peptides identified by mass spectrometry for sample B31.

sequence	modifications	psm score	mass	accession number and description
AHGSQLGR		32.83	824.425293	LECF_ASPFU Fucose-specific lectin OS= <i>Neosartorya fumigata</i> (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=fleA PE=1 SV=1
LGSPVAATSK		67.44	929.518188	
FQVAPYSR		53.67	966.49231	
IWFQTDDLK		52.17	1164.581528	
YTDYNGPSIR		68.96	1184.546188	
STPGAQQVLFRR		79.45	1202.640778	
STPGAQQVLFRR	Acetyl (N-term)	86.71	1244.651337	
LGSPVAATSKELK		85.24	1299.739792	
MSTPGAQQVLFRR		74.76	1333.681259	
VYFQDVYGSIR		74.2	1345.666656	
GWYPDLVTIFDR		80.36	1480.735077	
IAAVFLAGTDALQLR		100.63	1557.887863	
ESLYEGSWANGTEK		72.86	1569.694687	
NVIGNAKLGSPVAATSK		73.16	1625.910049	
IWFQTDDLKLVQR		105.28	1660.893692	
TAIAATSFAGNSSIYMR		147.63	1816.877762	
TAIAATSFAGNSSIYMR	Oxidation (M17)	112.39	1832.872681	
EGTNLGGALPGTGIGATSFRR		108.19	1874.948654	
GWYPDLVTIFDRAPRR		48.97	1901.978821	
AYDPHKGWYPDLVTIFDR		90.75	2192.069092	
APPRTAIAATSFAGNSSIYMR	Oxidation (M21)	96.96	2254.116425	
ESLYEGSWANGTEKRVIGNAK		134.29	2266.086548	
YTDYNGPSIRIWFQTDDLK		59.8	2331.117157	
IYFVNSDNTIYQVCWDHGK	Carbamidomethyl (C14)	72.2	2381.08992	
IYAQKPDNTIQEYMWNGDGWK	Oxidation (M14)	85.18	2572.16925	
AYDPHKGWYPDLVTIFDRAPRR		76.65	2613.312836	
YTDYNGPSIRIWFQTDDLKLVQR		71.43	2827.429321	
VYFQDVYGSIRESLYEGSWANGTEK		77.52	2897.350784	
VYTLTEGNTLQEFAYDSGTGWYNGGLGGAK		109.41	3168.467636	
IYAQKPDNTIQEYMWNGDGWKEGTNLGGA LPGTGIGATSFRR		81.13	4413.112427	

Table 5.15. List of peptides identified by mass spectrometry for sample B32.

sequence	modifications	psm score	mass	accession number and description
LGSPVAATSK		69.99	929.518188	<p>LECF_ASPFU Fucose-specific lectin OS=<i>Neosartorya fumigata</i> (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=fleA PE=1 SV=1</p>
FQVAPYSR		54.1	966.49231	
IWFQTDDLK		50.17	1164.581528	
YTDYNGPSIR		67.25	1184.546188	
STPGAQQVLF		89.62	1202.640778	
STPGAQQVLF	Acetyl (N-term)	86.62	1244.651337	
LGSPVAATSKELK		88.77	1299.739792	
MSTPGAQQVLF		75.84	1333.681259	
VYFQDVYGSIR		74.98	1345.666656	
GWYPDLVTIFDR		80.38	1480.735077	
IAAVFLAGTDALQLR		101.46	1557.887863	
ESLYEGSWANGTEK		71.82	1569.694687	
NVIGNAKLGSPVAATSK		73.53	1625.910049	
IWFQTDDLKLVQR		115.02	1660.893692	
TAIAATSFAGNSSIYMR		143.2	1816.877762	
TAIAATSFAGNSSIYMR	Oxidation (M17)	112.51	1832.872681	
EGTNLGGALPGTGIGATSF		129.01	1874.948654	
GWYPDLVTIFDRAPPR		45.63	1901.978821	
AYDPHKGWYPDLVTIFDR		90.7	2192.069092	
ESLYEGSWANGTEKNVIGNAK		96.98	2266.086548	
YTDYNGPSIRIWFQTDDLK		53.74	2331.117157	
IYFVNSDNTIWQVCWDHGK	Carbamidomethyl (C14)	28.23	2381.08992	
IYAQKPDNTIQEYMWNGDGWK		58.46	2556.174332	
AYDPHKGWYPDLVTIFDRAPPR		63.17	2613.312836	
YTDYNGPSIRIWFQTDDLKLVQR		85.94	2827.429321	
VYFQDVYGSIRESLYEGSWANGTEK		89.8	2897.350784	
VYTLTEGNTLQEFAYDSGTGWYNGGLGGAK		108.68	3168.467636	
IYAQKPDNTIQEYMWNGDGWKEGTNLGGA LPGTGIGATSF		57.96	4413.112427	
IYAQKPDNTIQEYMWNGDGWKEGTNLGGA LPGTGIGATSF	Oxidation (M14)	79.95	4429.107346	

Table 5.16. List of peptides identified by mass spectrometry for sample B40.

sequence	modifications	psm score	mass	accession number and description
FPTQGNLPK		58.59	1000.53418	<p>A0A084FZZ4_PSEDA Uncharacterized protein OS=<i>Pseudallescheria apiosperma</i> OX=563466 GN=SAPIO_CDS8583 PE=3 SV=1</p>
VEASAEQVDK		48.75	1074.519302	
VEASAEQVDKSNCG	Carbamidomethyl (C13)	90.09	1605.730438	
SMLAVSCSDGSNGLASR	Carbamidomethyl (C7)	109.89	1710.76651	
SMLAVSCSDGSNGLASR	Oxidation (M2); Carbamidomethyl (C7)	127.55	1726.761429	
SMLAVSCSDGSNGLASRFPTQGNLPK	Carbamidomethyl (C7)	73.2	2693.290131	
SMLAVSCSDGSNGLASRFPTQGNLPK	Oxidation (M2); Carbamidomethyl (C7)	91.94	2709.285049	

Table 5.17. List of peptides identified by mass spectrometry for sample B41.

sequence	modifications	psm score	mass	accession number and description
FPTQGNLPK		44.32	1000.53418	A0A084FZZ4_PSEDA Uncharacterized protein OS= <i>Pseudallescheria apiosperma</i> OX=563466 GN=SAPIO_CDS8583 PE=3 SV=1
VEASAEQVDKSNCGI	Carbamidomethyl (C13)	75.72	1605.730438	
SMLAVSCSDGSNGLASR	Carbamidomethyl (C7)	129.7	1710.76651	
SMLAVSCSDGSNGLASR	Oxidation (M2); Carbamidomethyl (C7)	127.13	1726.761429	
FPYIGGVPAIGGWNSPNCGSCWK	Carbamidomethyl (C18, C21)	58.63	2523.146393	
SMLAVSCSDGSNGLASRFPTQGNLPK	Carbamidomethyl (C7)	92.32	2693.290131	
SMLAVSCSDGSNGLASRFPTQGNLPK	Oxidation (M2); Carbamidomethyl (C7)	108.87	2709.285049	
SIYILGIDHSTSFNIALHAMNDLTNGNAVGLGR	Oxidation (M20)	62.23	3499.751785	
IGVGQVIR		29.82	840.518143	A0A084G113_PSEDA Peptidylprolyl isomerase OS= <i>Pseudallescheria apiosperma</i> OX=563466 GN=SAPIO_CDS7060 PE=4 SV=1
GWEEGVTQMK		44.68	1163.528107	
GWEEGVTQMK	Oxidation (M9)	42.57	1179.523026	
TVITEGSGPIPK		56.04	1197.660507	
FDSSVGRGPFVTR		41.86	1423.72084	
ALLDISSDFAYGER		83.18	1555.751816	
PGQTVTIEYTGWLK		99.48	1591.824615	
VGEKALLDISSDFAYGER		47.69	1968.979248	
IGVGQVIRGWEEGVTQMK		26.12	1986.03569	
GFTGHIPPADLLFEVELK		69.14	2096.094238	
PGQTVTIEYTGWLKGPNGEK		74.56	2174.100784	
GFTGHIPPADLLFEVELKK		75.08	2224.189194	
GFTGHIPPADLLFEVELKKVA		34.52	2394.294724	
TVITEGSGPIPKPGQTVTIEYTGWLK		111.24	2771.474564	

Table 5.18. List of peptides identified by mass spectrometry for sample B43.

sequence	modifications	psm score	mass	accession number and description
FPTQGNLPK		58.36	1000.53418	A0A084FZZ4_PSEDA Uncharacterized protein OS= <i>Pseudallescheria apiosperma</i> OX=563466 GN=SAPIO_CDS8583 PE=3 SV=1
VEASAEQVDKSNCGI	Carbamidomethyl (C13)	75.57	1605.730438	
SMLAVSCSDGSNGLASR	Carbamidomethyl (C7)	104.01	1710.76651	
SMLAVSCSDGSNGLASR	Oxidation (M2); Carbamidomethyl (C7)	126.97	1726.761429	
FPYIGGVPAIGGWNSPNCGSCWK	Carbamidomethyl (C18); Carbamidomethyl (C21)	65.96	2523.146393	
SMLAVSCSDGSNGLASRFPTQGNLPK	Carbamidomethyl (C7)	98.89	2693.290131	
SMLAVSCSDGSNGLASRFPTQGNLPK	Oxidation (M2); Carbamidomethyl (C7)	101.73	2709.285049	
SIYILGIDHSTSFNIALHAMNDLTNGNAVGLGR	Oxidation (M20)	44.98	3499.751785	

Table 5.19. List of peptides identified by mass spectrometry for sample B45.

sequence	modifications	psm score	mass	accession number and description
LGSPVAATSK		74.38	929.518188	LECF_AS PFU Fucose-specific lectin OS= <i>Neosartorya</i> <i>fumigata</i> (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=fleA PE=1 SV=1
FQVAPYSR		51.55	966.49231	
IWFQTDDLK		49.39	1164.581528	
YTDYNGPSIR		69.3	1184.546188	
STPGAQQVLF		89.59	1202.640778	
LGSPVAATSKELK		94.47	1299.739792	
MSTPGAQQVLF		74.81	1333.681259	
VYFQDVYGSIR		70.59	1345.666656	
GWYPDLVTIFDR		80.28	1480.735077	
IAAVFLAGTDALQLR		95.64	1557.887863	
ESLYEGSWANGTEK		70.69	1569.694687	
IWFQTDDLKLVQR		105.08	1660.893692	
TAIAATSFAGNSSIYMR		147.6	1816.877762	
EGTNLGGALPGTGIGATSFR		122.63	1874.948654	
GWYPDLVTIFDRAPPR		50.76	1901.978821	
AYDPHKGWYPDLVTIFDR		83.79	2192.069092	
ESLYEGSWANGTEKVNIGNAK		75.97	2266.086548	
YTDYNGPSIRIWFQTDDLK		53.86	2331.117157	
IYFVNSDNTIWQVCWDHGK	Carbamidomethyl (C14)	57.76	2381.08992	
FQVAPYSRIAIVFLAGTDALQLR		53.53	2506.369614	
IYAQKPDNTIQEYMWNGDGWK	Oxidation (M14)	64.18	2572.16925	
AYDPHKGWYPDLVTIFDRAPPR		85.95	2613.312836	
YTDYNGPSIRIWFQTDDLKLVQR		125.66	2827.429321	
EGTNLGGALPGTGIGATSFRTYTDYNGPSIR		76.68	3041.484283	
LYFQNGTYISAVSEWVWNRHAGSQLGR		90.34	3138.542389	
VYTLTEGNTLQEFAYDSGTGWYNGGLGGAK		101.72	3168.467636	
HIRVYTLTEGNTLQEFAYDSGTGWYNGGLGGAK		71.83	3574.711716	
VYFQDVYGSIRESLYEGSWANGTEKVNIGNAK		43.72	3593.742645	
IYAQKPDNTIQEYMWNGDGWKEGTNLGGALPGTGIGATSFR		72.45	4413.112427	
TGIAAVNSTNHLR		88.01	1352.716049	LECF_AS PFM Fucose-specific lectin OS= <i>Neosartorya</i> <i>fumigata</i> OX=746128 GN=fleA PE=3 SV=1
STPGAQQVLFRTGIAAVNSTNHLR	Acetyl (N-term)	93.45	2579.356827	
TGIAAVNSTNHLRVYFQDVYGSIR		105.76	2680.372147	

Table 5.20. List of peptides identified by mass spectrometry for sample B46.

sequence	modifications	psm score	mass	accession number and description
LGSPVAATSK		72.16	929.518188	<p>LECF_AS PFU Fucose-specific lectin OS=<i>Neosartorya fumigata</i> (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) OX=330879 GN=fleA PE=1 SV=1</p>
FQVAPYSR		51.59	966.49231	
IWFQTDDLK		45.25	1164.581528	
YTDYNGPSIR		67.27	1184.546188	
STPGAQQVLF		85.97	1202.640778	
LGSPVAATSKELK		89.55	1299.739792	
MSTPGAQQVLF		75.37	1333.681259	
VYFQDVYSIR		74.96	1345.666656	
GWYPDLVTIFDR		80.21	1480.735077	
IAAVFLAGTDALQLR		93.53	1557.887863	
ESLYEGSWANGTEK		80.91	1569.694687	
NVIGNAKLGSPVAATSK		72.22	1625.910049	
IWFQTDDLKLVQR		105.01	1660.893692	
TAIAATSFAGAGNSSIYMR		143.36	1816.877762	
EGTNLGGALPGTGIGATSF		111.87	1874.948654	
GWYPDLVTIFDRAPPR		44.89	1901.978821	
AYDPHKGWYPDLVTIFDR		78.38	2192.069092	
APPRTAIAATSFAGAGNSSIYMR	Oxidation (M21)	86.46	2254.116425	
ESLYEGSWANGTEKNVIGNAK		153.11	2266.086548	
YTDYNGPSIRIWFQTDDLK		45.91	2331.117157	
IYFVNSDNTIWQVCWDHGK	Carbamidomethyl (C14)	46.47	2381.08992	
IYAQKPDNTIQEYMWNGDGWK		53.45	2556.174332	
AYDPHKGWYPDLVTIFDRAPPR		74.68	2613.312836	
YTDYNGPSIRIWFQTDDLKLVQR		123.51	2827.429321	
GTITPVIQSEVAIISWGSFANNGPDLR		98.78	2898.487579	
VYTLTEGNTLQEFAYDSGTGWYNGGLGGAK		131.26	3168.467636	
HIRVYTLTEGNTLQEFAYDSGTGWYNGGLGGAK		98.97	3574.711716	
IYAQKPDNTIQEYMWNGDGWKEGTNLGGALPGTGIGATSF		87.22	4413.112427	
TGIAAVNSTNHLR		92.1	1352.716049	<p>LECF_AS PFM Fucose-specific lectin OS=<i>Neosartorya fumigata</i> OX=746128 GN=fleA PE=3 SV=1</p>
STPGAQQVLFRTGIAAVNSTNHLR	Acetyl (N-term)	106.9	2579.356827	
TGIAAVNSTNHLRVYFQDVYSIR		98.04	2680.372147	

Table 5.21. List of peptides identified by mass spectrometry for sample B47.

sequence	modifications	psm score	mass	accession number and description
WAVSFHNYTR		74.84	1366.64183	<p>A8NDX2_COPC7 Uncharacterized protein OS=<i>Coprinopsis cinerea</i> (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2</p>
DKVTTAYALPDNL		74.34	1419.724548	
LTLTWGGLNNLR		106.48	1469.835434	
SVIGAIPTENTGVTK		77.33	1485.803879	
AIFHTGSELFITR		80.3	1603.872223	
AIFHTGSELFITR	Acetyl (N-term)	48.45	1645.882782	
GPGKLLTWGGLNNLR		77.42	1809.026093	
AIFHTGSELFITRPGPK		53.69	1943.062881	
SVIGAIPTENTGVTKWAVSFHNYTR		89.63	2834.43515	
SWSSSTIHWGSSTVITEDVTSVVPNAVNR		90.53	3145.531662	

Table 5.22. List of peptides identified by mass spectrometry for sample B48.

sequence	modifications	psm score	mass	accession number and description
WAVSFSHNYTR		87.98	1366.64183	A8NDX2_COPC7 Uncharacterized protein OS= <i>Coprinopsis cinerea</i> (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) OX=240176 GN=CC1G_10558 PE=4 SV=2
DKVTTAYALPDNL		74.47	1419.724548	
LTLTGWGGLNNLR		94.78	1469.835434	
SVIGAIPTENTGVTK		69.31	1485.803879	
AIFHTGSELFITR		79.95	1603.872223	
AIFHTGSELFITR	Acetyl (N-term)	59.69	1645.882782	
GPGKLTLLTWGGLNNLR		83.02	1809.026093	
AIFHTGSELFITRGPVK		62.73	1943.062881	
AIFHTGSELFITRGPVK	Acetyl (N-term)	25.41	1985.073441	
FSFIWEGQGEACYQIGNGLTR	Carbamidomethyl (C12)	55.16	2432.121933	
SVIGAIPTENTGVTKWAVSFHNYTR		111.37	2834.43515	
SWSSSTIHWGSSTVITEDVTSVVPNAVNR		100.75	3145.531662	

Table 5.23. List of peptides identified by mass spectrometry for sample B53.

sequence	modifications	psm score	mass	accession number and description
FPTQGNLPK		44.32	1000.53418	A0A084FZZ4_PSEDA _PSEDA Uncharacterized protein OS= <i>Pseudallescheria apiosperma</i> OX=563466 GN=SAPIO_CDS8583 PE=3 SV=1
VEASAEQVDKSNCGI	Carbamidomethyl (C13)	75.72	1605.730438	
SMLAVSCSDGSNGLASR	Carbamidomethyl (C7)	129.7	1710.76651	
SMLAVSCSDGSNGLASR	Oxidation (M2); Carbamidomethyl (C7)	127.13	1726.761429	
FPYIGGVPAIGGWNSPNCGSCWK	Carbamidomethyl (C18, C21)	58.63	2523.146393	
SMLAVSCSDGSNGLASRFPTQGNLPK	Carbamidomethyl (C7)	92.32	2693.290131	
SMLAVSCSDGSNGLASRFPTQGNLPK	Oxidation (M2); Carbamidomethyl (C7)	108.87	2709.285049	
SIYILGIDHSTSFNIALHAMNDLTNGNAVGLGR	Oxidation (M20)	62.23	3499.751785	

6. GCFs; A NEW MINI GLYCANARRAY FOR IDENTIFICATION AND PURIFICATION OF LECTINS

6.1 Summary

Due to their biological relevance and possible applications, lectins have arisen the attention of the scientific community, either to explore their implementation as tools or to propose them as drug targets. Among the strategies for lectin purification, sugar-based affinity chromatography is by far the most widely used. However, this strategy has several disadvantages that slow down considerably the purification process, preventing the discovery of new lectins. For this reason, there is a growing interest in the development of non-column-based methodologies for lectin purification.

In this chapter, we report the development of a new tool that allows the identification and purification of unknown lectins directly from crude protein extracts. This tool assembles the fusion of the three classical steps of lectin purification : clarification, concentration and separation in a single operating unit. It also allows the simultaneous purification of lectins with different specificities in a quick and simple process. In addition, it can be used as a "mini glycan array" for the study of carbohydrate lectin interactions without the need for any labeling. Hence, we consider that this device might greatly contribute to the study of the lectinome.

6.2 Introduction

Lectins are powerful tools in the study of complex oligosaccharides and can be used for purification of industry-relevant glycans [150]. They also provide valuable insights for diagnostics and drug discovery. Therefore, it is expected that those proteins will occupy an important place in the biopharmaceutical industry in the coming decades [151]. Some of the most relevant biotechnology applications of lectins include: cell separation [152,153], mitogenic stimulation of immune cells [154], studies of glycoprotein biosynthesis, biomarkers [52], plant defense against pathogens and phytophagous insects [155,156], targeted drug delivery [52], monitoring alterations of normal and neoplastic cells [157], cancer drugs development [158], identification of blood groups and microorganisms [159,160], development of antiadhesive therapies against fungi, bacteria and virus [28,59–66], among others. For all these reasons, there is an emerging interest in the identification of new lectins and in the development of novel approaches to generate efficient alternatives for their purification.

6.2.2 Common strategies for lectins purification

Native lectins are purified from macerated tissues through a purification train that typically includes from three to five steps. Depending on the type of sample, it may be necessary to pre-treat the crude extracts, either by using ammonium sulphate or ethanol to precipitate proteins. Then, the process continues employing one or several chromatography steps based on gel-filtration, ion-exchange, hydrophobic interaction, etc. but affinity chromatography is, by far, the most commonly used. This latter one takes advantage of the carbohydrate-binding properties of lectins and use sugar-based matrices to purify them. Sugars are usually used for elution but EDTA can be also applied in the case of cation dependent binding [161,162]. The most popular matrices nowadays are glucose-based Sephadex™ and the galactose-based Sepharose™, however, the number of

commercial and homemade sugar-based resins is increasing every day. Table one collects some examples of matrices that have been used for purification of lectins.

Table 6.1. Carbohydrate-based matrix for lectins purification.

Matrix-coating	Purified lectins and references
Sugars	
Lactose	GLL[163], CTA[164], AKL[165], BiLec[166], TcL[167]; CRL[168], AKL[165], galatrox[169]; CNL[170], CRLII[171], BvCL[172], TcSL[173]
Sucrose and glucose	CNL[174]
Galactose	CVL[175], CTA[164], PSA[176], XEEL[161], CSA-II[177]
Mannose	Millectin[178], AFL[179], ATL[180], HTTL[181].
L-Rhamnose	TBL1, TBL2 and TBL3[182], SFL[183], ApJL[184]
Fucose	SauFBP32[185], TCL and CCL[186]
Melibiose	BI and BII from <i>P. scandens</i> [187], <i>B. simplicifolia</i> lectin[188]
Maltose	HTA-I and HTA-II[189]
N-acetylgalactosamine	CHA-II[190]; CAA-II[191], SBA[192]
N-acetylglucosamine	MBA[193], ADL[194], Bryohealin[195], FIA[196], <i>P. tuber-regium</i> lectin[197], WGA[192]
N,N'-diacetylchitobiose	CSA-I[177]; KL-15[198]
Polysaccharides	
Chitin	BEL[199]; PPL[200]; AaL[201], Jackin and frutackin[202].
Sephadex	PSL[203]
Sepharose	PPL[204]
Manann	MBL[205,206], ReMBL[207], CgMBP[208]; Millectin[178]; TFA[209]; pupal-MBL[210].
Alginate	Jacalin[211]
Chitosan	CTS-BL from <i>B. campestris</i> [212]
Glucan laminarin	b-1,3-glucan-specific lectin from <i>Blaberus discoidalis</i> [213]
Guar-gum	BmoLL[214], BIL[215]
Exopolysaccharides	<i>A. brasilense</i> Sp7 OMPs[216]
Galacto-mannan	Frutalin[217,218]
Glycoproteins	
Fetuin	CTA[164], MPL[219]; VL2, VL3, VL4[220]; OXYL[221] CTA[164]; FML[222]; LVL[223], PjLec[224].
Thyroglobulin	Isolectins of PHA; L4, L3 E1, L2 E2, L1 E3 and E4[225]
Mucin	SCL[226]; GCH[227]; MCL-4 and NnL[162]; HOL-I and HOL-II[228]; PPL[229], LCL[230]
Ovalbumin	DLL[231]
Gal β 1-4Fuc	NEX-1, NEX-2, NEX-3, LEC-1, LEC-2, LEC-4, LEC-6, LEC-9, LEC-10, and DC2.3a from <i>C. elegans</i> [232]
Glc1Man9GlcNAc2	AoClxA[233]
Asialofetuin	CTA[164], RVL[234], AVL[227]; CTA[164], RV[235].
Asialomucin	SBoL[236]

6.2.3 Drawbacks of sugar-based affinity chromatography

Although it is true that sugar-based columns are very useful and their variety is increasing every day, their current diversity does not meet the actual

requirements for the purification and/or identification of lectins that bind exclusively to complex glycans, such as in the case of PHA:L, for example. This generates a gap in the study and identification of an immeasurable number of unknown lectins that could have important applications or could be a key piece to elucidate cellular processes that remain unclear. Furthermore, this system requires hard maintenance when crude extracts are loaded into the columns because such samples contain pigments, oily components, proteolytic enzymes, and other complex substances that could damage the column and impair the purification [237].

One of the strategies that have been most widely used to avoid the drawbacks associated with the purification of native lectins is their heterologous production in genetically modified organisms (GMO) such as bacteria and yeast, among others. These recombinant proteins generally include fusion tag to facilitate their expression in soluble form (MBP, GST tags) and their purification (6-His tag). However, since it is necessary to know the sequence of the protein *a priori*, this technology is limited to lectins that have been previously identified. In exceptional cases, their sequences can also be tracked by data mining when the genome of the source organism is available. For all these reasons, there is a growing interest in the development of non-column-based methodologies for lectin purification, with a particular focus on those with the plasticity to adapt to specific requirements of particular cases.

Here, we report the development of a new tool that assembles the fusion of the three classic steps of lectin purification train: clarification, concentration and separation in a single operating unit. Furthermore, this tool allows the identification of unknown lectins from crude extracts and can also be used as a mini array to identify their carbohydrate specificity.

6.2.4 Brief description of the device

The device comprises the modification of borosilicate microfiber filters by coating their surfaces with a homogeneous layer of epoxides that can be covalently attached to any glycan, protein, antibody or compound with a nucleophilic nature. In this particular work, we have use glycosides attached to spacers with primary amines to guaranty their proper exposure, however, any sugar is susceptible to coupled. In the case of non-modified sugars, it is important to consider that, even when the alpha hydroxyl is the most susceptible to react, all the hydroxyls are susceptible to coupling.

The novelty of the device lies in the fact that, due to its porous nature, once the filter is modified it can be coupled to a vacuum filtration system to treat large volumes of samples. This allows lectins, even those found in very small concentrations, to be retained by interactions with the sugar-coated surface. Once the sample is eluted, the filter is rinsed and sectioned (in the case of multi-glycans filters). Then, bounded lectins can be eluted by two different ways:

- 1) **Competitive inhibition** where the filter is immersed in a solution containing the same sugar that is on the surface.
- 2) **Denaturation** where the filter is immersed in PBS and heated to later proceed to SDS-PAGE gel electrophoresis.

Elution by competitive inhibition allows recovering of native lectins without affecting their ability to bind to glycans, making it very useful in the cases where future activity studies are required. The elution by denaturation is exclusively used for the identification of lectins (Figure 6.1). In this case, the lectin profile obtained from each organism will contribute to lay the foundations for the study of its lectinome. Finally and most important of all, we have discovered that a simple protein staining can be applied to identify the area of the surface where lectins are bound. This allows to identify carbohydrate specificity in a similar

way to other glycan arrays but without the need for labeling nor specialized equipment. This makes it an accessible strategy for any laboratory.

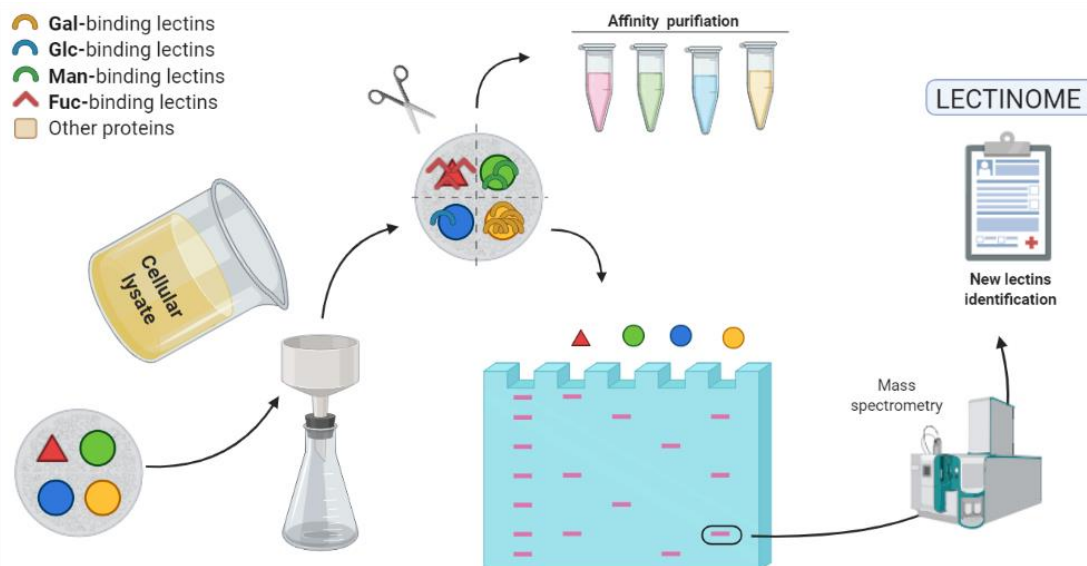


Figure 6.1: Schematic representation of the device. The red triangles represent fucose; green circles mannose; blue circles glucose and yellow circles galactose. Figure created using BioRender.com.

6.3 Material and methods

6.3.1 Synthesis of 2'-Chloroethyl 2,3,4-Tri-O-acetyl- α -L-fucopyranoside

A mixture of L-fucose (1 g), 2-chloroethoxyethanol (7.5 mL), and Amberlite IR-120 Resin (H+) (0.6 g) was stirred under reflux at 80 °C for 5 h. Then, the mixture was cooled to room temperature (RT) and the resin was removed by filtration. The excess of solvent was evaporated at low pressure and residues were dissolved in pyridine (12 mL) and placed in a cool bath with stirring, while acetic anhydride was periodically added (6 mL). The ice bath was removed and the reaction proceeded at RT overnight. After the reaction was complete, a liquid extraction was performed with water and ethyl acetate. The organic layer was dried by adding sodium sulfate, filtered and evaporated. Then, compounds were purified by flash column chromatography on silica gel using EtOAc-Petroleum ether (8:2, v/v) as mobile phase to obtain separately the α and β conformations of

the per-O-acetylated fucoside, which was obtained in a $\alpha:\beta$ ratio of approximately 7:3.

6.3.2 Synthesis of 2-(2-chloroethoxy)ethylper-O-acetylated glucosides

Mixtures containing 2-(2-chloroethoxy)ethanol (7.5 mL), Amberlite IR-120 Resin (H⁺) (0.6 g) and 1 gr of the respective glycoside (D-mannose, D-glucose and D-galactose) were stirred at 80 °C for 5 h. The mixtures were cooled to RT and filtrated. Then, solvent was evaporated and glycosides were purified by flash column chromatography on silica gel using DCM:MeOH (8:2, v/v). Acetylation was carried out as described previously for the fucoside, followed by liquid extraction (also described previously). Per-O-acetylated glycosides were purified by flash column chromatography on silica gel using EtOAc-Petroleum ether (8:2, v/v).

6.3.3 General procedures for azidation of glycosides

A suspension of the respective glycoside (0.5 mM, 1 eq), TBAI (0.2 eq), and sodium azide (5 eq) in N,N-dimethylformamide was stirred at 90 °C for 18 h. The N,Ndimethylformamide was evaporated at reduced pressure and the residue was partitioned between dichloromethane (150 mL) and water (30 mL). The organic layer was separated, washed with brine (3X 30 mL), dried over sodium sulfate, filtered and evaporated.

6.3.4 General procedures for deprotection of glycosides

The per-O-acetylated glycosides were treated with a catalytic amount of MeONa in MeOH (10 mL) for 30 min and then the reaction was neutralized with Amberlite IR-120 Resin (H⁺). The resin was removed by filtration and the solvent was evaporated.

6.3.5 General procedures for the reduction of azido glycosides

Solutions of the respective 2-(2-aminoethoxy)ethyl galactoside, 2-(2-aminoethoxy)ethyl mannoside, 2-(2-aminoethoxy)ethyl glucoside and 2'-

azidoethyl fucoside were prepared in methanol at 1 mM. Then, palladium (Pd) was added in catalytic amount and solutions were stirred at room temperature under hydrogen atmosphere for 12-18 h, until the disappearance of the starting material. Mixtures were filtered through a pad of celite, solids were washed with methanol and the filtrates were evaporated at a reduced pressure.

6.3.6 Hydroxylation of borosilicate fibers

The borosilicate microfiber filters were rinsed with acetone for 5 min in orbital agitation to remove any adsorbed chemical species from the surface. Then, filters were allowed to dry at RT until complete evaporation of acetone, since traces of organic solvent react explosively during the next step. Once acetone was completely evaporated, filters were immersed into sulphuric acid and hydrogen peroxide (3:1 ratio) for 10 s. After, filters were immediately submerged in toluene (99.8%, v/v) for 15 seconds and rinsed twice with sterile PBS buffer pH 7.4. Finally, the filters were dry at 110° C for 45 minutes.

6.3.7 Silane coupling (Self-assembled monolayer)

Filters were covered with 4% GPS ((3-glycidyoxypropyl)trimethoxysilane) dissolved in dry toluene and placed in a nitrogen environment in absence of light. This allows a homogeneous coupling of the silane with the hydroxides previously generated on the borosilicate microfibers and prevents self-polymerization. After 16 h, the GPS-coated filters were removed from this solution, rinsed with ethanol and were allowed to dry at RT.

6.3.8 Monodeposition of glycosides on GPS-coated filters

Activated filters were placed in individual containers and covered by a 5 mM solution of the respective glycoside (**1**, **2**, **3** and **4**) dissolved in sterile PBS pH 7.4. Containers were closed and incubated at slow orbital agitation at room temperature for 2 h. Afterwards, the substrate was rinsed several times with PBS buffer to remove any unbound compound from the surface. Finally, the

glycoside-coated filters (GCFs) were incubated 2 h in a solution containing 50 mM of urea and 50 mM of glycine to open remaining epoxides.

6.3.9 Multi-spot deposition of glycosides on GPS-coated filters

Compounds **1**, **2**, **3** and **4** were dissolved at 50 mM in MeOH and a 2 µl of each solution were pipetted onto different areas of the GPS-coated filters surface. Spots were allowed to dry at RT and the unbound glycosides were removed by eluting 100 mL of PBS buffer through the filters by applying vacuum into a Buchner funnel. Finally, filters were incubated 2 h in a solution containing 50 mM of urea and 50 mM of glycine to open remaining epoxides.

6.3.10 Lectin binding to glycoside-coated filters

Solutions containing 200 µg/mL of purified lectins that have affinity for the prepared glycosides. Each sample also contained 200 µg/mL BSA as a negative control for testing protein adsorption. The GCFs were placed in a Buchner funnel, incubated and 10 mL of the respective proteins cocktail was eluted applying vacuum. To visualize proteins attached to the surface InstantBlue^R protein stain was used.

6.3.11 Protein denaturing electrophoresis (SDS-PAGE)

All SDS-PAGE electrophoresis were performed on 15% polyacrylamide gels with a current of 180 mV in a Mini-Protean Tetra Electrophoresis Cell kit from Bio-Rad (Hercules, CA, EU.), using a molecular weight marker wide range. To visualize the result of the runs, the gels were stained with InstantBlue^R 10 min.

6.4 Results and discussion

6.4.1 Glycosides synthesis

The derivatized monosaccharides **1**, **2**, **3** and **4** were synthesized through a sequence of reactions that involves attaching a chloro-alcohol to the anomeric position of the sugars, followed by nucleophilic substitution of chlorine for an

azido group (N_3) and its subsequent reduction to a primary amine. Intermediate steps of protection and deprotection were performed to facilitate purification of the intermediate products by flash column chromatography and to avoid unspecific reactions with the remaining OHs of the ring (Figure 6.2).

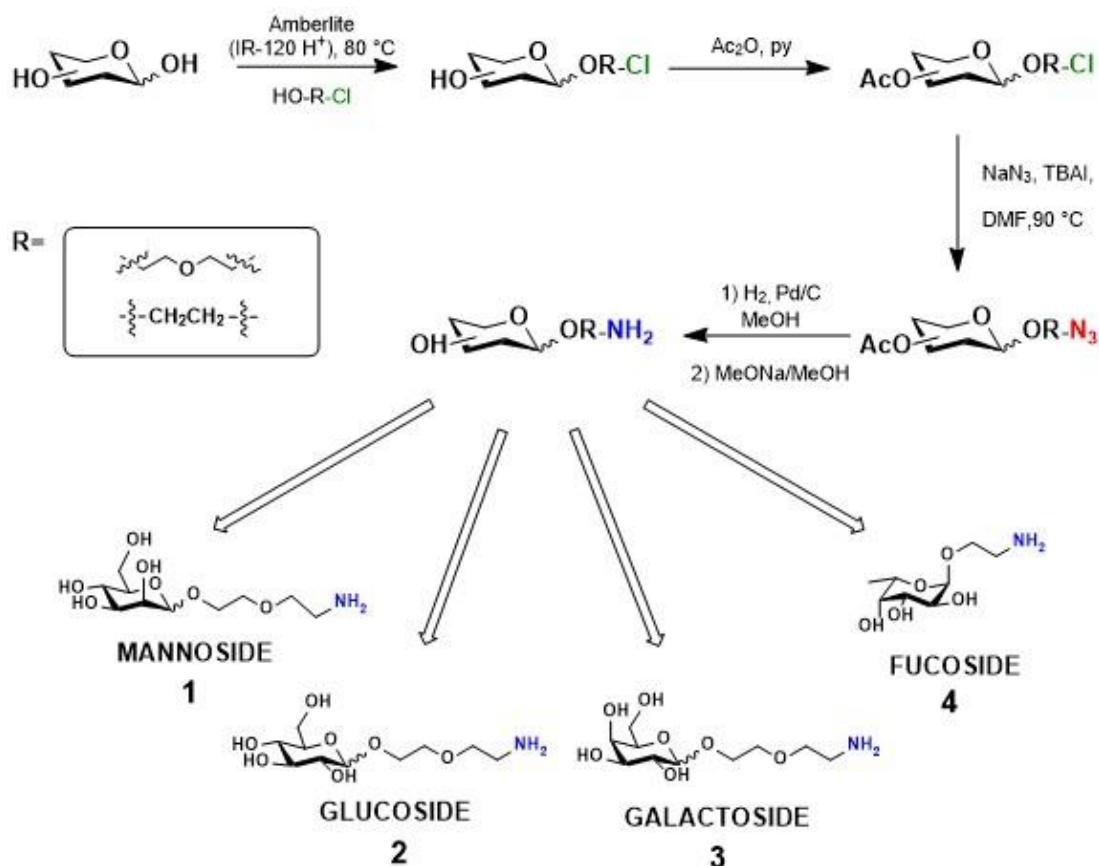


Figure 6.2: Glycosides synthesis. Schematic representation of the synthetic pathway of 2-(2-aminoethoxy)ethyl mannoside (1), 2-(2-aminoethoxy)ethyl glucoside (2), 2-(2-aminoethoxy)ethyl galactoside (3) and 2'-azidoethyl fucoside (4). Figure created using ChemDraw Version 15.

6.4.2 Borosilicate surface modification

Borosilicate (ZnO) microfiber filters were hydroxylated by immersion in piranha solution. Then, a GPS-Self Assembled Monolayer (GPS-SAM) was achieved by liquid deposition under N₂ atmosphere. This reaction leaves available epoxide residue to interact with the amino group of the previously prepared glycosides to covalently attach them to the surface (Figure 6.3). Glycosides-coated filters (GCFs) with compounds **1**, **2**, **3** and **4** were prepared to generate mannose-coated

Filters (Man-CF), glucose-coated filters (Glu-CF), galactose-coated filters (Gal-CF) and fucose coated filters (Fuc-CF), respectively.

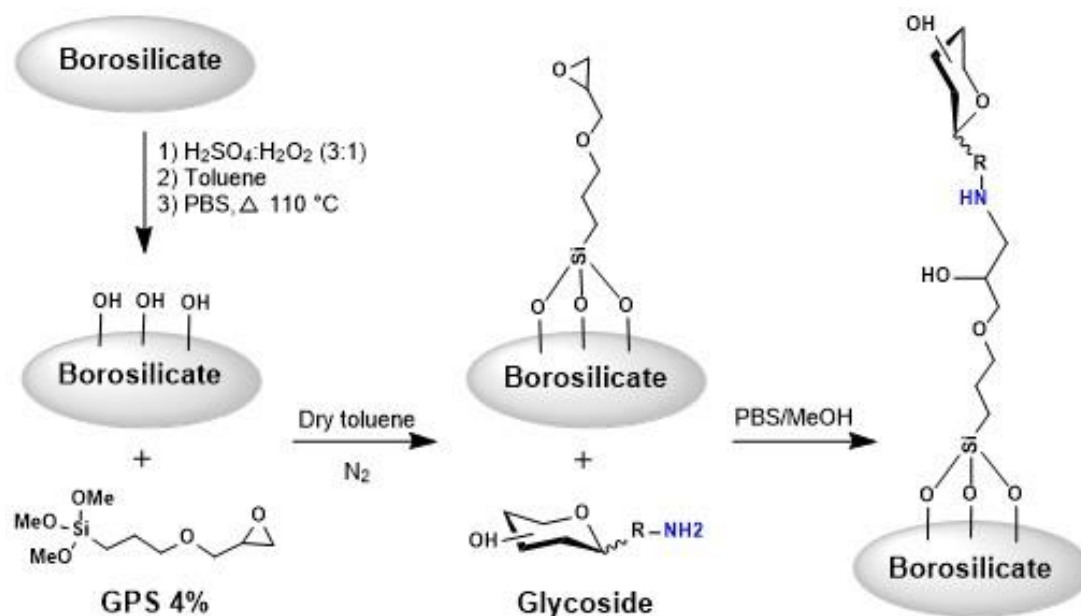


Figure 6.3: Borosilicate surface functionalization and glycosides coating. Schematic representation of the synthetic pathway to covalently attach glycosides to the borosilicate microfiber filters. Figure created using ChemDraw Version 15.

6.4.3 Specific interaction between lectins and glycosides-coated filters

To determine if lectins can bind to the carbohydrates on borosilicate surfaces, the previously coated filters (Man-CF, Glc-CF, Gal-CF and Fuc-CF) were incubated in a PBS solution containing a lectin with known affinity for the respective glycoside. Thus, FleA from *A. fumigatus* [40] was used for Fuc-CF, LecA from *Pseudomonas aeruginosa* [238] for Gal-CF and due to its dual recognition for glucose and mannose, ConA from *Canavalia ensiformis* [239] was used for Man-CF and Glc-CF. Each solution also contained Bovine Serum Albumin (BSA) as a negative control to rule out protein adsorption. Then, filters were rinsed, immersed in 1 mL of PBS and heated at 100 °C for 10 minutes. The presence of proteins in all fractions was verified by SDS-Page, though it was found that only the respective lectin was attached to the filters after the washing steps (Figure 6.5A-D). This result indicates that the interaction lectin-sugar is taking place and

confirms that non-interacting proteins can be removed through simple washing steps. It also evidences that the rest of the matrix does not display unspecific interactions with common proteins.

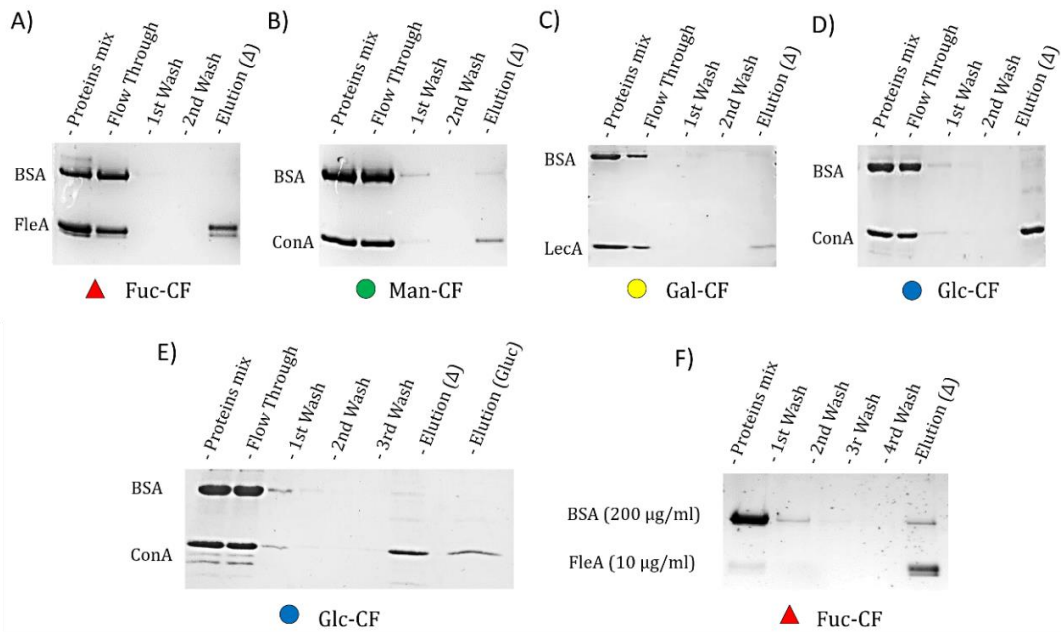


Figure 6.4: Electrophoretic profile of fractions recovered from the GCFs after protein incubation. Well 1, freshly prepared proteins mix containing 200 μg/mL of lectin and BSA; well 2, unbound proteins after 1 h incubation; wells 3 and 4 washing steps with PBS; well 4, proteins recovered from the filter after heat treatment at 100 °C. Fractions recovered from the fucose-coated filter (A), from the mannose-coated filter (B), from the galactose-coated filter (C) and from the glucose-coated filter (D). E) Fractions recovered from the glucose-coated filter with two elution methods. Well 1, freshly prepared proteins mix at 200 mg/mL; well 2, proteins mix recovered from the filter after 1 h incubation; wells 3 to 5 washing steps with PBS; well 5, proteins recovered after heat treatment at 100 °C; well 6 proteins eluted from the filter with 50 mM glucose. F) Lectins concentration. Well 1, freshly prepared proteins mix containing (10 μg/ml of FleA and 200 μg/ml of BSA); wells 2 to 5 washing steps with PBS; well 6, proteins recovered by heat treatment at 100 °C.

Once the interaction between lectins and our glycosides coated filters (GCFs) was established, the next question to address was if the proteins could be recovered from the surface. Therefore, a preliminary study was conducted with Glc-CF and ConA where, the filter was split into two equivalent portions after washing and one of them was incubated in PBS containing 50 mM glucose, instead of heating. ConA was recovered when glucose was added to the buffer as expected (Figure

6.4E). In a later study, a Fuc-CF was incubated with a solution of PBS containing FleA at 10 $\mu\text{g/mL}$ and BSA at 200 mg/mL . After elution with fucose, the electrophoretic profile shows that FleA was retained on the filter and that its concentration in the elution fraction was considerable higher than in the original solution. This confirms that the filters can bind lectins, even when they are in a very low concentration with respect to other proteins in the sample and establishes the filter utility as concentration tools (Figure 6.4 F).

6.4.4 Mini chip for detection of protein-glycan interactions on unlabeled samples

One of the main characteristics of glass microfiber filters is that they are inert, do not retain small molecules and do not establish electrostatic interactions. So, theoretically, they should not interact with the anionic dye Coomassie R-250, present in the InstantBlue solution. If this was correct and we manage to find the right conditions to deposit several glycosides on different areas of the same filter, it would mean that the lectins attached to them could be *in situ* stained. Thus, generating a system for detection of lectin-carbohydrate interactions, equivalent to the glycan arrays but without the need for any labelling. From a practical point of view, this strategy has many advantages compared to conventional arrays, not only because it can be performed in unlabeled samples, but also, it is simple and can be carried out in any laboratory with minimal requirement of equipment.

To confirm the absence of interactions between filter and colorant, we dipped untreated filters in the InstantBlue solution during 1 h and rinsed them in water. They were still completely white indicating that no dye particles remained on the surface after treatment (data not shown). Then, we proceeded to find the ideal conditions to achieve the localized deposition of glycosides on the surface of the filters (data not shown). In the final experiment, we use a pencil to draw a circle in both, a GPS-coated filter and an untreated filter. Then 2 μl of mannoside (1) (50 mM in MeOH) were pipetted inside of each circle. Solutions were allowed to

dry and the excess of glycoside was removed with PBS. The filters were incubated 2 h in a solution containing urea and glycine to deactivate unreactive GPS (from the surface that was not in contact with the mannoside solution). Each filter was latterly incubated with ConA for 1 hour, followed by washing steps and staining with InstantBlue solution. After 15 minutes, the Man-CF displayed an intense blue coloration on the surface where mannoside had been spotted. The back also showed an intense coloration in this area, indicating that when the glycoside was deposited on the filter surface, it penetrated to deeper layers and was homogeneously attached throughout the thickness of the filter. In the case of the untreated filter, no deposition of the dye was observed, either in the front nor the backside (Figure 6.5A). The slight coloration on the rest of the Man-CF may be attributed to the interaction of the colorant with the free carboxyl group of glycine used to deactivate unreactive epoxides. Now, we are working on new protocols to deactivate the remaining epoxides through the use of other bases such as ethanolamine. This experiment evidences that the glycosides deposition can be guided and confirms that GCFs could be used as mini arrays for the study of carbohydrate-protein interactions.

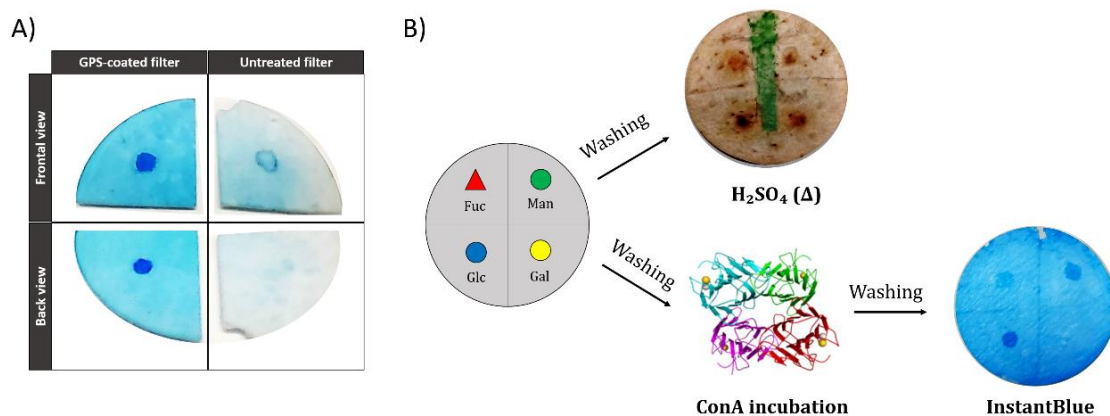


Figure 6.5: GCFs mini chips for the detection of carbohydrate-lectin interactions. **A)** *In situ* labeling of ConA attached to Man-CF. Untreated filter spotted with mannoside (1) was used as a negative control. **B)** GCFs Miniarrays printing scheme and stainings. In the upper part is shown the filter stained with H_2SO_4 after glycosides deposition and in the lower part the filter stained with InstantBlue after incubation with ConA.

With the evidence provided by the previous experiments, we decided to make our first mini array prototype. For this, we attach the glycosides **1**, **2**, **3** and **4** to different quadrants of GPS-coated filters. This was done by duplicate and after the glycosides were attached, one of the filters was immersed in 4% sulphuric acid (H₂SO₄) and dry with a heat gun to reveal the presence of sugar spots on the surface. The other filter was incubated in a PBS solution containing 200 µg/mL of ConA.

Sulphuric acid staining confirmed that all glycosides were properly attached to the surface and the protein staining revealed that, as expected, ConA was mostly linked to the glucose/mannose-coated spots (Figure 6.5B). A slight change in coloration was also detected in the fucose-coated spot suggesting a possible interaction with this sugar. Further experiments will address the study of this interaction by ITC. On the other hand, no signal was detected at all on the galactose-spotted quadrant, despite the fact that its successful deposition had been previously confirmed. This result evidences that the differential binding of ConA to the surface is ruled by the sugar in the respective area and sets the GCFs as useful mini arrays for unlabeled samples.

6.4.5 Recovery of lectins from complex protein mixtures

The next step in our research was to validate if the GCFs would keep their functionality when complex mixtures of proteins are applied. For this, we expressed recombinant FleA in *E. coli* and treated the extract containing soluble intracellular proteins with a Fuc-CF. On the resulting electrophoretic profile, most of the protein bound to the filter corresponds to FleA, however, it is still necessary to optimize the washing steps to get rid of the contaminant proteins observed on the gel (Figure 6.6). The possibility that those other proteins correspond to endogenous fucose binding lectins from *E. coli* is not ruled out and it will be addressed in the near future.

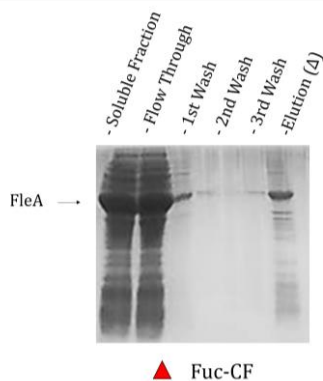


Figure 6.6. Electrophoretic profile of fractions recovered from the GCFs after incubation with total proteins extracts. Well 1, soluble extract; well 2, unbound proteins after 1 h incubation; wells 3 and 5 washing steps with PBS; well 4, proteins eluted by heat treatment at 100 °C.

6.5 Conclusions

The GCFs represent an interesting multipurpose tool that can be used both, as a lectin purification system for identification purposes and as a device to study carbohydrate-protein interactions.

Some of the advantages of GCFs as a purification tool are that: i) they avoid the inconveniences associated with prepackaging, maintenance and cleaning of columns; ii) the recovery of surface-bound lectins is easily accessible due to the flexible nature and easy handling of the filters; iii) the strategy allows the process of lectin purification quickly and efficiently since it merges several steps of the classical train of lectin purification and iv) the use of this tool is easy and cost-effective. The most significant expenses are given by the cost of the sugars attached to the surface.

On the other hand, concerning its potential use as a glycan matrix, the device has the advantages that i) no prior purification and labeling of proteins is required; ii) identification of interactions *in situ* is simple and iii) no specialized equipment is required for data acquisition nor manipulation.

Finally, since any nucleophilic molecule can bind to the surface (proteins, nucleic acids, natural glycans, antibodies, etc.), the adaptation of this methodology could

be used for enzyme activity/inhibition assays, antibody screening, interactions of whole cells with surfaces, enzyme activities, DNA / RNA interactions and non-carbohydrate molecule binding assays. Therefore, we consider that this device could contribute greatly not only to the study of the lectinome of species, but also that our strategy could be used beyond the glycobiology approach and applied in many other areas of science.

6.6 Contributions

Dania Martínez Alarcón (DMA) performed the synthesis of glycosidic derivatives, the borosilicate surface modification, prepared the Glycoside Coated Filters (GCFs) and performed protein binding experiments. DMA prepared all the figures, discussed results and conceived the design of the study. Roland J. Pieters (RJP) and Annabelle Varrot (AV) administered the project, obtained funding, contributed to data analysis, and supervised DMA.

7. GENERAL CONSLUSIONS AND PERSPECTIVES

The emergence of new pathogenic species represents a global health problem that claims millions of lives each year. Furthermore, the limited susceptibility of these emerging pathogens to conventional drugs and the rapid generation of resistance are urgent health problems. According to the Antimicrobial Resistance Review "By 2050, 10 million deaths per year are expected to be attributable to microbial resistance." Therefore, the World Health Organization (WHO) has declared antimicrobial resistance to be one of the top 10 public health threats facing humanity and has established the fight against antimicrobial resistance (AMR) as a priority worldwide.

In this context, the European collaborative network PhD4GlycoDrug has emerged. The main objective of this program is the discovery and development of new Glyco-Drugs as a therapeutic alternative to evade the selective pressure that leads to AMR. For this, 12 independent projects have been established; each project seeks to contribute to the development of antiadhesive therapies against different human pathogens from varied perspectives; from exploration to evaluation. For pathogens whose virulence factors are known, the program aims to develop glycomimetics to antagonize microbial lectins involved in host binding. However, in the case of emerging pathogens, such as *S. apiospermum*, there is very limited information about its anchoring mechanism. Hence, it is first necessary to identify the virulence mechanisms involved in carbohydrate-lectin interactions during host recognition.

This project arises from the urgent need to generate information to guide the development of antifungals against the emerging pathogenic fungus *Scedosporium apiospermum*. The main objective of the thesis was the identification and characterization of its lectins, as a strategy to contribute to the general

understanding of its host-pathogen interactions during the early stages of infection. For this, we have developed 4 independent but related projects:

Project 1- Identification and characterization of SapL1; a new target for the development of antiadhesive glycodrugs from *S. apiospermum*.

In this study, we present the identification, production, and characterization of a new carbohydrate-binding protein from *S. apiospermum*. To the best of our knowledge, this is the first studied lectin for this opportunistic pathogen. SapL1 is homologous to the conidial surface lectin FleA from *A. fumigatus*, known to be involved in adhesion to the host glycoconjugates present in mucins and human lung epithelium.

Here, we demonstrated that SapL1 is strictly specific for fucosylated carbohydrates and recognized all blood group types. These results are particularly interesting since it has been shown that there is six times more content of fucose α 1,3/4 linked to glycoproteins in the CF airways, where *Scedosporium* colonization is more frequent. SapL1 adopts the 6-bladed β -propeller fold and forms a dimer with 6 carbohydrate-binding pockets in each face. Within our analysis, we have found that these binding pockets are non-equivalent, but they all share the features necessary for fucose recognition. The specificity and affinity of each one of the six binding pockets of SapL1 have been deeply analyzed to highlight the features that must be explored for the design of efficient inhibitors.

The detailed information exposed here places SapL1 as a promising target for the treatment of *Scedosporium* infections and will be of great value to guide the development of antiadhesive glycodrugs against this pathogen.

Project 2- Biochemical and structural characterization of a Cyanovirin-like lectin from the human opportunistic fungi *S. apiospermum*

In this study, we report the identification of a Cyanovirin-like lectin from *Scedosporium apiospermum* (SapL6). SapL6 belongs to the fungal Cyanovirin–N homolog (CVNH) subfamily III. Although the specificity and biological function of this group of proteins are not clear, it is believed that they play a fundamental role in fungal life. Consequently, they are considered potential targets for the development of new antifungals.

SapL6 was successfully produced and characterized, both biochemically and structurally, however, despite our efforts it was not possible to identify any carbohydrate-binding activity. Nevertheless, we believe that the identification of new CVNH such as SapL6 will contribute to the general understanding of this family of proteins. Furthermore, the structural information gathered here will be useful to address possible interactions with glycans through *in silico* analysis and might guide towards the discovery of its natural ligand. Which, we consider, is the ultimate clue to finally understanding the biological role of CVNHs in fungi and to validate their potential uses as drug targets.

Project 3- Identification of new native lectins from protein extracts of medical relevant fungal strains

The main aim of this project was to identify new native lectins from the human opportunistic fungi *S. apiospermum* and *A. fumigatus*. The analysis of their intracellular proteins and secretomes led us to the identification of 8 different proteins. Five of them were found in the protein extracts of *S. apiospermum* and 4 were identified from *A. fumigatus* extracts. The preliminary analysis of those proteins revealed that some might be attractive pharmacological targets and others could find application as biotechnological tools.

SapCP: This protein was the most abundant protein secreted to the culture media of *S. apiospermum*. SapCP belongs to the family of cerato-platanin proteins (CPPs) whose members bind to multiple polysaccharides and are essential for the fungal life cycle. Some of them also participate in the colonization and host recognition process, while others have found biotechnological applications as allergens or as antigens for detection of specific fungal pathogens.

PVL-like lectins: two GlcNAc-binding lectins were identified from the intracellular fraction of proteins of *S. apiospermum* and *A. fumigatus*. Both proteins display high similarity with the protein PVL from the saprophytic fungi *Psathyrella velutina*, which is a GlcNAc-binding lectin involved in the colonization of soils and host recognition. PVL-homologs are widely distributed in some branches of the fungal kingdom but it has never been identified in human pathogens. Therefore, in case of being confirmed, this finding might have important implications for the subsequent study of infections caused by those two microfungi.

FleA: The fucose binding lectin FleA from *A. fumigatus* was unambiguously identified in both, the intracellular and extracellular fraction of proteins of *Aspergillus*. FleA is a lectin found on the surface of the *Aspergillus* conidia and shows a strong pro-inflammatory effect on human bronchial cells. This protein displays affinity by fucosylated carbohydrates and recognizes human blood group antigens (BGA). Therefore, it has been suggested that FleA mediates the binding of *A. fumigatus* to airways in a fucose-dependent manner.

AfFG-GAP: The mannose/galactose binding protein AfFG-GAP was highly secreted to the extracellular media by *A. fumigatus*. The overall fold this protein is a β -propeller composed of two FG-GAP repeats. This domain is commonly found in the N-terminal region of the α -chains of integrins, a region shown to be important for ligand binding and protein-protein interactions with some components of the

extracellular matrix. Hence, it has been proposed that AFG-GAP could be involved in cell adhesion and/or the transduction of signals for immune recognition of *A. fumigatus*.

CcMBL-like proteins: we identify six samples that display high similarity to a hypothetical mucin-binding lectin from the mushroom *Coprinopsis cinerea* (CcMBL). CcMBL-like proteins were recovered from both, fucose and galactose affinity chromatographies. This fact supports the mucin-binding activity predicted for CcMBL since it has been shown that mucins of cystic fibrosis lungs (CF) have an increased content of those monosaccharides. Therefore, considering the well-known role that aberrant glycosylation of mucins displays during fungal pathogenesis in CF patients, these proteins might represent interesting drug targets.

SapFKBP12: SapFKBP12 belongs to the family of cytoplasmic immunophilins FKBP12. These proteins are highly conserved among fungal pathogens and can interfere with the transcription of virulence factors and genes involved in hyphal growth. FKBP12 proteins own their name to their capacity to bind the immunosuppressant FK506. This drug has shown to be very effective against deadly fungal pathogens but is still contraindicated in systemic infections due to its immunosuppressant side effects. Hence, the synthesis of FK506 analogous with greater specificity is needed and the discovery of FKBP12s from emerging fungal pathogens would significantly aid their development. On the other hand, we have shown that residues critical for FK506 binding are well conserved between SapFKBP12 and FKBP12 from other FK506-susceptible fungal pathogens, such as *A. fumigatus* and *Candida albicans*. This information suggests that *Scedosporium* might also be susceptible to this drug and future clinical approaches must explore its use as a treatment in non-systemic infections, such as the increasingly common topical scedosporiosis. The identification of

SapFKBP12 and its future characterization will greatly contribute to explore new treatments not only against *Scedosporium* infections but also contribute to the development of a broad-spectrum antifungal therapy.

Project 4- GCFs; a new mini glycan array for identification and purification of lectins

The challenges we faced during the development of the first 3 projects of this thesis led us to raise the need to generate an efficient tool for the identification of new lectins. The original idea was to generate a device that would allow the simultaneous purification of lectins with different specificities, exclusively for identification purposes. However, as the project progressed, we realized that the tool we had created (GCFs) would not only allow the purification of multiple lectins but also facilitate the process and, additionally, could be used as a “miniglycan matrix”. Hence, the GCFs represent an interesting multipurpose tool that can be used both, as a lectin purification system and to study carbohydrate lectin interactions.

Concerning its potential use as a glycan matrix, the device has the advantage that no prior protein labeling is required. Furthermore, the identification of binding interactions is simple and does not require specialized equipment for data acquisition nor manipulation. On the other hand, as a purification tool, the most outstanding advantage of GCFs is that they allow the process to be carried out in a fast and efficient way since it merges several steps of the classic lectin purification train. This might impulse the characterization of lectinomes in the coming decades. But farther than that, this methodology can be adapted to study any other interactions *e.g.*, enzyme activity/inhibition assays, antibody screening, interactions of whole cells with surfaces, enzyme activities, DNA / RNA, etc. Hence, we consider our strategy could be used beyond the glycobiology approach and to be applied in many other areas of science.

Overall, this thesis represents a general strategy to address the study *S. apiospermum* lectinome. Our findings have revealed the first insights about the recognition of human glycoconjugates by lectins from this microfungus and contribute to the general understanding of the host-binding process during the early stages of infection. In addition, we have developed a new tool that could mean a milestone in glycobiology and exponentially boost the discovery of new lectins in the coming years.

Finally, we expect that the information gathered here will contribute to design targeted therapies against this pathogen and will encourage the study and characterization of the new proteins discovered, in the search for new biomedical applications and drug targets.

REFERENCES

1. Thornton, C.R. Detection of the 'Big Five' mold killers of humans: *Aspergillus*, *Fusarium*, *Lomentospora*, *Scedosporium* and Mucormycete. *Adv. Appl. Microbiol.* **2020**, *110*, 1–61.
2. Ellis, D.; Marriott, D.; Hajjeh, R.A.; Warnock, D.; Meyer, W.; Barton, R. Epidemiology: Surveillance of fungal infections. *Med. Mycol.* **2000**, *38*, 173–182.
3. Garnica, M.; Nucci, M. Infecciones fúngicas emergentes. *Ann. Nestlé* **2010**, *67*, 135–142.
4. Tuite, N.L.; Lacey, K. Overview of invasive fungal infections. In *Fungal Diagnostics*; O'Connor, L., Glynn, B., Eds.; Humana Press, Totowa, NJ: NY, **2013**; Vol. 968, pp. 1–23 ISBN 9781627032575.
5. Ramirez-Garcia, A.; Pellon, A.; Rementeria, A.; Buldain, I.; Barreto-Bergter, E.; Rollin-Pinheiro, R.; De Meirelles, J.V.; Xisto, M.I.D.S.; Ranque, S.; Havlicek, V.; *et al.* *Scedosporium* and *Lomentospora*: An updated overview of underrated opportunists. *Med. Mycol.* **2018**, *56*, 102–125.
6. Gilgado, F.; Cano, J.; Gené, J.; Guarro, J. Molecular phylogeny of the *Pseudallescheria boydii* species complex: proposal of two new species. *J. Clin. Microbiol.* **2005**, *43*, 4930–4942.
7. Duvaux, L.; Shiller, J.; Vandeputte, P.; Bernonville, T.D. De; Thornton, C.; Papon, N.; Le Cam, B.; Bouchara, J.-P.; Gastebois, A. Draft genome sequence of the human-pathogenic fungus *Scedosporium boydii*. *Genome Announc.* **2017**, *5*, e00871-17.
8. Pérez-Bercoff, Å.; Papanicolaou, A.; Ramsperger, M.; Kaur, J.; Patel, H.R.; Harun, A.; Duan, S.Y.; Elbourne, L.; Bouchara, J.P.; Paulsen, I.T.; *et al.* Draft genome of Australian environmental strain WM 09.24 of the opportunistic human pathogen *Scedosporium aurantiacum*. *Genome Announc.* **2015**, *3*,

- e01526-14.
9. Vandeputte, P.; Ghamrawi, S.; Rechenmann, M.; Iltis, A.; Giraud, S.; Fleury, M.; Thornton, C.; Delhaes, L.; Meyer, W.; Papon, N.; *et al.* Draft genome sequence of the pathogenic fungus *Scedosporium apiospermum*. *Genome Announc.* **2014**, *2*, e00988-14.
 10. Luplertlop, N. *Pseudallescheria/Scedosporium* complex species: From saprobic to pathogenic fungus. *J. Mycol. Med.* **2018**, *28*, 249–256.
 11. Cortez, K.J.; Roilides, E.; Quiroz-telles, F.; Meletiadis, J.; Antachopoulos, C.; Knudsen, T.; Buchanan, W.; Milanovich, J.; Sutton, D.A.; Fothergill, A.; *et al.* Infections Caused by *Scedosporium* spp. *Clin. Microbiol. Rev.* **2008**, *21*, 157–197.
 12. Pereira De Mello, T.; Aor, A.C.; Santiago, S.; De Oliveira, C.; Branquinha, M.H.; Luis, A. Conidial germination in *Scedosporium apiospermum*, *S. aurantiacum*, *S. minutisporum* and *Lomentospora prolificans*: influence of growth conditions and antifungal susceptibility profiles. *Mem Inst Oswaldo Cruz* **2016**, *111*, 484–494.
 13. Goldman, C.; Akiyama, M.J.; Torres, J.; Louie, E.; Meehan, S.A. *Scedosporium apiospermum* infections and the role of combination antifungal therapy and GM-CSF: a case report and review of the literature. *Med. Mycol. Case Rep.* **2016**, *11*, 40–43.
 14. Santiago Reis, C.; Reis-filho, E.G.D.M. Mycetomas: an epidemiological, etiological, clinical, laboratory and therapeutic review. *An Bras Dermatol* **2018**, *93*, 8–18.
 15. Branscomb, R. Mycetoma: An Overview. *Lab. Med.* **2003**, *34*, 803–808.
 16. Borghi, E.; Iatta, R.; Manca, A.; Montagna, M.T.; Morace, G. Chronic airway colonization by *Scedosporium apiospermum* with a fatal outcome in a patient with cystic fibrosis. *Med. Mycol.* **2010**, *48*, 108–113.
 17. Badiie, P.; Hashemizadeh, Z. Opportunistic invasive fungal infections:

- Diagnosis & clinical management. *Indian J. Med. Res.* **2014**, *139*, 195–204.
18. Ellis, S.; Ong, E. Disseminated infections: A clinical overview. *Mol. Med. Microbiol. Second Ed.* **2014**, *2–3*, 637–653.
 19. Bouchara, J.P.; Papon, N. *Scedosporium apiospermum*. *Trends Microbiol.* **2019**, *27*, 1045–1046.
 20. Rammaert, B.; Puyade, M.; Cornely, O.A.; Seidel, D.; Grossi, P.; Husain, S.; Picard, C.; Lass-Flörl, C.; Manuel, O.; Le Pavec, J.; *et al.* Perspectives on *Scedosporium* species and *Lomentospora prolificans* in lung transplantation: results of an international practice survey from ESCMID fungal infection study group and study group for infections in compromised hosts, and European confederation of medical mycology. *Transpl. Infect. Dis.* **2019**, *21*, 1–8.
 21. Turcios, N.L. Cystic fibrosis lung disease: An overview. *Respir. Care* **2020**, *65*, 233–251.
 22. Scanlin, T.F.; Glick, M.C. Terminal glycosylation in cystic fibrosis. *Biochim. Biophys. Acta* **1999**, *1455*, 241–253.
 23. Lamblin, G.; Degroote, S.; Perini, J.M.; Delmotte, P.; Scharfman, A.; Davril, M.; Lo-Guidice, J.M.; Houdret, N.; Dumur, V.; Klein, A.; *et al.* Human airway mucin glycosylation: a combinatorial of carbohydrate determinants which vary in cystic fibrosis. *Glycoconj. J.* **2001**, *18*, 661–684.
 24. Tortorano, A.M.; Richardson, M.; Roilides, E.; van Diepeningen, A.; Caira, M.; Munoz, P.; Johnson, E.; Meletiadiis, J.; Pana, Z.D.; Lackner, M.; *et al.* ESCMID and ECMM joint guidelines on diagnosis and management of hyalohyphomycosis: *Fusarium spp.*, *Scedosporium spp.* and others. *Clin. Microbiol. Infect.* **2014**, *20*, 27–46.
 25. Troke, P.; Aguirrebengoa, K.; Arteaga, C.; Ellis, D.; Heath, C.H.; Lutsar, I.; Rovira, M.; Nguyen, Q.; Slavin, M.; Chen, S.C.A. Treatment of scedosporiosis with voriconazole: clinical experience with 107 patients.

-
- Antimicrob. Agents Chemother.* **2008**, *52*, 1743–1750.
26. Rodriguez-tudela, J.L.; Guarro, J.; Kantarcioglu, A.S.; Horre, R.; Estrella, M.C.; Berenguer, J.; Hoog, G.S.D.E. *Scedosporium apiospermum*: changing clinical spectrum of a therapy-refractory opportunist. *Med. Mycol.* **2006**, *44*, 295–327.
 27. Theuretzbacher, U.; Piddock, L.J. V Non-traditional antibacterial therapeutic options and challenges. *Cell Host Microbe* **2019**, *26*, 61–72.
 28. Krachler, A.M.; Orth, K. Targeting the bacteria-host interface strategies in anti-adhesion therapy. *Virulence* **2013**, *4*, 284–294.
 29. Ofek, I.; Hasty, D.L.; Sharon, N. Anti-adhesion therapy of bacterial diseases: prospects and problems. *FEMS Immunol. Med. Microbiol.* **2003**, *38*, 181–191.
 30. Pinto, M.R.; De Sá, A.C.M.; Limongi, C.L.; Rozental, S.; Santos, A.L.S.; Barreto-Bergter, E. Involvement of peptidorhamnomannan in the interaction of *Pseudallescheria boydii* and HEp2 cells. *Microbes Infect.* **2004**, *6*, 1259–1267.
 31. Bittencourt, V.C.B.; Figueiredo, R.T.; Silva, R.B.; Moura, D.S.; Fernandez, P.L.; Sasaki, G.L.; Mulloy, B.; Bozza, M.T.; Barreto-bergter, E. An α -glucan of *Pseudallescheria boydii* is involved in fungal phagocytosis and Toll-like receptor activation. *J. Biol. Chem.* **2006**, *281*, 22614–22623.
 32. Ghamrawi, S.; Renier, G.; Saulnier, P.; Cuenot, S.; Zykwiniska, A.; Dutilh, B.E.; Thornton, C.; Faure, S.; Bouchara, J. Cell wall modifications during conidial maturation of the human pathogenic fungus *Pseudallescheria boydii*. *PLoS One* **2014**, *9*, e100290.
 33. Rollin-Pinheiro, R.; Liporagi-lobes, L.C.; Meirelles, J.V. De; Souza, L.M. De Characterization of *Scedosporium apiospermum* glucosylceramides and their involvement in fungal development and macrophage functions. *PLoS One* **2014**, *9*, e98149.

-
34. Mello, T.P. De; Aor, A.C.; Gonçalves, D.D.S.; Seabra, S.H.; Branquinha, M.H.; Luis, A. *Scedosporium apiospermum*, *Scedosporium aurantiacum*, *Scedosporium minutisporum* and *Lomentospora prolificans*: a comparative study of surface molecules produced by conidial and germinated conidial cells. *Mem Inst Oswaldo Cruz* **2018**, *113*, 1–8.
 35. Sharon, N.; Lis, H. Lectins as cell recognition molecules. *Science* **1989**, *246*, 227–234.
 36. Kocourek, J.; Hořejší, V. A note on the recent discussion on definition of the term “LECTIN.” In Proceedings of the Fifth Lectin Meeting Bern, 1982; De Gruyter, **2020**; pp. 3–6.
 37. Sharon, N.; Lis, H. How proteins bind carbohydrates: Lessons from legume lectins. *J. Agric. Food Chem.* **2002**, *50*, 6586–6591.
 38. Weis, W.I.; Drickamer, K. Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **1996**, *65*, 441–473.
 39. Dodd, R.B.; Drickamer, K. Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. *Glycobiology* **2001**, *11*, 71–79.
 40. Sakai, K.; Hiemori, K.; Tateno, H.; Hirabayashi, J.; Gono, T. Fucose-specific lectin of *Aspergillus fumigatus*: binding properties and effects on immune response stimulation. *Med. Mycol.* **2019**, *57*, 71–83.
 41. Notova, S.; Bonnardel, F.; Lisacek, F.; Varrot, A.; Imberty, A. Structure and engineering of tandem repeat lectins. *Curr. Opin. Struct. Biol.* **2020**, *62*, 39–47.
 42. Raemaekers, J.M.; De Muro, L.; Gatehouse, J.A.; Fordham-Skelton, A.P. Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris*. Correct N-terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide. *Eur. J. Biochem.* **1999**, *265*, 394–403.

-
43. Nagae, M.; Soga, K.; Morita, K. Phytohemagglutinin from *Phaseolus vulgaris* (PHA-E) displays a novel glycan recognition mode using a common legume lectin fold. *Glycobiology* **2014**, *24*, 368–378.
 44. Martínez-Alarcón, D.; Blanco-Labra, A.; García-Gasca, T. Expression of lectins in heterologous systems. *Int. J. Mol. Sci.* **2018**, *19*, 616.
 45. Bermeo, R.; Bernardi, A.; Varrot, A. BC2L-C N-Terminal lectin domain complexed with histo blood group oligosaccharides provides new structural information. *Molecules* **2020**, *25*, 248.
 46. Šulák, O.; Cioci, G.; Lameignère, E.; Balloy, V.; Round, A.; Gutsche, I.; Malinovská, L.; Chignard, M.; Kosma, P.; Aubert, D.F.; *et al.* *Burkholderia cenocepacia* BC2L-C is a super lectin with dual specificity and proinflammatory activity. *PLoS Pathog.* **2011**, *7*, e1002238.
 47. Branco, A.T.; Bernabé, R.B.; Dos Santos Ferreira, B.; De Oliveira, M.V.V.; Garcia, A.B.; De Souza Filho, G.A. Expression and purification of the recombinant SALT lectin from rice (*Oryza sativa* L.). *Protein Expr. Purif.* **2004**, *33*, 34–38.
 48. Lannoo, N.; Vervecken, W.; Proost, P.; Rougé, P.; Van Damme, E.J.M. Expression of the nucleocytoplasmic tobacco lectin in the yeast *Pichia pastoris*. *Protein Expr. Purif.* **2007**, *53*, 275–282.
 49. Lakhtin, V.; Lakhtin, M.; Alyoshkin, V. Lectins of living organisms. The overview. *Anaerobe* **2011**, *17*, 452–455.
 50. Peumans WJ1, Van Damme EJ, Barre A, R.P. Classification of plant lectins in families of structurally and evolutionary related proteins. In *Advances in Experimental Medicine and Biology*; **2001**; Vol. 491, pp. 27–54.
 51. Bonnardel, F.; Mariethoz, J.; Salentin, S.; Robin, X.; Schroeder, M.; Perez, S.; Lisacek, F.D.S.; Imberty, A. Unilectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. *Nucleic Acids Res.* **2019**, *47*, D1236–D1244.

-
52. Varrot, A.; Basheer, S.M.; Imberty, A. Fungal lectins : structure , function and potential applications. *Curr. Opin. Struct. Biol.* **2013**, *23*, 678–685.
 53. Cioci, G.; Mitchell, E.P.; Chazalet, V.; Debray, H.; Oscarson, S.; Lahmann, M.; Gautier, C.; Breton, C.; Perez, S.; Imberty, A. β -Propeller crystal structure of *Psathyrella velutina* lectin: an integrin like fungal protein interacting with monosaccharides and calcium. *J. Mol. Biol.* **2006**, *357*, 1575–1591.
 54. Houser, J.; Komarek, J.; Kostlanova, N.; Cioci, G.; Varrot, A.; Kerr, S.C.; Lahmann, M.; Balloy, V.; Fahy, J. V; Chignard, M.; *et al.* A soluble fucose-specific lectin from *Aspergillus fumigatus* conidia - structure, specificity and possible role in fungal pathogenicity. *PLoS One* **2013**, *8*, e83077.
 55. Monzo, A.; Bonn, G.K.; Guttman, A. Lectin-immobilization strategies for affinity purification and separation of glycoconjugates. *TrAC - Trends Anal. Chem.* **2007**, *26*, 423–432.
 56. Khan, F.; Islam Khan, M. Fungal Lectins: Current molecular and biochemical perspectives. *Int. J. Biol. Chem.* **2011**, *5*, 1–20.
 57. Kobayashi, Y. and Kawagishi, H. Fungal lectins: a growing family. In *Lectins*; 2014; Vol. 1200, pp. 15–38 ISBN 9781493912926.
 58. Sarup Singh, R.; Preet Kaur, H.; Rakesh Kanwar, J. Mushroom lectins as promising anticancer substances. *Curr. Protein Pept. Sci.* **2016**, *17*, 797–807.
 59. Koharudin, L.M.I.; Viscomi, A.R.; Jee, J.; Ottonello, S.; Gronenborn, A.M. The evolutionarily conserved family of cyanovirin-N homologs : structures and carbohydrate specificity. *Structure* **2008**, *4*, 570–584.
 60. Mohamed, H.; Rouf, R.; Tiralongo, E.; May, T.W. Mushroom lectins : Specificity , structure and bioactivity relevant to human disease. *Int. J. Mol. Sci.* **2015**, *16*, 7802–7838.
 61. Shoaf-sweeney, K.D.; Hutkins, R.W. Adherence, anti-adherence, and oligosaccharides: preventing pathogens from sticking to the host. In

- Advances in Food and Nutrition Research*; 2009; Vol. 55, pp. 101–161.
62. Varki A, Cummings RD, Esko JD, *et al.* Microbial lectins: hemagglutinins, adhesins, and toxins. In *Essentials of Glycobiology*; NY, **2009**.
 63. Lehot, V.; Brissonnet, Y.; Dussouy, C.; Brument, S.; Cabanettes, A.; Gillon, E.; Deniaud, D.; Varrot, A.; Le Pape, P.; Gouin, S.G. Multivalent fucosides with nanomolar affinity for the *Aspergillus fumigatus* lectin FleA prevent spore adhesion to pneumocytes. *Chem. – A Eur. J.* **2018**, *24*, 19243–19249.
 64. Goyard, D.; Baldoneschi, V.; Varrot, A.; Fiore, M.; Imberty, A.; Richichi, B.; Renaudet, O.; Nativi, C. Multivalent glycomimetics with affinity and selectivity toward fucose-binding receptors from emerging pathogens. *Bioconjug. Chem.* **2018**, *29*, 83–88.
 65. Akkouch, O.; Ng, T.B.; Singh, S.S.; Yin, C.; Dan, X.; Chan, Y.S.; Pan, W.; Cheung, R.C.F. Lectins with anti-HIV activity: A review. *Molecules* **2015**, *20*, 648–668.
 66. Fang, E.F.; Lin, P.; Wong, J.H.; Tsao, S.W.; Ng, T.B. A lectin with anti-HIV-1 reverse transcriptase, antitumor, and nitric oxide inducing activities from seeds of *Phaseolus vulgaris* cv. Extralong Autumn Purple Bean. *J. Agric. Food Chem.* **2010**, *58*, 2221–2229.
 67. Sharon, N.; Ofek, I. Safe as mother's milk: carbohydrates as future anti-adhesion drugs for bacterial diseases. *Glycoconj. J.* **2000**, *17*, 659–664.
 68. Sattin, S.; Bernardi, A. Glycoconjugates and glycomimetics as microbial anti-adhesives. *Trends Biotechnol.* **2016**, *34*, 483–495.
 69. Sharon, N. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochim. Biophys. Acta* **2006**, *1760*, 527–537.
 70. Asadi, A.; Razavi, S.; Talebi, M.; Gholami, M. A review on anti-adhesion therapies of bacterial diseases. *Infection* **2019**, *47*, 13–23.
 71. Kunz, C. Historical aspects of human milk oligosaccharides. *Adv. Nutr.* **2012**, *3*, 430S–439S.

-
72. Ayechu-Muruzabal, V.; van Stigt, A.H.; Mank, M.; Willemsen, L.E.M.; Stahl, B.; Garssen, J.; van't Land, B. Diversity of human milk oligosaccharides and effects on early life immune development. *Front. Pediatr.* **2018**, *6*, 1–9.
73. Aronson, M.; Medalia, O.; Schori, L.; Mirelman, D.; Sharon, N.; Ofek, I. Prevention of colonization of the urinary tract of mice with *Escherichia coli* by blocking of bacterial adherence with methyl α -D-mannopyranoside. *J. Infect. Dis.* **1979**, *139*, 329–332.
74. Ruiz-Palacios, G.M.; Cervantes, L.E.; Ramos, P.; Chavez-Munguia, B.; Newburg, D.S. Campylobacter jejuni binds intestinal H(O) antigen (Fuc α 1, 2Gal β 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. *J. Biol. Chem.* **2003**, *278*, 14112–14120.
75. Edén, C.S.; Freter, R.; Hagberg, L.; Hull, R.; Hull, S.; Leffler, H.; Schoolnik, G. Inhibition of experimental ascending urinary tract infection by an epithelial cell-surface receptor analogue. *Nature* **1982**, *298*, 560–562.
76. Mouricout, M.; Petit, J.M.; Carias, J.R.; Julien, R. Glycoprotein glycans that inhibit adhesion of *Escherichia coli* mediated by K99 fimbriae: Treatment of experimental colibacillosis. *Infect. Immun.* **1990**, *58*, 98–106.
77. Fader, R.C.; Davis, C.P. Effect of piliation on *Klebsiella pneumoniae* infection in rat bladders. *Infect. Immun.* **1980**, *30*, 554–561.
78. Mysore, J. V.; Wigginton, T.; Simon, P.M.; Zopf, D.; Heman-Ackah, L.M.; Dubois, A. Treatment of *Helicobacter pylori* infection in rhesus monkeys using a novel antiadhesion compound. *Gastroenterology* **1999**, *117*, 1316–1325.
79. Izhar, M.; Nuchamowitz, Y.; Mirelman, D. Adherence of *Shigella flexneri* to guinea pig intestinal cells is mediated by a mucosal adhesin. *Infect. Immun.* **1982**, *35*, 1110–1118.
80. Idänpään-Heikkilä, I.; Simon, P.M.; Zopf, D.; Vullo, T.; Cahill, P.; Sokol, K.;

-
- Tuomanen, E. Oligosaccharides interfere with the establishment and progression of experimental pneumococcal pneumonia. *J. Infect. Dis.* **1997**, *176*, 704–712.
81. Wang, Q.; Singh, S.; Taylor, K.G.; Doyle, R.J. Anti-adhesins of *Streptococcus sobrinus*. *Adv Exp Med Biol* **1996**, *408*, 249–250.
82. Cywes, C.; Stamenkovic, I.; Wessels, M.R. CD44 as a receptor for colonization of the pharynx by group A *Streptococcus*. *J. Clin. Invest.* **2000**, *106*, 995–1002.
83. Ernst, B.; Magnani, J.L. From carbohydrate leads to glycomimetic drugs. *Nat. Publ. Gr.* **2009**, *8*, 661–677.
84. Arastehfar, A.; Carvalho, A.; Houbraken, J.; Lombardi, L.; Garcia-Rubio, R. *Aspergillus fumigatus* and aspergillosis: From basics to clinics. *Studies in Mycology* **2021**, *10*, 100115.
85. Percudani, R.; Montanini, B.; Ottonello, S. The anti-HIV Cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins* **2005**, *60*, 670–678.
86. Yu, H.; Liu, Z.T.; Lv, R.; Zhang, W.Q. Antiviral activity of recombinant cyanovirin-N against HSV-1. *Virol. Sin.* **2010**, *25*, 432–439.
87. Vandeputte, P.; Dugé de Bernonville, T.; Le Govic, Y.; Le Gal, S.; Nevez, G.; Papon, N.; Bouchara, J.P. Comparative transcriptome analysis unveils the adaptive mechanisms of *Scedosporium apiospermum* to the microenvironment encountered in the lungs of patients with cystic fibrosis. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3468–3483.
88. Matei, E.; Basu, R.; Furey, W.; Shi, J.; Calnan, C.; Aiken, C.; Gronenborn, A.M. Structure and glycan binding of a new cyanovirin-N homolog. *J. Biol. Chem.* **2016**, *291*, 18967–18976.
89. Qi, X.; Yang, Y.; Su, Y.; Wang, T. Molecular cloning and sequence analysis of cyanovirin-N homology gene in *Ceratopteris thalictroides*. *Am. Fern J.*

- 2009, 99, 79–92.
90. Bewley, C.A. Solution Structure of a Cyanovirin-N: Mana 1-2Mana Complex: Structural Basis for High-Affinity Carbohydrate-Mediated Binding to gp120. *Structure* **2001**, 9, 931–940.
 91. Kehr, J.C.; Zilliges, Y.; Springer, A.; Disney, M.D.; Ratner, D.D.; Bouchier, C.; Seeberger, P.H.; De Marsac, N.T.; Dittmann, E. A mannan binding lectin is involved in cell-cell attachment in a toxic strain of *Microcystis aeruginosa*. *Mol. Microbiol.* **2006**, 59, 893–906.
 92. Lakin-Thomas, P.L.; Brody, S. Circadian rhythms in microorganisms: New complexities. *Annu. Rev. Microbiol.* **2004**, 58, 489–519.
 93. Soragni, E.; Bolchi, A.; Balestrini, R.; Gambaretto, C.; Percudani, R.; Bonfante, P.; Ottonello, S. A nutrient-regulated, dual localization phospholipase A2 in the symbiotic fungus *Tuber borchii*. *EMBO J.* **2001**, 20, 5079–5090.
 94. Hündling, D. Biochemical and structural studies of lectines from opportunistic filamentous fungi. *Biochemistry, Molecular Biology. Université Grenoble Alpes*, **2015**. English. NNT : 2015GREAV055.
 95. Houben, K.; Marion, D.; Tarbouriech, N.; Ruigrok, R.W.H.; Blanchard, L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J. Virol.* **2007**, 81, 6807–6816.
 96. Kabsch, W. XDS. *Acta Crystallogr.* **2010**, 66, 125–132.
 97. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.W.; McCoy, A.; *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, 67, 235–242.
 98. McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**,

-
- 40, 658–674.
99. Murshudov, G.N.; Nicholls, R.A. REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 355–367.
 100. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
 101. Robert, X.; Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, *42*, 320–324.
 102. Jones, D.T.; Taylor, W.R.; Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **1992**, *8*, 275–282.
 103. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874.
 104. Geissner, A.; Reinhardt, A.; Rademacher, C.; Johannssen, T.; Monteiro, J.; Lepenies, B.; Thépaut, M.; Fieschi, F.; Mrázková, J.; Wimmerova, M.; *et al.* Microbe-focused glycan array screening platform. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 1958–1967.
 105. Koharudin, L.; Furey, W.; Gronenborn, A. A. A designed chimeric Cyanovirin-N homolog lectin: structure and molecular basis of sucrose binding. *Proteins* **2010**, *77*, 904–915.
 106. Romero, M.; Messina, F.; Marin, E.; Arechavala, A.; Depardo, R.; Walker, L.; Negroni, R.; Santiso, G. Antifungal resistance in clinical isolates of *Aspergillus* spp.: When local epidemiology breaks the norm. *J. Fungi* **2019**, *5*, 41.
 107. Källberg, M.; Wang, H.; Wang, S.; Jian, P.; Wang, Z.; Hui Lu, J. Template-based protein structure modeling using the RaptorX web server. *Nat.*

- Protoc.* **2012**, *7*, 1511–1522.
108. Chen, H.; Kovalchuk, A.; Keriö, S.; Asiegbu, F.O. Distribution and bioinformatic analysis of the cerato-platanin protein family in Dikarya. *Mycologia* **2013**, *105*, 1479–1488.
109. Luti, S.; Sella, L.; Quarantin, A.; Pazzagli, L.; Baccelli, I. Twenty years of research on cerato-platanin familyproteins: clues, conclusions, and unsolved issues. *Fungal Biol. Rev.* **2019**, 1–41.
110. Pazzagli, L.; Cappugi, G.; Manao, G.; Camici, G.; Santini, A.; Scala, A. Purification, characterization, and amino acid sequence of cerato- platanin, a new phytotoxic protein from *Ceratocystis fimbriata f. sp. platani*. *J. Biol. Chem.* **1999**, *274*, 24959–24964.
111. Boddi, S.; Comparini, C.; Calamassi, R.; Pazzagli, L.; Cappugi, G.; Scala, A. Cerato-platanin protein is located in the cell walls of ascospores, conidia and hyphae of *Ceratocystis fimbriata f. sp. platani*. *FEMS Microbiol. Lett.* **2004**, *233*, 341–346.
112. Seidl, V.; Marchetti, M.; Schandl, R.; Allmaier, G.; Kubicek, C.P. Epl1, the major secreted protein of *Hypocrea atroviridis* on glucose, is a member of a strongly conserved protein family comprising plant defense response elicitors. *FEBS J.* **2006**, *273*, 4346–4359.
113. De Oliveira, A.L.; Gallo, M.; Pazzagli, L.; Benedetti, C.E.; Cappugi, G.; Scala, A.; Pantera, B.; Spisni, A.; Pertinhez, T.A.; Cicero, D.O. The structure of the elicitor cerato-platanin (CP), the first member of the CP fungal protein family, reveals a double $\psi\beta$ -barrel fold and carbohydrate binding. *J. Biol. Chem.* **2011**, *286*, 17560–17568.
114. Baccelli, I.; Comparini, C.; Bettini, P.P.; Martellini, F.; Ruocco, M.; Pazzagli, L.; Bernardi, R.; Scala, A. The expression of the cerato-platanin gene is related to hyphal growth and chlamydospores formation in *Ceratocystis platani*. *FEMS Microbiol. Lett.* **2012**, *327*, 155–163.

-
115. Pan, Y.; Wei, J.; Yao, C.; Reng, H.; Gao, Z. SsSm1, a Cerato-platanin family protein, is involved in the hyphal development and pathogenic process of *Sclerotinia sclerotiorum*. *Plant Sci.* **2018**, *270*, 37–46.
 116. Baccelli, I. Cerato-platanin family proteins: One function for multiple biological roles? *Front. Plant Sci.* **2015**, *5*, 2013–2016.
 117. Jeong, J.S.; Mitchell, T.K.; Dean, R.A. The magnaporthe grisea snodprot1 homolog ,MSP1, isrequired for virulence. *FEMS Microbiol. Lett.* **2007**, *273*, 157–165.
 118. Djonović, S.; Pozo, M.J.; Dangott, L.J.; Howell, C.R.; Kenerley, C.M. Sm1, a proteinaceous elicitor secreted by the biocontrol fungus *Trichoderma virens* induces plant defense responses and systemic resistance. *Mol. Plant-Microbe Interact.* **2006**, *19*, 838–853.
 119. Samolski, I.; De Luis, A.; Vizcaíno, J.A.; Monte, E.; Suárez, M.B. Gene expression analysis of the biocontrol fungus *Trichoderma harzianum* in the presence of tomato plants, chitin, or glucose using a high-density oligonucleotide microarray. *BMC Microbiol.* **2009**, *9*, 1–14.
 120. Jeong, J.S.; Mitchell, T.K.; Dean, R.A. The Magnaporthe grisea snodprot1 homolog, MSP1, is required for virulence. *FEMS Microbiol. Lett.* **2007**, *273*, 157–165.
 121. Frías, M.; González, C.; Brito, N. BcSpl1, a cerato-platanin family protein, contributes to *Botrytis cinerea* virulence and elicits the hypersensitive response in the host. *New Phytol.* **2011**, *192*, 483–495.
 122. Frischmann, A.; Neudl, S.; Gaderer, R.; Bonazza, K.; Zach, S.; Gruber, S.; Spadiut, O.; Friedbacher, G.; Grothe, H.; Seidl-Seiboth, V. Self-assembly at air/water interfaces and carbohydrate binding properties of the small secreted protein EPL1 from the fungus *Trichoderma atroviride*. *J. Biol. Chem.* **2013**, *288*, 4278–4287.
 123. Barsottini, M.R. de O.; de Oliveira, J.; Adamoski, D.; Teixeira, P.J.P.L.;

-
- Tiezzi, H.O.; Sforça, M.L.; Cassago, A.; Portugal, R. V; de Oliveira, P.S.L.; Zeri, A.C.M.; *et al.* Functional diversification of cerato-platanins in *Moniliophthora perniciosa* as seen by differential expression and protein function specialization. *Mol. Plant. Microbe. Interact.* **2013**, *26*, 1281–1293.
124. Quarantin, A.; Castiglioni, C.; Schäfer, W.; Favaron, F.; Sella, L. The *Fusarium graminearum* cerato-platanins loosen cellulose substrates enhancing fungal cellulase activity as expansin-like proteins. *Plant Physiol. Biochem.* **2019**, *139*, 229–238.
125. Baccelli, I.; Luti, S.; Bernardi, R.; Scala, A.; Pazzagli, L. Cerato-platanin shows expansin-like activity on cellulosic materials. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 175–184.
126. Pan, S.; Cole, G.T. Molecular and biochemical characterization of a *Coccidioides immitis*- specific antigen. *Infect. Immun.* **1995**, *63*, 3994–4002.
127. Chaudhary, N.; Staab, J.F.; Marr, K.A. Healthy human T-cell responses to *Aspergillus fumigatus* antigens. *PLoS One* **2010**, *5*.
128. Vargas, W.A.; Djonović, S.; Sukno, S.A.; Kenerley, C.M. Dimerization controls the activity of fungal elicitors that trigger systemic resistance in plants. *J. Biol. Chem.* **2008**, *283*, 19804–19815.
129. Endo, T.; Ohbayashi, H.; Kanazawa, K.; Kochibes, N.; Kobata, A. Carbohydrate binding specificity of immobilized *Psathyrella velutina* lectin. *J. Biol. Chem.* **1992**, *267*, 707–713.
130. R. B. Parekh, R. A. Dwek, B. J. Sutton, D. L. Fernandes, A. Leung, D. Stanworth, T. W. Rademacher, T. Mizuochi, T. Taniguchi, K. Matsuta, F. Takeuchi, Y. Nagano, T.M.& A.K. Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **1985**, *316*, 452-457.
131. Liljeblad, M.; Lundblad, A.; Pålsson, P. Analysis of agalacto-IgG in rheumatoid arthritis using surface plasmon resonance. *Glycoconj. J.* **2000**,

- 17, 323–329.
132. Moore, J.S.M.; Wu, X.; Kulhavy, R.; Tomana, M.; Novak, J.; Moldoveanu, Z.; Brown, R.; Goepfert, P.A.; Mestecky, J. Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals. *AIDS* **2005**, *19*, 381–389.
133. Ueda, H.; Kojima, K.; Saitoh, T.; Ogawa, H. Interaction of a lectin from *Psathyrella velutina* mushroom with N-acetylneuraminic acid. *FEBS Lett.* **1999**, *448*, 75–80.
134. Ueda, H.; Saitoh, T.; Kojima, K.; Ogawa, H. Multi-specificity of a *Psathyrella velutina* mushroom lectin: Heparin/pectin binding occurs at a site different from the N-acetylglucosamine/N-acetylneuraminic acid-specific site. *J. Biochem.* **1999**, *126*, 530–537.
135. Loftus, J.C.; Smith, J.W.; Ginsberg, M.H. Integrin-mediated cell adhesion: The extracellular face. *J. Biol. Chem.* **1994**, *269*, 25235–25238.
136. Springer, T.A. Folding of the N-terminal, ligand-binding region of integrin α -subunits into a β -propeller domain. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 65–72.
137. Kumar, A.; Ahmed, R.; Singh, P.K.; Shukla, P.K. Identification of virulence factors and diagnostic markers using immunosecretome of *Aspergillus fumigatus*. *J. Proteomics* **2011**, *74*, 1104–1112.
138. Lucke, C.; Weiwad, M. Insights into immunophilin structure and function. *Curr. Med. Chem.* **2011**, *18*, 5333–5354.
139. M. Christopher, A.M.L.S. FK506-Binding proteins and their diverse functions. *Physiol. Behav.* **2016**, *176*, 100–106.
140. Harikishore, A.; Sup Yoon, H. Immunophilins: Structures, mechanisms and ligands. *Curr. Mol. Pharmacol.* **2015**, *9*, 37–47.
141. Pinto, D.; Duarte, M.; Soares, S.; Tropschug, M.; Videira, A. Identification of all FK506-binding proteins from *Neurospora crassa*. *Fungal Genet. Biol.* **2008**, *45*, 1600–1607.

-
142. Barik, S. Immunophilins: For the love of proteins. *Cell. Mol. Life Sci.* **2006**, *63*, 2889–2900.
143. Tonthat, N.K.; Juvvadi, P.R.; Zhang, H.; Lee, S.C.; Venters, R.; Spicer, L.; Steinbach, W.J.; Heitman, J.; Schumacher, M.A. Structures of pathogenic fungal FKBP12s reveal possible self-catalysis function. *MBio* **2016**, *7*, 1–11.
144. Falloon, K.; Juvvadi, P.R.; Richards, A.D.; Vargas-Muñiz, J.M.; Renshaw, H.; Steinbach, W.J. Characterization of the FKBP12-encoding genes in *Aspergillus fumigatus*. *PLoS One* **2015**, *10*, e0137869.
145. Nambu, M.; Covell, J.A.; Kapoor, M.; Li, X.; Moloney, M.K.; Numa, M.M.; Soltow, Q.A.; Trzoss, M.; Webb, P.; Webb, R.R.; *et al.* A calcineurin antifungal strategy with analogs of FK506. *Bioorganic Med. Chem. Lett.* **2017**, *27*, 2465–2471.
146. Jung, J.A.; Yoon, Y.J. Development of non-immunosuppressive FK506 derivatives as antifungal and neurotrophic agents. *J. Microbiol. Biotechnol.* **2020**, *30*, 1–10.
147. Karababa, M.; Valentino, E.; Pardini, G.; Coste, A.T.; Bille, J.; Sanglard, D. CRZ1, a target of the calcineurin pathway in *Candida albicans*. *Mol. Microbiol.* **2006**, *59*, 1429–1451.
148. Juvvadi, P.R.; Fox, D.; Bobay, B.G.; Hoy, M.J.; Gobeil, S.M.C.; Venters, R.A.; Chang, Z.; Lin, J.J.; Averette, A.F.; Cole, D.C.; *et al.* Harnessing calcineurin-FK506-FKBP12 crystal structures from invasive fungal pathogens to develop antifungal agents. *Nat. Commun.* **2019**, *10*, 1–18.
149. Griffith, J.P.; Kim, J.L.; Kim, E.E.; Sintchak, M.D.; Thomson, J.A.; Fitzgibbon, M.J.; Fleming, M.A.; Caron, P.R.; Hsiao, K.; Navia, M.A. X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex. *Cell* **1995**, *82*, 507–522.
150. O'Connor, B.; Monaghan, D.; Cawley, J. Lectin affinity chromatography (LAC). In *Methods in Molecular Biology*; Humana Press Inc., **2017**; Vol. 1485,

- pp. 411–420.
151. Nascimento, K.S.; Cunha, A.I.; Nascimento, K.S.; Cavada, B.S.; Azevedo, A.M.; Aires-Barros, M.R. An overview of lectins purification strategies. *J. Mol. Recognit.* **2012**, *25*, 527–541.
 152. Zheng, T.; Yu, H.; Alexander, C.M.; Beebe, D.J.; Smith, L.M. Lectin-modified microchannels for mammalian cell capture and purification. *Biomed. Microdevices* **2007**, *9*, 611–617.
 153. McCoy, J.P. The application of lectins to the characterization and isolation of mammalian cell populations. *Cancer Metastasis Rev.* **1987**, *6*, 595–613.
 154. Ashraf, M.T.; Khan, R.H. Mitogenic Lectins. *Med. Sci. Monit.* **2003**, *9*, RA265–269.
 155. Vandenborre, G.; Smagghe, G.; Van Damme, E.J.M. Plant lectins as defense proteins against phytophagous insects. *Phytochemistry* **2011**, *72*, 1538–1550.
 156. Chrispeels, M.J.; Raikhel, N. V; Raikhelb, N. V Lectins, lectin genes, and their role in plant defense. *Plant Cell* **1991**, *3*, 1–9.
 157. Mereiter, S.; Balmaña, M.; Campos, D.; Gomes, J.; Reis, C.A. Glycosylation in the era of cancer-targeted therapy: Where are we heading? *Cancer Cell* **2019**, *36*, 6–16.
 158. Martínez-Alarcón, D.; Varrot, A.; Fitches, E.; Gatehouse, J.A.; Cao, M.; Pyati, P.; Blanco-Labra, A.; Garcia-Gasca, T. Recombinant lectin from Tepary Bean (*Phaseolus acutifolius*) with specific recognition for cancer-associated glycans: production, structural characterization, and target identification. *Biomolecules* **2020**, *10*, 654.
 159. Khan, F.; Khan, R.H.; Sherwani, A.; Mohmood, S.; Azfer, M.A. Lectins as markers for blood grouping. *Med. Sci. Monit.* **2002**, *8*, 293–301.
 160. Campanero-Rhodes, M.A.; Palma, A.S.; Menéndez, M.; Solís, D. Microarray strategies for exploring bacterial surface glycans and their interactions with glycan-binding proteins. *Front. Microbiol.* **2020**, *10*, 2909.

-
161. Nagata, S. Isolation, characterization, and extra-embryonic secretion of the *Xenopus laevis* embryonic epidermal lectin, XEEL. *Glycobiology* **2005**, *15*, 281–290.
162. Imamichi, Y.; Yokoyama, Y. Purification, characterization and cDNA cloning of a novel lectin from the jellyfish *Nemopilema nomurai*. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **2010**, *156*, 12–18.
163. Almanza, M.; Vega, N.; Pérez, G. Isolating and characterising a lectin from *Galactia lindenii* seeds that recognises blood group H determinants. *Arch. Biochem. Biophys.* **2004**, *429*, 180–190.
164. Naeem, A.; Haque, S.; Khan, R.H. Purification and characterization of a novel b-D-galactosides-specific lectin from *Clitoria ternatea*. *Protein J.* **2007**, *26*, 403–413.
165. Kawsar, S.M.A.; Matsumoto, R.; Fujii, Y.; Yasumitsu, H.; Dogasaki, C.; Hosono, M.; Nitta, K.; Hamako, J.; Matsui, T.; Kojima, N.; *et al.* Purification and biochemical characterization of a D-galactose binding lectin from Japanese sea hare (*Aplysia kurodai*) eggs. *Biochem.* **2009**, *74*, 709–716.
166. Braga, M.D.M.; Martins, A.M.C.; Amora, D.N.; de Menezes, D.B.; Toyama, M.H.; Toyama, D.O.; Marangoni, S.; Barbosa, P.S.F.; de Sousa Alves, R.; Fonteles, M.C.; *et al.* Purification and biological effects of C-type lectin isolated from *Bothrops insularis* venom. *Toxicon* **2006**, *47*, 859–867.
167. Fujii, Y.; Kawsar, S.M.A.; Matsumoto, R.; Yasumitsu, H.; Ishizaki, N.; Dogasaki, C.; Hosono, M.; Nitta, K.; Hamako, J.; Tabei, M.; *et al.* A d-galactose-binding lectin purified from coronate moon turban, *Turbo (Lunella) coreensis*, with a unique amino acid sequence and the ability to recognize lacto-series glycosphingolipids. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **2011**, *158*, 30–37.
168. Hamako, J.; Suzuki, Y.; Hayashi, N.; Kimura, M.; Ozeki, Y.; Hashimoto, K.; Matsui, T. Amino acid sequence and characterization of C-type lectin

- purified from the snake venom of *Crotalus ruber*. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **2007**, *146*, 299–306.
169. Mendonça-Franqueiro, E.D.P.; Alves-Paiva, R.D.M.; Sartim, M.A.; Callejon, D.R.; Paiva, H.H.; Antonucci, G.A.; Rosa, J.C.; Cintra, A.C.O.; Franco, J.J.; Arantes, E.C.; *et al.* Isolation, functional, and partial biochemical characterization of galatrox, an acidic lectin from *Bothrops atrox* snake venom. *Acta Biochim. Biophys. Sin. (Shanghai)*. **2011**, *43*, 181–192.
170. Pohleven, J.; Obermajer, N.; Sabotič, J.; Anžlovar, S.; Sepčič, K.; Kosa, J.; Kralj, B.; Štrukelj, B.; Brzina, J. Purification, characterization and cloning of a ricin b-like lectin from mushroom *clitocybe nebularis* with antiproliferative activity against human le. *Biochim Biophys Acta* **2009**, *1790*, 173–181.
171. Rocha, B.A.M.; Moreno, F.B.M.B.; Delatorre, P.; Souza, E.P.; Marinho, E.S.; Benevides, R.G.; Rustiguel, J.K.R.; Souza, L.A.G.; Nagano, C.S.; Debray, H.; *et al.* Purification, characterization, and preliminary x-ray diffraction analysis of a lactose-specific lectin from *Cymbosema roseum* seeds. *Appl. Biochem. Biotechnol.* **2009**, *152*, 383–393.
172. Silva, J.A.; Damico, D.C.S.; Baldasso, P.A.; Mattioli, M.A.; Winck, F. V.; Fraceto, L.F.; Novello, J.C.; Marangoni, S. Isolation and biochemical characterization of a galactoside binding lectin from *Bauhinia variegata* candida (BvcL) seeds. *Protein J.* **2007**, *26*, 193–201.
173. Oladokun, B.O.; Omisore, O.N.; Osukoya, O.A.; Kuku, A. Anti-nociceptive and anti-inflammatory activities of *Tetracarpidium conophorum* seed lectin. *Sci. African* **2019**, *3*, e00073.
174. Pohleven, J.; Brzin, J.; Vrabc, L.; Leonardi, A.; Čokl, A.; Štrukelj, B.; Kos, J.; Sabotič, J. Basidiomycete *Clitocybe nebularis* is rich in lectins with insecticidal activities. *Appl. Microbiol. Biotechnol.* **2011**, *91*, 1141–1148.
175. Wang, J.H.; Kong, J.; Li, W.; Molchanova, V.; Chikalovets, I.; Belogortseva,

- N.; Luk'yanov, P.; Zheng, Y.T. A B-galactose-specific lectin isolated from the marine worm *Chaetopterus variopedatus* possesses anti-HIV-1 activity. *Comp. Biochem. Physiol. - C Toxicol. Pharmacol.* **2006**, *142*, 111–117.
176. Mo, H.; Winter, H.C.; Goldstein, I.J. Purification and characterization of a Neu5Aca2-6Gal β 1-4Glc/GlcNAc- specific lectin from the fruiting body of the polypore mushroom *Polyporus squamosus*. *J. Biol. Chem.* **2000**, *275*, 10623–10629.
177. Konami, Y.; Yamamoto, K.; Osawa, T. Purification and characterization of two types of *Cytisus sessilifolius* anti-H(O) lectins by affinity chromatography. *Biol. Chem. Hoppe. Seyler.* **1991**, *372*, 103–112.
178. Kvennefors, E.C.E.; Leggat, W.; Hoegh-Guldberg, O.; Degnan, B.M.; Barnes, A.C. An ancient and variable mannose-binding lectin from the coral *Acropora millepora* binds both pathogens and symbionts. *Dev. Comp. Immunol.* **2008**, *32*, 1582–1592.
179. Houser, J.; Komarek, J.; Cioci, G.; Varrot, A.; Imberty, A.; Wimmerova, M. Structural insights into *Aspergillus fumigatus* lectin specificity : AFL binding sites are functionally non-equivalent. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2015**, *442*–453.
180. Ooi, L.S.M.; Yu, H.; Chen, C.M.; Sun, S.S.M.; Ooi, V.E.C. Isolation and characterization of a bioactive mannose-binding protein from the Chinese chive *Allium tuberosum*. *J. Agric. Food Chem.* **2002**, *50*, 696–700.
181. Suseelan, K.N.; Mitra, R.; Pandey, R.; Sainis, K.B.; Krishna, T.G. Purification and characterization of a lectin from wild sunflower (*Helianthus tuberosus* L.) tubers. *Arch. Biochem. Biophys.* **2002**, *407*, 241–247.
182. Jimbo, M.; Usui, R.; Sakai, R.; Muramoto, K.; Kamiya, H. Purification, cloning and characterization of egg lectins from the teleost *Tribolodon brandti*. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **2007**, *147*, 164–171.
183. Watanabe, Y.; Shiina, N.; Shinozaki, F.; Yokoyama, H.; Kominami, J.;

- Nakamura-Tsuruta, S.; Hirabayashi, J.; Sugahara, K.; Kamiya, H.; Matsubara, H.; *et al.* Isolation and characterization of l-rhamnose-binding lectin, which binds to microsporidian *Glugea plecoglossi*, from ayu (*Plecoglossus altivelis*) eggs. *Dev. Comp. Immunol.* **2008**, *32*, 487–499.
184. Devi, H.K.; Devi, S.K.; Rully, H.; Singh, S.J.; Singh, W.S.; Thongam, H.; Singh, L.R. Purification and characterization of a novel rhamnose/fucose-specific lectin from the hemolymph of Oak Tasar (*Antheraea proylei* J.) silkworm. *Protein Pept. Lett.* **2020**, *27*, 649 - 657.
185. Cammarata, M.; Benenati, G.; Odom, E.W.; Salerno, G.; Vizzini, A.; Vasta, G.R.; Parrinello, N. Isolation and characterization of a fish F-type lectin from gilt head bream (*Sparus aurata*) serum. *Biochim. Biophys. Acta - Gen. Subj.* **2007**, *1770*, 150–155.
186. Mansour, M.H.; Abdul-Salam, F. Characterization of fucose-binding lectins in rock- and mud-dwelling snails inhabiting Kuwait Bay. *Immunobiology* **2009**, *214*, 77–85.
187. Kortt, A.A. Isolation and characterization of the lectins from the seeds of *Psophocarpus scandens*. *Phytochemistry* **1988**, *27*, 2847–2855.
188. Baues, R.J.; Gray, G.R. Lectin purification on affinity columns containing reductively aminated disaccharides. *J. Biol. Chem.* **1977**, *252*, 57–60.
189. Nakagawa, R.; Yasokawa, D.; Ikeda, T.; Nagashima, K. Purification and characterization of two lectins from callus of *Helianthus tuberosus*. *Biosci. Biotechnol. Biochem.* **1996**, *60*, 259–262.
190. Gerlach, D.; Schlott, B.; Zaehringerr, U.; Schmidt, K.-H. N-acetyl-d-galactosamine/N-acetyl-d-glucosamine – recognizing lectin from the snail *Cepaea hortensis*: purification, chemical characterization, cloning and expression in *E. coli*. *FEMS Immunol. Med. Microbiol.* **2005**, *43*, 223–232.
191. Qureshi, I.A.; Dash, P.; Srivastava, P.S.; Koundal, K.R. Purification and characterization of anN-acetyl-d-galactosamine-specific lectin from seeds

- of chickpea (*Cicer arietinum* L.). *Phytochem. Anal.* **2006**, *17*, 350–356.
192. Vretblad, P. Purification of lectins by biospecific affinity chromatography. *BBA - Protein Struct.* **1976**, *434*, 169–176.
193. Adhya, M.; Singha, B.; Chatterjee, B.P. Purification and characterization of an N-acetylglucosamine specific lectin from marine bivalve *Macoma birmanica*. *Fish Shellfish Immunol.* **2009**, *27*, 1–8.
194. Kaur, A.; Singh, J.; Kamboj, S.S.; Sexana, A.K.; Pandita, R.M.; Shamnugavel, M. Isolation of an N-acetyl-d-glucosamine specific lectin from the rhizomes of *Arundo donax* with antiproliferative activity. *Phytochemistry* **2005**, *66*, 1933–1940.
195. Kim, G.H.; Klochkova, T.A.; Yoon, K.-S.; Song, Y.-S.; Lee, K.P. Purification and characterization of a lectin, bryohealin, involved in the protoplast formation of a marine green alga *Bryopsis plumosa* (chlorophyta). *J. Phycol.* **2006**, *42*, 86–95.
196. Maheswari, R.; Mullainadhan, P.; Arumugam, M. Isolation and characterization of an acetyl group-recognizing agglutinin from the serum of the Indian white shrimp *Fenneropenaeus indicus*. *Arch. Biochem. Biophys.* **2002**, *402*, 65–76.
197. Wang, H.; Ng, T.B. Isolation of a novel N-acetylglucosamine-specific lectin from fresh sclerotia of the edible mushroom *Pleurotus tuberregium*. *Protein Expr. Purif.* **2003**, *29*, 156–160.
198. Koyama, Y.; Katsuno, Y.; Miyoshi, N.; Hayakawa, S.; Mita, T.; Muto, H.; Isemura, S.; Aoyagi, Y.; Isemura, M. Apoptosis induction by lectin isolated from the mushroom *Boletopsis leucomelas* in U937 cells. *Biosci. Biotechnol. Biochem.* **2002**, *66*, 784–789.
199. Bovi, M.; Carrizo, M.E.; Capaldi, S.; Perduca, M.; Chiarelli, L.R.; Galliano, M.; Monaco, H.L. Structure of a lectin with antitumoral properties in king bolete (*Boletus edulis*) mushrooms. *Glycobiology* **2011**, *21*, 1000–1009.

-
200. Narahari, A.; Swamy, M.J.; Statement, F. Rapid affinity-purification and physico-chemical characterization of pumpkin (*Cucurbita maxima*) phloem exudate lectin. *Biosci. Rep.* **2010**, *30*, 15–34.
201. Santi-Gadelha, T.; de Almeida Gadelha, C.A.; Aragão, K.S.; de Oliveira, C.C.; Lima Mota, M.R.; Gomes, R.C.; de Freitas Pires, A.; Toyama, M.H.; de Oliveira Toyama, D.; de Alencar, N.M.N.; *et al.* Purification and biological effects of *Araucaria angustifolia* (Araucariaceae) seed lectin. *Biochem. Biophys. Res. Commun.* **2006**, *350*, 1050–1055.
202. Trindade, M.B.; Lopes, J.L.S.; Soares-Costa, A.; Monteiro-Moreira, A.C.; Moreira, R.A.; Oliva, M.L. V.; Beltramini, L.M. Structural characterization of novel chitin-binding lectins from the genus *Artocarpus* and their antifungal activity. *Biochim. Biophys. Acta - Proteins Proteomics* **2006**, *1764*, 146–152.
203. Díaz, C.L.; Logman, T.J.J.; Stam, H.C.; Kijne, J.W. Sugar-binding activity of pea lectin expressed in white clover hairy roots. *Plant Physiol* **1995**, *109*, 1167–1177.
204. Suzuki, T.; Amano, Y.; Fujita, M.; Kobayashi, Y.; Dohra, H.; Hirai, H.; Murata, T.; Usui, T.; Morita, T.; Kawagishi, H. Purification, characterization, and cDNA cloning of a lectin from the mushroom *Pleurocybella porrigens*. *Biosci. Biotechnol. Biochem.* **2009**, *73*, 702–709.
205. Kase, T.; Suzuki, Y.; Kawai, T.; Sakamoto, T.; Ohtani, K.; Eda, S.; Maeda, A.; Okuno, Y. Human mannan-binding lectin inhibits the infection of influenza A virus. *Immunology* **1999**, *97*, 385–392.
206. Jensenius, J.C.; Jensen, P.H.; McGuire, K.; Larsen, J.L.; Thiel, S. Recombinant mannan-binding lectin (MBL) for therapy. **2003**, 763–767.
207. Zhang, W.; van Eijk, M.; Guo, H.; van Dijk, A.; Bleijerveld, O.B.; Verheije, M.H.; Wang, G.; Haagsman, H.P.; Veldhuizen, E.J.A. Expression and characterization of recombinant chicken mannose binding lectin.

- Immunobiology* **2017**, *222*, 518–528.
208. Argayosa, A.M.; Bernal, R.A.D.; Luczon, A.U.; Arboleda, J.S. Characterization of mannose-binding protein isolated from the African catfish (*Clarias gariepinus* B.) serum. *Aquaculture* **2011**, *310*, 274–280.
209. Naeem, A.; Ahmad, E.; Ashraf, M.T.; Khan, R.H. Purification and characterization of mannose/glucose-specific lectin from seeds of *Trigonella foenumgraecum*. *Biochem.* **2007**, *72*, 44–48.
210. Ourth, D.D.; Narra, M.B.; Chung, K.T. Isolation of mannose-binding C-type lectin from *Heliothis virescens* pupae. *Biochem. Biophys. Res. Commun.* **2005**, *335*, 1085–1089.
211. Roy, I.; Sardar, M.; Gupta, M.N. Cross-linked alginate-guar gum beads as fluidized bed affinity media for purification of jacalin. *Biochem. Eng. J.* **2005**, *23*, 193–198.
212. Chen, H.-P.; Xu, L.-L. Isolation and characterization of a novel chitosan-binding protein from non-heading chinese cabbage leaves. *J. Integr. Plant Biol.* **2005**, *47*, 452–456.
213. Chen, C.; Rowley, A.F.; Newton, R.P.; Ratcliffe, N.A. Identification, purification and properties of a β -1,3-glucan-specific lectin from the serum of the cockroach, *Blaberus discoidalis* which is implicated in immune defence reactions. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **1999**, *122*, 309–319.
214. Macedo, M.L.R.; Freire, M. das G.M.; da Silva, M.B.R.; Coelho, L.C.B.B. Insecticidal action of Bauhinia monandra leaf lectin (BmoLL) against *Anagasta kuehniella* (Lepidoptera: Pyralidae), *Zabrotes subfasciatus* and *Callosobruchus maculatus* (Coleoptera: Bruchidae). *Comp. Biochem. Physiol. - A Mol. Integr. Physiol.* **2007**, *146*, 486–498.
215. Nunes, E. dos S.; de Souza, M.A.A.; Vaz, A.F. de M.; Santana, G.M. de S.; Gomes, F.S.; Coelho, L.C.B.B.; Paiva, P.M.G.; da Silva, R.M.L.; Silva-Lucca, R.A.; Oliva, M.L.V.; *et al.* Purification of a lectin with antibacterial activity

- from *Bothrops leucurus* snake venom. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **2011**, *159*, 57–63.
216. Mora, P.; Rosconi, F.; Franco Fraguas, L.; Castro-Sowinski, S. *Azospirillum brasilense* Sp7 produces an outer-membrane lectin that specifically binds to surface-exposed extracellular polysaccharide produced by the bacterium. *Arch. Microbiol.* **2008**, *189*, 519–524.
217. Moreira, R.A.; Castelo-Branco, C.C.; Monteiro, A.C.O.; Tavares, R.O.; Beltramini, L.M. Isolation and partial characterization of a lectin from *Artocarpus incisa* L. seeds. *Phytochemistry* **1998**, *47*, 1183–1188.
218. Oliveira, C.; Nicolau, A.; Teixeira, J.A.; Domingues, L. Cytotoxic effects of native and recombinant frutalin, a plant galactose-binding lectin, on Hela cervical cancer cells. *J. Biomed. Biotechnol.* **2011**, 568932.
219. Bhowal, J.; Guha, A.K.; Chatterjee, B.P. Purification and molecular characterization of a sialic acid specific lectin from the phytopathogenic fungus *Macrophomina phaseolina*. *Carbohydr. Res.* **2005**, *340*, 1973–1982.
220. Guzmán-Partida, A.M.; Robles-Burgueño, M.R.; Ortega-Nieblas, M.; Vázquez-Moreno, I. Purification and characterization of complex carbohydrate specific isolectins from wild legume seeds: *Acacia constricta* is (vinorama) highly homologous to *Phaseolus vulgaris* lectins. *Biochimie* **2004**, *86*, 335–342.
221. Matsumoto, R.; Shibata, T.F.; Kohtsuka, H.; Sekifuji, M.; Sugii, N.; Nakajima, H.; Kojima, N.; Fujii, Y.; Kawsar, S.M.A.; Yasumitsu, H.; *et al.* Glycomics of a novel type-2 N-acetyllactosamine-specific lectin purified from the feather star, *Oxycomanthus japonicus* (Pelmatozoa: Crinoidea). *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **2011**, *158*, 266–273.
222. Rittidach, W.; Pajit, N.; Utarabhand, P. Purification and characterization of a lectin from the banana shrimp *Fenneropenaeus merguensis* hemolymph. *Biochim. Biophys. Acta - Gen. Subj.* **2007**, *1770*, 106–114.

-
223. Sun, J.; Wang, L.; Wang, B.; Guo, Z.; Liu, M.; Jiang, K.; Luo, Z. Purification and characterisation of a natural lectin from the serum of the shrimp *Litopenaeus vannamei*. *Fish Shellfish Immunol.* **2007**, *23*, 292–299.
224. Yang, H.; Luo, T.; Li, F.; Li, S.; Xu, X. Purification and characterisation of a calcium-independent lectin (PjLec) from the haemolymph of the shrimp *Penaeus japonicus*. *Fish Shellfish Immunol.* **2007**, *22*, 88–97.
225. Ren, J.; Shi, J.; Kakuda, Y.; Kim, D.; Xue, S.J.; Zhao, M.; Jiang, Y. Phytohemagglutinin isolectins extracted and purified from red kidney beans and its cytotoxicity on human H9 lymphoma cell line. *Sep. Purif. Technol.* **2008**, *63*, 122–128.
226. Chumkhunthod, P.; Rodtong, S.; Lambert, S.J.; Fordham-Skelton, A.P.; Rizkallah, P.J.; Wilkinson, M.C.; Reynolds, C.D. Purification and characterization of an N-acetyl-D-galactosamine-specific lectin from the edible mushroom *Schizophyllum commune*. *Biochim. Biophys. Acta - Gen. Subj.* **2006**, *1760*, 326–332.
227. Kaur, Navjot; Dhuna, Vikram; Kamboj, Sukhdev Singh; Agrewala, Javed N; Singh, J.; Kaur, N.; Dhuna, V.; Kamboj, S.S.; Agrewala, J.; Singh, J. A novel antiproliferative and antifungal lectin from *Amaranthus viridis* linn seeds. *Protein Pept. Lett.* **2006**, *13*, 897–905.
228. Yamawaki, M.; Isobe, S.; Usui, T.; Kimura, A.; Chibao, S. Two lectins from the marine sponge *Haliichondria okadai*. *J. Biol. Chem.* **1994**, *269*, 0–4.
229. Naganuma, T.; Ogawa, T.; Hirabayashi, J.; Kasai, K.; Kamiya, H.; Muramoto, K. Isolation, characterization and molecular evolution of a novel pearl shell lectin from a marine bivalve, *Pteria penguin*. *Mol. Divers.* **2006**, *10*, 607–618.
230. Hiremath, K.Y.; Jagadeesh, N.; Belur, S.; Kulkarni, S.S.; Inamdar, S.R. A lectin with anti-microbial and anti proliferative activities from *Lantana camara*, a medicinal plant. *Protein Expr. Purif.* **2020**, *170*, 105574.

-
231. Mo, H.; Meah, Y.; Moore, J.G.; Goldstein, I.J. Purification and characterization of *Dolichos lablab* lectin. *Glycobiology* **1999**, *9*, 173–179.
232. Takeuchi, T.; Nishiyama, K.; Yamada, A.; Tamura, M.; Takahashi, H.; Natsugari, H.; Aikawa, J.I.; Kojima-Aikawa, K.; Arata, Y.; Kasai, K.I. *Caenorhabditis elegans* proteins captured by immobilized Gal β 1-4Fuc disaccharide units: Assignment of 3 annexins. *Carbohydr. Res.* **2011**, *346*, 1837–1841.
233. Watanabe, T.; Matsuo, I.; Maruyama, J.I.; Kitamoto, K.; Ito, Y. Identification and characterization of an intracellular lectin, calnexin, from *Aspergillus oryzae* using N-glycan-conjugated beads. *Biosci. Biotechnol. Biochem.* **2007**, *71*, 2688–2696.
234. Bhat, G.G.; Shetty, K.N.; Nagre, N.N.; Neekhara, V. V.; Lingaraju, S.; Bhat, R.S.; Inamdar, S.R.; Suguna, K.; Swamy, B.M. Purification, characterization and molecular cloning of a monocot mannose-binding lectin from *Remusatia vivipara* with nematocidal activity. *Glycoconj. J.* **2010**, *27*, 309–320.
235. Nagre, N.N.; Chachadi, V.B.; Sundaram, P.M.; Naik, R.S.; Pujari, R.; Shastry, P.; Swamy, B.M.; Inamdar, S.R. A potent mitogenic lectin from the mycelia of a phytopathogenic fungus, *Rhizoctonia bataticola*, with complex sugar specificity and cytotoxic effect on human ovarian cancer cells. *Glycoconj. J.* **2010**, *27*, 375–386.
236. Vega, N.; Pérez, G. Isolation and characterisation of a *Salvia bogotensis* seed lectin specific for the Tn antigen. *Phytochemistry* **2006**, *67*, 347–355.
237. de Santana, M.A.; Santos, A.M.C.; Oliveira, M.E.; de Oliveira, J.S.; Baba, E.H.; Santoro, M.M.; de Andrade, M.H.G. A novel and efficient and low-cost methodology for purification of *Macrotyloma axillare* (Leguminosae) seed lectin. *Int. J. Biol. Macromol.* **2008**, *43*, 352–358.
238. Visini, R.; Jin, X.; Bergmann, M.; Michaud, G.; Pertici, F.; Fu, O.; Pukin, A.; Branson, T.R.; Thies-Weesie, D.M.E.; Kemmink, J.; *et al.* Structural insight





- into multivalent galactoside binding to *Pseudomonas aeruginosa* lectin LecA. *ACS Chem. Biol.* **2015**, *10*, 2455–2462.
239. Surolia, A.; Bishayee, S.; Ahmad, A.; Balasubramanian, K.A.; Thambi-Dorai, D.; Podder, S.K.; Bachhawat, B.K. Studies on the interaction of concanavalin A with glycoproteins. *Adv. Exp. Med. Biol.* **1975**, *55*, 95–115.

APPENDIX

During the Ph.D., I also worked on three additional projects that are not in direct line with the topic of the thesis. One of these projects has already been already published and it is included in this annex.

Article

Recombinant Lectin from Tepary Bean (*Phaseolus acutifolius*) with Specific Recognition for Cancer-Associated Glycans: Production, Structural Characterization, and Target Identification

Dania Martínez-Alarcón ^{1,2} , Annabelle Varrot ² , Elaine Fitches ³ , John A. Gatehouse ³,
Mín Cao ³, Prashant Pyati ³, Alejandro Blanco-Labra ^{1,*} and Teresa Garcia-Gasca ^{4,*} 

¹ Centro de Investigación y de Estudios Avanzados Unidad Irapuato, Departamento de Biotecnología y Bioquímica, Irapuato 36821, Guanajuato, Mexico; dania.martinez.alarcon@gmail.com

² University of Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France; annabelle.varrot@cermav.cnrs.fr

³ Department of Biosciences, Durham University, Durham DH1 3LE, UK; e.c.fitches@durham.ac.uk (E.F.); j.a.gatehouse@durham.ac.uk (J.A.G.); mincao.chn@gmail.com (M.C.); prashpya@googlemail.com (P.P.)

⁴ Facultad de Ciencias Naturales, Universidad Autónoma de Querétaro, Santiago de Querétaro 76230, Querétaro, Mexico

* Correspondence: tggasca@uaq.edu.mx (T.G.-G.); alejandrobancolabra@gmail.com (A.B.-L.); Tel.: +52-442-192-1200 (ext. 5308) (T.G.-G.); +52-461-613-9640 (A.B.-L.)

Received: 22 March 2020; Accepted: 20 April 2020; Published: 23 April 2020



Abstract: Herein, we report the production of a recombinant Tepary bean lectin (*rTBL-1*), its three-dimensional (3D) structure, and its differential recognition for cancer-type glycoconjugates. *rTBL-1* was expressed in *Pichia pastoris*, yielding 316 mg per liter of culture, and was purified by nickel affinity chromatography. Characterization of the protein showed that *rTBL-1* is a stable 120 kDa homo-tetramer folded as a canonical leguminous lectin with two divalent cations (Ca^{2+} and Mn^{2+}) attached to each subunit, confirmed in its 3D structure solved by X-ray diffraction at 1.9 Å resolution. Monomers also presented a ~2.5 kDa *N*-linked glycan located on the opposite face of the binding pocket. It does not participate in carbohydrate recognition but contributes to the stabilization of the interfaces between protomers. Screening for potential *rTBL-1* targets by glycan array identified 14 positive binders, all of which correspond to β 1-6 branched *N*-glycans' characteristics of cancer cells. The presence of α 1-6 core fucose, also tumor-associated, improved carbohydrate recognition. *rTBL-1* affinity for a broad spectrum of mono- and disaccharides was evaluated by isothermal titration calorimetry (ITC); however, no interaction was detected, corroborating that carbohydrate recognition is highly specific and requires larger ligands for binding. This would explain the differential recognition between healthy and cancer cells by Tepary bean lectins.

Keywords: recombinant lectins; Tepary bean; *Pichia pastoris*; MGAT5; glycan array; structure; cancer

1. Introduction

Lectins are proteins that bind to carbohydrates, either free or bound to cell membranes as part of glycoproteins, glycolipids, or polysaccharides. Since these interactions are reversible and highly specific, lectins have been widely used to elucidate alterations in the composition of carbohydrates during pathological processes such as cancer, where the cell glycosylation machinery is frequently altered, leading to the exposure of aberrant carbohydrates on cellular surfaces [1–5].

Given that cancer is a microevolutionary process, anomalous carbohydrate edition does not occur randomly; in fact, there is a limited subset of changes that correlate with malignant

transformation [5]. These changes vary depending on the cancer type and are often related to tumor progression and lifespan [4]. Some of the most common tumor-associated glycan changes described to date include (1) increased β 1-4 branched tetra-antennary *N*-glycans [4,5]; (2) α 2-6 sialylation on the outer poly-*N*-acetylglucosamines (LacNAc) of *N*-glycans [6]; (3) increase of α 1-6 fucosylation [7,8]; (4) overexpression of mucins and truncated *O*-glycans [9]; (5) altered expression of blood groups and sialyl-Lewis antigens [10]; (6) increase in bisecting GlcNAc [11]; (7) overexpression of hyaluronan [12,13]; and (8) increased β 1-6 branching of *N*-glycans [4,5]. The latter is one of the most important glycan modifications in colorectal cancer cells. It is derived from an upregulation of the *N*-acetylglucosaminyltransferase V (MGAT5) by oncogenic transcription factors from the RAS–RAF–MAPK signaling pathways [4], which are highly associated with cancer metastasis [11].

A possible mechanism by which MGAT5 upregulation enhances cancer progression is by causing a high frequency of branched *N*-glycans on the extracellular domains of Receptor Tyrosine Kinases (RTKs) [14,15]. RTKs are the second main type of cell surface receptors, and they are critical actors in cancer through regulation of multiple cellular processes such as growth, migration, proliferation, differentiation, and apoptosis [15]. Activation of these receptors occurs through cross-linked phosphorylation after oligomerization induced by ligand binding. Their action is regulated by a fine balance between their synthesis/exposure and internalization/degradation. Alterations in the glycosylation pattern of RTKs lead to the disruption of this balance, allowing, for example, longer retention on the surface of cancer cells. Such is the case of the epidermal growth factor receptor (EGFR), of which recognition by galectins promotes lattice formation, enhances phosphorylation, and causes a high rate of cell proliferation via signal transduction [14,16]. Consequently, targeting RTK recognition represents an interesting therapeutic strategy to block cell growth and may provide new strategies for unique and combinatorial therapies.

It has been shown that a semi-pure fraction of lectins from Tepary bean (TBLF), mainly composed of two different glycoproteins (TBL-1 and TBL-2), can recognize a broad spectrum of human cancer cells. It induces apoptosis at low concentrations; the most sensitive are colon and breast cancer cells. TBLF lectins are highly specific in distinguishing between normal and cancer cells; thus, as a consequence, a differential cytotoxic effect is displayed [17] by apoptosis induction and cell cycle arrest [18]. These lectins are particularly resistant and can cross the gastrointestinal tract without being degraded by pH conditions and/or digestive enzymes [19]. Furthermore, TBLF presents low toxicity and inhibits early tumorigenesis in rats with chemically induced colorectal cancer [19,20]. When administered intragastrically to rats, TBLF has been shown to have a stimulatory effect on the immune system and displays few side effects, such as atrophy in small intestine villus and colonic crypts and pancreatic hypertrophy, that can be partially reverted after a two-week rest period [21]. These results suggest the potential therapeutic use of Tepary bean lectins against colon cancer.

The main drawback of the use of TBLF as an anticancer therapy is that its purification from seeds is slow and expensive and provides low yields. Its purification comprises a six-step train, where large amounts of methanol, chloroform, and ammonium sulphate are needed, among others. One of the major limitations of this process is that the gel filtration columns, which are approximately 2 m long, are run by gravity and that only a small amount of the sample can be processed at a time, resulting in a very time-consuming step. Our general estimates indicate that obtaining 1 g of semi-pure TBLF requires 30–41 weeks of work for one person, 1 kg of beans, 15 L of chloroform, 5 L of methanol, and 1.7 kg of ammonium sulphate, among other reagents on a smaller scale. When a high-purity sample is required, it is necessary to perform an additional step of purification by ion exchange chromatography and HPLC, which further reduces the performance.

TBLF lectins are complex glycoproteins, and preliminary data suggest that posttranslational modifications could influence folding and, consequently, could alter their biological function. Previously, we reported the cisgenic expression of a TBLF lectin using a strategy that allowed its secretion through root exudates of genetically modified Tepary bean plants [22]. However, complete plant regeneration is a complex process and the yield obtained using this approach (~21 μ g of protein per gram of dry

root) was insufficient to meet the protein requirements for in vivo assays, requiring investigation of alternative means of production.

Since bacteria lack organelles for posttranslational processing, yeasts are the most common system for the expression of glycoproteins, such as legume lectins. Initially, the most promising system for this purpose was *Saccharomyces cerevisiae* due the knowledge of its genetic manipulation [23,24]. However, *S. cerevisiae* has the disadvantage of producing high mannose *N*-glycans with outer chains of typically 50–150 residues in length (~9 to ~27 kDa), which are highly antigenic for humans [23,25]. Thus, the methylotrophic yeast *Pichia pastoris*, which does not appear to add extra α 1-3 mannose residues or to carry out hyperglycosylation on its *N*-glycans, has been successfully used as expression system for leguminous lectins [26,27]. In addition, recombinant proteins produced by *Pichia* can be directly recovered from the culture media by coupling them to the *S. cerevisiae* α -factor, which is a signal peptide that redirects the protein to the secretory pathway and is subsequently cleaved in the Golgi by *Pichia* endopeptidases [24].

In this work, we present the production of a recombinant lectin from Tepary bean (rTBL-1) by using *P. pastoris* engineering, its biochemical characterization, crystallographic structure, and an analysis of its specificity for cancer-type glycans.

2. Materials and Methods

2.1. Materials

Chemicals and reagents were of analytical grade and were supplied by Sigma (Sigma-Aldrich, Gillingham, UK) or BDH Chemical Company, unless stated otherwise. Restriction enzymes were supplied by Fermentas (Ontario, Canada) and antibodies from Invitrogen, Thermo Fisher Scientific (Carlsbad, CA, USA), and Bio-Rad Laboratories (Hercules, CA, USA).

2.2. Strains

Gene constructs were prepared in the TOP10 strain of *Escherichia coli*, and the production of proteins was carried out using the *P. pastoris* protease-deficient strain SMD1168H.

2.3. Plasmids

pGAPZ α B plasmid was provided by Invitrogen Life Technologies (Carlsbad, CA, USA). Briefly, this commercial plasmid allows the transformation of *P. pastoris* by recombining the polylinker flanking sequences. Expression of the cloned genes is under regulation of the constitutive Glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter. This vector also contains replication sites for both *E. coli* and *P. pastoris*, a zeocin (Zn) resistance marker, and a *S. cerevisiae* signal peptide for protein secretion (α -factor) fused to the *N*-terminal of the protein and subsequently cleaved by endopeptidases Kex2 and Ste13 in the Golgi apparatus.

2.4. Tepary Bean Lectin Fraction (TBLF)

TBLF was obtained by purification from Tepary bean seeds. Briefly, beans were finely ground and the resulting flour was defatted by several washes with a mixture of CHCl₃/MeOH (3:1) until the filtering was clear. Then, total proteins were extracted by using Tris-HCl pH 8.0 at 4 °C, and a sequential precipitation was performed with ammonium sulphate at 40% and 60% (*w/v*) saturation, followed by intensive dialysis. Finally, the protein extract was separated by molecular size exclusion chromatography using a G-75 Sephadex column [17,18].

2.5. Vector Construction

The coding sequence of rTBL-1 was amplified with primers (forward 5'-TATCTGCAGCATCAGCCAACGACATCTC-3' and reverse 5'-ATATCTAGACTA ATGATGATGA TGATGATGATGATTC-3'), where PstI and XbaI restriction sites to the 5' and 3' end of the sequence

were added (underlined sequences). Next, both the amplicon and pGAPZ α B vector were digested with PstI and XbaI enzymes for 3 h at 37 °C and then purified from gel electrophoresis using the commercial kit of BioLabs (Hitchin, UK) "Monarch DNA Gel Extraction." The products obtained were ligated using T4 ligase in a 3:1 insert/vector ratio. The reaction was incubated at 25 °C for 1 h, and the product was used to transform the electrocompetent cells of *E. coli* TOP10 using an electroporator with a resistance of 100 Ohms, capacitance of 125 μ Fa, and 1.8 Volts. The transformed cells were recovered for growth on plates of Luria–Bertani (LB) medium with 25 μ g/mL zeocin as a selection agent and incubated at 37 °C for 24 h. The transformants were screened by colony PCR using a set of primers designed to hybridize the flanking sequences of the pGAPZ α B polylinker (pGAP Forward 5'-GTCCCTATTTC AATCAATTGAA-3' and AOX1 5'-GCAAATGGCATTCTGACATCC3-3'). Plasmid extraction was carried out from one of the positive colonies, and the in-frame insertion and the correct coding sequence were verified by digestion and DNA sequencing.

2.6. Transformation of *Pichia pastoris*

Once the sequence was confirmed, the plasmid was digested with AvrII enzyme at 37 °C for 16 h and complete DNA linearization was confirmed by nucleic acid electrophoresis. Subsequently, the DNA was precipitated (100% ethanol and 5 M ammonium acetate at –20 °C for 16 h), centrifuged, and resuspended in 20 μ L of nuclease-free sterile water, and then, a concentration >5 μ g/ μ L was confirmed by NanoDrop spectrophotometry (NanoDrop 2000c from Thermo Fisher Scientific; Hercules, CA, USA) and electrophoresis. This material was used to transform *P. pastoris* by means of the Pichia EasyCompTM Transformation Kit (Thermo Fisher Scientific; Pittsburgh, PA, USA), according to the manufacturer's protocol. Transformed cells were recovered for growth on plates of Yeast Extract–Peptone–Glycerol (YPG) medium with 50 μ g/mL of zeocin and were incubated at 30 °C for 48 h.

2.7. Production

Ten of the transformed *P. pastoris* colonies were inoculated in 10 mL of liquid YPG medium containing 25 μ g/mL of zeocin. Cultures were allowed to grow at 30 °C with shaking for 48 h, and then, the supernatants were recovered by centrifugation at 11,800 \times *g* for 10 min. Aliquots of 25 μ L of each of the supernatants were used for SDS-PAGE electrophoresis [28], followed by semi-dry transfer (ATTO blotter, Tokyo, Japan) for 1 h at constant voltage (10 V) to nitrocellulose membranes previously soaked in 1X TBS. The membranes were stained with Ponceau red to indicate the position of the molecular weight markers and then subsequently immersed in blocking solution (defatted milk powder 5% *w/v* in 0.1% *v/v* Tween TBS) for 1 h. After blocking, the membranes were incubated overnight with a 6XHis Tag mouse monoclonal antibody (Invitrogen, Carlsbad, CA, USA) in a 1:1000 dilution prepared in blocking solution. After 16 h, the primary antibodies were removed by several washes with blocking solution, and then, the membranes were incubated for 2 h in blocking solution containing HRP-conjugated secondary goat anti-mouse antibodies (Bio-Rad, Hercules, CA, USA). Finally, immunoreactivity was visualized by chemiluminescence.

2.8. Heterologous Expression and Purification of rTBL-1

For protein production, *P. pastoris* (SMD1168H) cells expressing recombinant rTBL-1 were grown in an Applikon ez-control laboratory fermenter (7.5 L vessel) as described previously [29], except that the pH was maintained at 5.0. The secreted proteins were separated from cells by centrifugation (30 min, 7500 \times *g*, 4 °C), and the supernatant was clarified by sequential filtering with 2.7 and 0.7 μ m glass fiber filters (Whatmann, Maidstone, UK). Recombinant rTBL-1 was purified from the supernatant by nickel-affinity chromatography by using HisTrap HP columns (GE Healthcare, Maidstone, UK) dialysed and freeze-dried as described [30]. The protein contents in lyophilized samples were determined from SDS-PAGE gels based on bands corresponding to intact proteins, which were compared to GNA (Sigma) standards by visual inspection, and by capturing an image of the de-stained gel using a

commercial flat-bed scanner; image analysis was carried out with a custom-written software program (ProQuantify version 2.0 supplied by Rodrigo Guerrero).

2.9. Molecular Size Exclusion Chromatography

Molecular size exclusion chromatography (SEC) was performed using a High-Resolution ENrich™ SEC 650 column (Bio-Rad, Marnes-la-Coquette, France) on the NGC™ chromatography system (Bio-Rad, Marnes-la-Coquette, France). Prior to assay, the column was calibrated using the gel filtration standard #15119001 (Bio-Rad, Marnes-la-Coquette, France), according to the supplier's instructions, and protein samples at 10 mg/mL were centrifuged for 30 min at 12,000 rpm. The column was equilibrated with 50 mL of buffer D (20 mM (2-(*N*-morpholino)ethanesulfonic acid, MES) pH 6.5 and 100 mM NaCl), and 200 µL of the sample was injected into the system, followed by 30 mL isocratic elution on buffer D. The fractions were monitored by absorbance at 280 nm and collected every 0.5 mL.

2.10. Thermal Shift Assay (TSA)

The thermal stability of rTBL-1 was analyzed by TSA using a MiniOpticon real-time PCR system (Bio-Rad, Marnes-la-Coquette, France). Prior to assay, buffer stocks at 100 mM and a mixture containing 70 µL of rTBL-1 at 1 mg/mL, 7 µL of 500X Sypro Orange (Sigma-Aldrich, Saint Quentin Fallavier, France), and 63 µL of H₂O were prepared. Then, 7.5 µL of H₂O, 12.5 µL of the corresponding buffer and 5 µL of the protein/Sypro mixture were mixed in 96-well PCR microplates. The heat exchange test was then carried out from 20 to 100 °C with a heating rate of 1 °C/min. Fluorescence intensity was measured with Ex/Em: 490/530 nm, and data processing was performed using CFX Manager software version 1.6.

2.11. Determination of the Polypeptidic Sequence

The product of the translation as well as the most frequent C- and N-terminal residues were determined by nano-LC-MS/MS, depending on standard data of tryptic and chemotryptic peptides generated from the digestion of the samples, where a method of acquisition of data-dependent mass spectrometers was used. Along a reverse phase gradient used for the separation of peptides, the MS scans were followed using an MS/MS spectrum from the selected precursors. The mobile exclusion windows prevented the reacquisition of the same precursors during a fixed period and allowed to collect less-abundant ion data. For detection of glycosylated peptides, the same methodology was followed on tryptic peptides of deglycosylated protein.

2.12. Identification of the Glycosylations Present in the Recombinant Lectin

Proteins were subjected to glycan digestion using combinations of five different specific glycosidases for O- and N-glycosylations using the Glycoprotein Deglycosylation kit (EMD Millipore, Danvers, MA, USA). The antennas' size was determined through the retention factor (RF) of samples on SDS-PAGE, both before and after deglycosylation. To validate these results, the periodic acid-Schiff reagent (PAS) staining technique was used (Sigma Chemical, St. Louis, MO, USA), as described by the supplier. The glycosidases were (1) N-Glycosidase F: this enzyme cleaves all N-linked oligosaccharides, unless their core is α1-3 fucosylated; (2) Endo-α-N-acetylgalactosaminidase that separates common nucleus (Galβ1-3GalNAc) from all O-glycosylations (it only works when no additional sugars adhere to the main nucleus); (3) α2-3,6,8,9-neuraminidase that cuts all sialic acids that may be attached to the common core of the O-glycosylations; (4) β1-4galactosidase that releases only non-reducing terminal galactose, bound in β1-4 to O-glycosylations; and (5) β-N-acetylglucosaminidase that cleaves all GlcNAc residues bound to β-non-reducing terminal O-glycosylations.

2.13. Crystallization and Data Collection

Crystal screening was performed using the hanging-drop vapor diffusion technique by mixing equal volumes of pure protein at 5 mg/mL and precipitant solutions from the BCS Screen (Molecular Dimensions, Newmarket, Suffolk, UK). Drops were incubated at 19 °C until crystals appeared. A subsequent optimization of positive conditions was carried out, and crystals suitable for X-ray diffraction analysis were obtained under a solution containing 100 mM Tris-HCl buffer pH 7.5 and 18% (v/v) PEG Smear Low (mix containing PEG 400, PEG 550 MME, PEG 600, and PEG 1000). The crystals were soaked in mother liquor supplemented with 30% (v/v) of PEG Smear Low (Molecular Dimensions, Newmarket, Suffolk, UK) prior to flash cooling in liquid nitrogen. Data collection was performed on the beamline Proxima-1 at SOLEIL Synchrotron, Saint Aubin, France using an Eiger X 16M detector.

2.14. Structure Determination

Diffraction data were processed using XDS [31] and converted to structure factors using the CCP4 program package v.6.1 [32] with 5% of the data reserved for R_{free} calculation. The structure of the *r*TBL-1 was solved by molecular replacement using Phaser MR v.2.5 [33], utilizing the tetramer coordinates of PHA:L (PDB entry 1 FAT A) [34] as the search model after trimming with Chainsaw. Restrain refinement was performed using REFMAC5 [35] alternated with manual model building in Coot v.0.7 [36]. The sugar residues and other compounds that were present were placed manually using Coot. Water molecules were added automatically in Coot and checked manually. The final structure was validated using the validation server from the Protein Data Bank (PDB) ([https://validate-rcsb-1 wwpdb.org/](https://validate.rcsb-1 wwpdb.org/)) and was deposited in the PDB as entry 6TT9.

2.15. Glycan Array

*r*TBL-1 was labeled with fluorescein isothiocyanate (FITC) (Sigma-Aldrich, St Louis, MO) according to the supplier's instructions with slight modifications. Briefly, 2 mg of protein was dissolved in 1 mL of buffer E (100 mM Na_2CO_3 and 100 mM NaCl, pH 9); then, 40 μL of FITC at 1 mg/mL in dimethyl sulfoxide (DMSO) was gradually added to the protein solution, and the mixture was gently stirred at room temperature overnight. The next day, the solution was supplemented with NH_4Cl to a final concentration of 50 mM and free FITC was removed using a PD10 desalting column (GE Healthcare Life Sciences) with PBS as the mobile phase. The protein concentration was determined at ABS_{280} and FITC at ABS_{490} using a NanoDropTM 200 (Thermo Fisher Scientific), and the Fluorescein/Protein molar ratio (F/P) was estimated by the following formula:

$$\text{Molar } \frac{\text{F}}{\text{P}} = \frac{\text{MW}}{389} \times \frac{\frac{A_{495}}{195}}{[A_{280} - (0.35 \times A_{495})] E^{0.1\%}}$$

where MW is the protein molecular weight, 389 is the FITC molecular weight, 195 is the absorption $E_{0.1\%}$ of bound FITC at 490 nm and pH 13.0, $(0.35 \times A_{495})$ is the correction factor due to the absorbance of FITC at 280 nm, and $E_{0.1\%}$ is the absorption at 280 nm of a protein at 1.0 mg/mL (an ideal F/P should be $0.3 > 1$).

Labeled *r*TBL-1 was sent to the Consortium for Functional Glycomics (CFG) (Boston, MA, USA), and the binding properties were assayed at 5 and 50 $\mu\text{g}/\text{mL}$ on a "Mammalian Glycan Array version 5.4" that contain 585 glycans in replicates of six. The highest and lowest signals of each set of replicates were eliminated, and the average of the remaining data was normalized to the percentages of the highest relative fluorescent units (RFU) value for each analysis; finally, the percentages for each glycan were averaged at different lectin concentrations.

2.16. Isothermal Titration Calorimetry (ITC)

Experiments were performed using a Microcal ITC200 calorimeter (Malvern Panalytical, Malvern, UK) with 40 μ L of ligand and 200 μ L of rTBL-1 100 μ M. Both the protein and the ligand were dissolved in MES buffer pH 6.5, and the sugar was gradually added to the sample cell by 2 μ L injections in a range of 120 s while stirring at 1000 rpm. The experimental data were adjusted to a theoretical titration curve by the Origin ITC analysis software.

3. Results

3.1. Production of Recombinant TBL1 (rTBL-1)

A synthetic gene encoding for the mature sequence of TBL-1 was inserted into the pGAPZ α B expression vector under regulation of the constitutive GAP promoter by merging its *N*-terminus to the yeast α -factor sequence using XbaI and PstI restriction enzymes (Figure 1A). Plasmid DNA from a sequence-verified clone was then linearized and the expression cassette integrated at the *GAPDH* locus of *P. pastoris* by DNA recombination. Western blotting of shake flask culture supernatants showed that eight of the ten selected colonies expressed and secreted the lectin, with at least seven colonies displaying relatively high expression levels (Figure 1B). The expression under shake flask conditions was estimated to be 2–32 μ g/mL, with colonies 5 and 10 being the lowest and highest producers, respectively. Production was then scaled in a 7-L bench top fermenter, where a final yield of 316 ± 0.27 mg recombinant lectin/L of culture supernatant was obtained using colony 10 as the pre-culture. rTBL-1 was purified by nickel affinity chromatography, and fractions were collected after elution with 200 mM imidazole. The analysis of purified rTBL-1 by SDS-PAGE electrophoresis showed the presence of a single band of \sim 30 kDa in gels stained for total proteins (Figure 1C). No additional protein bands or degradation products were identified, even when the wells were loaded with \sim 50 μ g of protein.

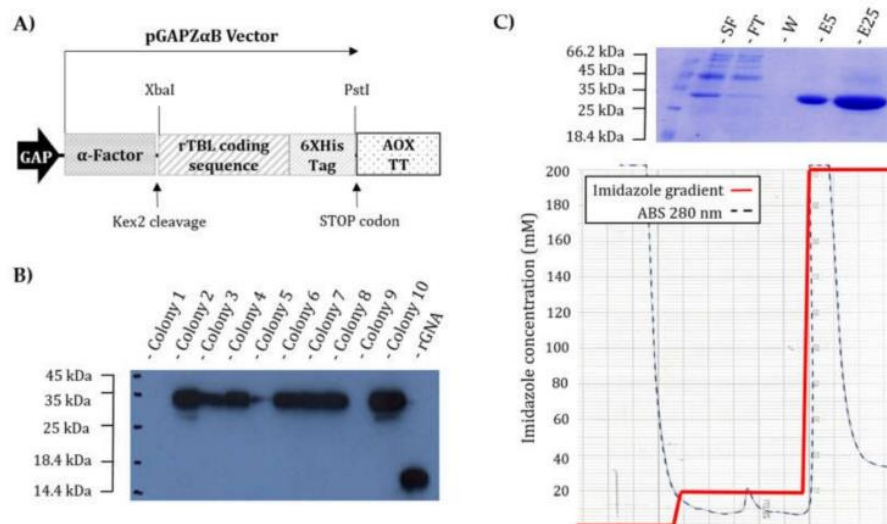


Figure 1. rTBL-1 production and purification: (A) Schematic representation of the construct encoding for rTBL-1 in the vector pGAP α B-rTBL-1. The α -factor pre-pro sequence directs the expressed protein to the yeast secretory pathway, enabling purification from fermented culture supernatants. (B) Western blot screening with anti-His antibodies to identify yeast clones expressing rTBL-1: lane 1, ladder; lanes 2–11, transformed colonies; and lane 12, 500 ng of recombinant histidine-tagged GNA as a positive control. (C) rTBL-1 purification process: top: SDS-PAGE electrophoretic profiles with gel stained for total protein. Lane 1, ladder; lane 2, 25 μ L of culture supernatant; lane 3, 25 μ L of flow through; lane 4, 25 μ L of the wash with 10 mM imidazole buffer; and lanes 5 and 6, 5 and 25 μ L of elution with 200 mM imidazole buffer, respectively. Bottom: chromatogram with imidazole gradient and Abs₂₈₀ represented with red and black dotted lines, respectively.

3.2. *rTBL-1* Characterization

A thermal shift assay (TSA) was used to evaluate the stability of the protein through a temperature gradient from 20 to 100 °C using 26 different buffers in the pH range of 5–10. The most stable conditions were obtained with MES, Tris, and HEPES buffers in a pH range of 6–7.5, with melting temperatures of 76, 70, and 75 °C, respectively. Citrate, phosphate, and malonate buffers in the same pH range displayed melting temperatures <45 °C. Interestingly, *rTBL-1* was stable up to 55 °C at pH 8.5 in Tris-HCl buffer, whereas it was denatured at a temperature <25 °C in 3-(Cyclohexylamino)-1-propanesulfonic acid (CAPS) and Bicine buffers at the same pH (Figure 2A).

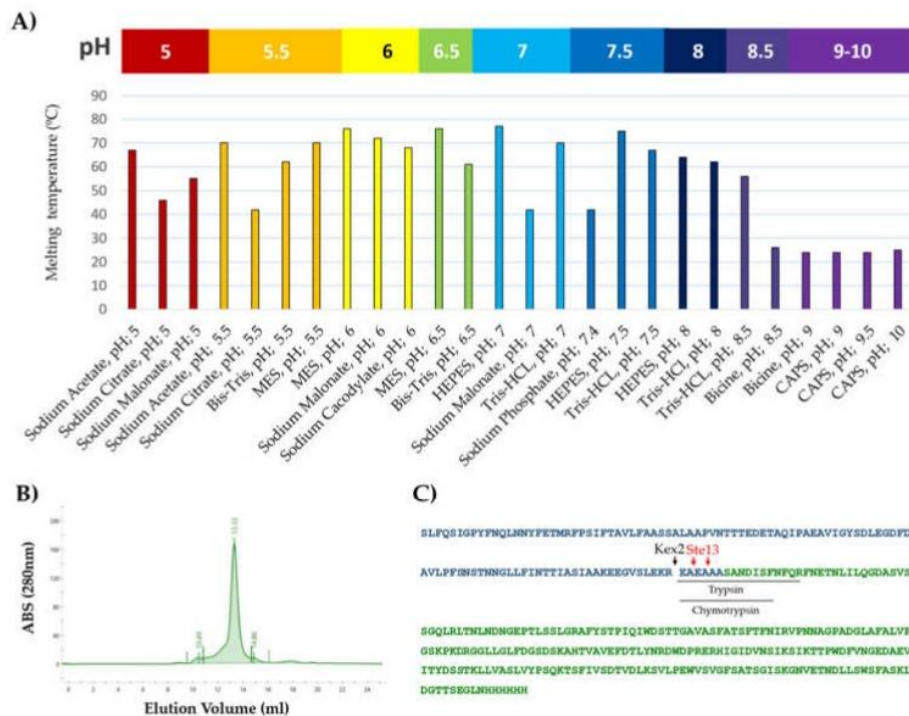


Figure 2. Characterization of *rTBL-1*: (A) Melting temperatures of *rTBL-1* obtained through thermal shift assay (TSA). A temperature gradient from 20 to 100 °C was applied under 26 different buffer conditions in a pH range from 5 to 10. (B) Size exclusion chromatogram of *rTBL-1*: The column was calibrated using the gel filtration standard #15119001 (Bio-Rad). (C) Sequence of *rTBL-1* by mass spectrometry; green depicts *rTBL-1* sequence; blue represents α -factor sequence; black lines indicate the most abundant tryptic and chymotryptic N-terminal peptides found in the digested samples; and cleavage sites for Kex2 and Ste13 are indicated by black and red arrows, respectively.

To evaluate if the heterologous expression of *rTBL-1* could affect its ability to form tetramers, its molecular weight was determined by gel filtration using a high-resolution size exclusion column, ENrich™ SEC 650 (Bio-Rad). The elution profile shows a single peak corresponding to a protein of ~120 kDa and indicating that the protein formed a homo-tetramer in the solution (Figure 2B).

The *rTBL-1* sequence and α -factor excision were verified by mass spectrometry in two samples independently digested with trypsin and chymotrypsin. The MS/MS spectra acquired after trypsin digestion matched the expected sequence of the protein. Regarding the α -factor, different peptides overlapping the Kex2p and Ste13p cleavage sites were observed, but by far, the most abundant spectra was for the N-terminal peptide EAEAAASANDISFNFQR (Figure 2C). This peptide can be generated by trypsinolysis itself or by combined trypsin and Kex2p activity, since Kex2p also cleaves at the C-terminal of arginine residues. Nevertheless, the prevalence of the spectrum matches the peptide EAEAAASANDISF, produced by chymotrypsin digestion, confirming the N-terminal sequence of

the protein. The peptide AASANDISE, derived from Ste13p cleavage, was also identified in both samples; however, its small proportion indicates that most of the protein was cleaved exclusively by the Kex2p enzyme.

3.3. Comparison of *rTBL-1* with Native *TBL-1*

According to the electrophoretic profile, *rTBL-1* is ~2.5 kDa bigger than native *TBL-1* (Figure 3A). This difference in size is mainly attributed to three factors: (1) the addition of a 6XHis tag (~840.92 Da); (2) the sequence EAEAAA at the *N*-terminal, derived from Kex2p cleavage of the α -factor (~561 Da); and (3) the presence of glycosidic antennas ~640 Da bigger than those on *TBL-1*.

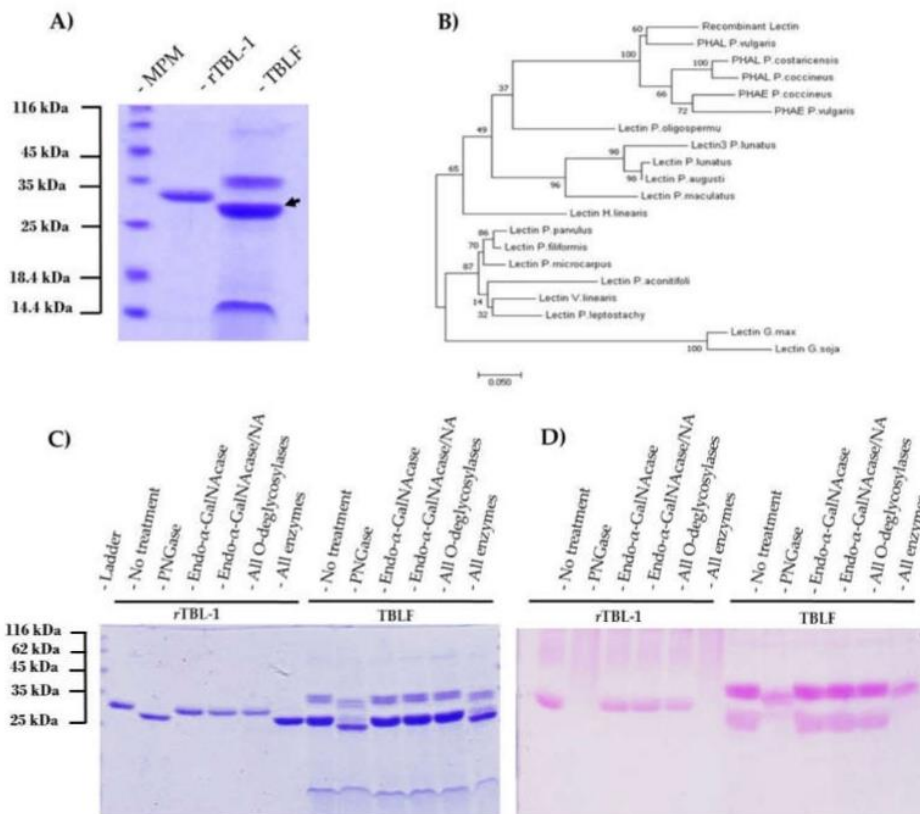


Figure 3. Molecular comparison of recombinant and native *TBL-1*: (A) SDS-PAGE electrophoresis of *rTBL-1* and Tepary Bean Lectin Fraction (TBLF) (gels stained for total protein). The black arrow shows native *TBL-1*. (B) Comparative phylogenetic sequence analysis of several legume lectins. (C) SDS-PAGE of *rTBL-1* and Tepary bean lectin fraction (TBLF) +/- deglycosylation treatments. Lane 1, *rTBL-1* no enzyme control; lanes 2–5, *rTBL-1* treated with PNGaseF, endo- α -GalNAcase, endo- α -GalNAcase, and neuraminidase (NA), respectively; lane 6, *rTBL-1* treated with all *O*-deglycosylases (β 1,4-galactosidase, endo- α -GalNAcase, NA, and β -*N*-acetylglucosaminidase); lane 7, *rTBL-1* treated with all deglycosylases (*N*-glycosidase F, β 1,4-galactosidase, endo- α -GalNAcase, NA, and β -*N*-acetylglucosaminidase); lane 8, TBLF without treatment; lanes 9–11 TBLF treated with PNGaseF, endo- α -GalNAcase, endo- α -GalNAcase, and NA, respectively; lane 12, TBLF treated with all *O*-deglycosylases (β 1,4-galactosidase, endo- α -GlcNAcase, NA, and β -*N*-acetylglucosaminidase); lane 13, TBLF treated with all deglycosylases (*N*-glycosidase-F, β 1,4-galactosidase, endo- α -GalNAcase, NA, and β -*N*-Acetylglucosaminidase). (D) SDS-PAGE of *rTBL-1* and TBLF +/- deglycosylation treatments stained with Schiff-PAS reagent for the detection of glycoproteins. Sample arrangement is as described for SDS-PAGE gel.

A comparative sequence analysis of several legume lectins with high homology to *TBL-1* was performed (Figure 3B). The taxa inferred by the evolutionary analysis showed that *TBL-1* belongs to

the group of phytohemagglutinins of the *Phaseolus* genus; more specifically, the data suggest that TBL-1 is actually a homologue of PHA:L, a leucoagglutinin from *Phaseolus vulgaris*, present in four of the five different populations of PHA heterotetramers (E4, E3L, E2L2, EL3, and L4) [37,38].

It is important to highlight that, while TBL-1 is a protein that binds carbohydrates, it is also a glycoprotein that contains covalently attached glycans. The structure and size of these carbohydrates play an important role in protein folding and, consequently, a gross alteration of its pattern during heterologous expression into *Pichia pastoris* could disrupt activity. Therefore, glycan characterization via treatment of rTBL-1 and TBLF with glycosidases, followed by SDS-PAGE and Schiff-PASS staining, was conducted (Figure 3C,D). When treated with *N*-glycosidase-F, a size reduction of ~2.48 and ~1.84 kDa for rTBL-1 and TBL-1, respectively, was observed, accompanied by the complete loss of their carbohydrate moiety. No loss of intensity or size was detected after treatment with *O*-glycosidases in any sample, suggesting that both proteins contain exclusively *N*-linked glycans. To validate the absence of *O*-linked glycans and to confirm the position of the carbohydrate antennas of rTBL-1, LC-MS/MS was performed after trypsinolysis of samples +/- deglycosylation. Peptides with confidence scores >95% in both samples were compared to identify those present only after deglycosylation, and it was found that 63.45% corresponded to peptides containing the 19-NETN-23 sequence. It is noteworthy that almost all of the present peptides showed deamidation at Asn19, known to be a product of *N*-glycosidase-F treatment.

3.4. Structure Determination

rTBL-1 was crystallized and its structure was solved by molecular replacement at 1.9 Å using the PHA:L tetramer structure as the search model (PDB entry 1FAT) (Table 1). The crystal space group was found to be P1, and the asymmetric unit contained an rTBL-1 homo-tetramer composed of two antiparallel dimers (Figure 4A). Each monomer was composed of 15 antiparallel β-strands arranged in the classical β-sandwich or jellyroll fold observed for legume lectins such as in the case of their homologous PHA-E, PHA-L, SBA, PNA, ConA, and LOL, among others. The jellyroll fold consists of (1) a “back-sheet” with six β-strands, (2) a concave seven-stranded “front-sheet”, and (3) two “stranded-top” connecting sheets. Two divalent cations, calcium Ca²⁺ and manganese Mn²⁺, are attached to each protomer through coordination with Asp131, Glu129, Asp139, His144, Leu133, and Asn135 as well as two structural water molecules per ion. They are located at the top of the “front sheet” in the vicinity of the predicted carbohydrate binding site and are indispensable for lectin activity [37]. It is remarkable that the residues found to be involved coordination bonding of cations were identical to the ones previously described for its homologous PHA-L and very similar to the ones found in ConA, where the only mismatch corresponded to the substitution of Leu by Tyr [39].

Side-by-side, canonical dimers were formed by the antiparallel alignment of the “back” sheets of protomers A/B or C/D along β-strand 1 (amino acids 3–11; Figure 4AII). The tetramer was obtained from the back-to-back association of both dimers stabilized through interactions between the β-strands 10 (amino acids 186–194) of the corresponding chains (Figure 4AIII).

In accordance with the mass spectrometry data, we observed electron density that corresponds to the *N*-glycosylation site at Asn19. This permitted to build the proximal and the distal GlcNAc of the common core of *N*-glycans in all chains as well as six and four additional residues belonging to the antennas of chains A and D, respectively (Figure 4B). Those additional residues seem to be part of the structure Glc₃Man₉GlcNAc₂, which is the precursor of all *N*-glycans in eukaryotes. No evidence of hyperglycosylation or additional manα1-3 outer chains was identified. Considering the molecular weight obtained for the complete *N*-glycan of rTBL-1 during its characterization (~2.48 kDa) and the estimated molecular weight of the *N*-glycan identified by X-ray diffraction on chain A (~1.23 kDa), we assumed that there should be another ~7 residues on the antennas of rTBL-1. These could be a mixture of mannose and glucose which is isobaric to resemble the general structure of the *N*-glycan precursor. No structural evidence was found to suggest that the binding pockets could be affected by the presence and/or size of this posttranslational modification. However, the *N*-glycan seems to play a

role in the folding of the protein and, in particular, in the stabilization of the dimer, since it favors the interfaces between the A/B and C/D chains (Figure 4AIV).

Table 1. Data collection and refinement statistics.

Data Collection				
Beamline	PX1, SOLEIL			
Wavelength (Å)	0.97918			
Space group	P1			
a, b, c (Å)	57.57, 64.47, 67.59			
α, β, γ (°)	96.35, 101.95, 97.27			
No. of monomers in asymmetric unit	4			
Resolution (Å)	32.73–1.9 (1.94–1.9)			
Total no. of reflections	451,693 (25,605)			
No. of unique reflections	73,400 (4491)			
Completeness (%)	99.7 (98.8)			
Multiplicity	6.2 (5.7)			
Mean I/ σ (I)	11 (2.8)			
R _{merge}	0.073 (0.529)			
R _{meas}	0.088 (0.649)			
CC1/2	0.997 (0.902)			
Wilson B factor (Å ²)	23.8			
Refinement				
Resolution (Å)	32.73–1.90			
No. of reflections	73,390			
No. of reflections in test set	3643			
R _{work} /R _{free}	0.169/0.213			
Rmsd bonds (Å)	0.014			
Rmsd angles (°)	1.846			
Rmsd Chiral (Å ³)	0.091			
No. Atoms/Bfac (Å ²)	Chain A	Chain B	Chain C	Chain D
Protein	1805/30.1	1779/30.7	1805/30.7	1747/31.7
Glycan	94/40.7	28/51.8	28/58.0	72/47.9
Metal ions	2/22.7	2/23.0	2/22.4	2/23.0
Water	129/37.1	131/36.5	136/37.0	124/36.5
Ramachandran				
Favored (%)	97			
Allowed (%)	3			
Outliers (%)	0			
PDB Code	6TT9			

Note: Values in parentheses are for the outer shell. PDB: Protein data bank.

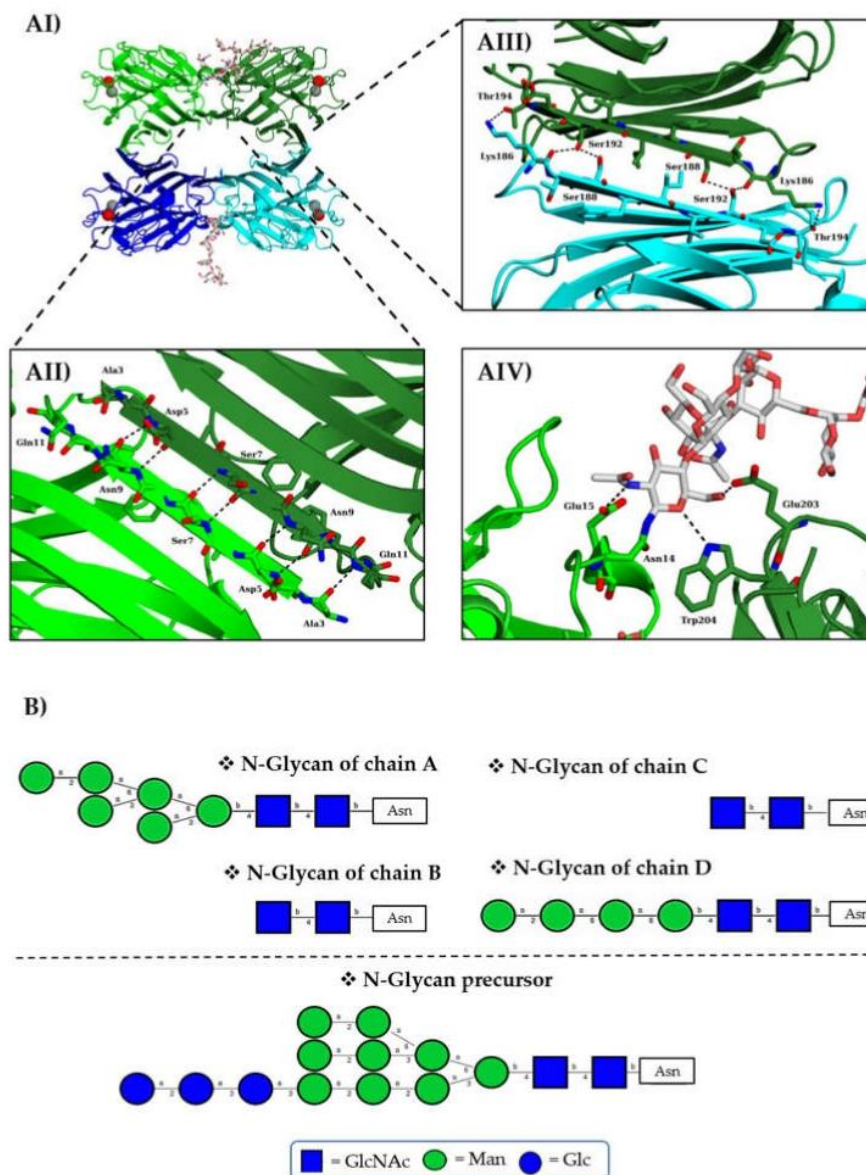


Figure 4. Structure, N-glycosylation, and interface interactions in *rTBL-1* tetramer: **(AI)** *rTBL-1* tetramer colored by chains with chains A, B, C, and D in light green, dark green, dark blue, and light blue, respectively, with Ca^{2+} depicted in red spheres and Mn^{2+} depicted in grey spheres. **(AII)** Stabilizing interactions of the interface between A and B chains. **(AIII)** Stabilizing interactions of the interface between B and D chains. **(AIV)** GlcNAc interactions in the interface of two adjacent chains. **(B)** Structure of the *N*-glycans identified in each monomer of *rTBL-1*.

3.5. *rTBL-1* Binding Properties

The glycan array was done by labeling the *rTBL-1* with fluorescein isothiocyanate (FITC), followed by analysis at 5 and 50 $\mu\text{g}/\text{mL}$ on the mammalian array “version 5.4” of the Consortium for Functional Glycomics. From the 585 glycans printed on the chip, only 14 were identified as binders (Figure 5). In all cases, *rTBL-1* recognized β 1-6 branched *N*-glycans independently of their size. Twelve of the binders contained a galactose residue linked to this branch; however, it does not seem to be a strict requirement for affinity, since glycans 2, 7, and 12 do not present this residue but were also recognized. This information contrasts with previous reports for other leucoagglutinins, as in the case of PHA:L from *P. vulgaris*, of which the minimal recognition unit for high affinity is the pentasaccharide

Gal β 1-4GlcNAc β 1-2[Gal β 1-4GlcNAc β 1-6]Man α - and where galactose seems to influence the complex behavior [36,37].

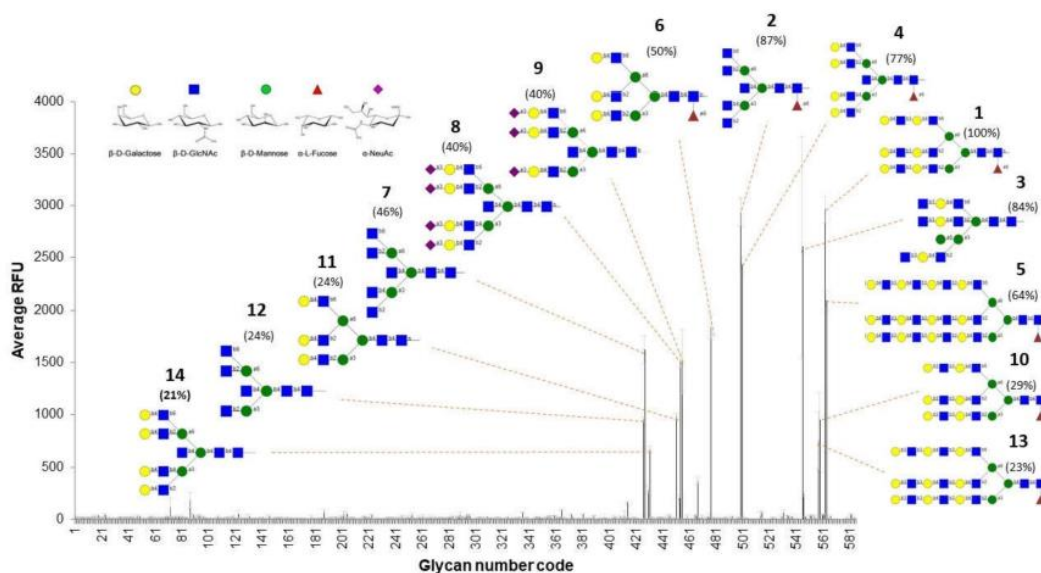


Figure 5. Glycan array results for rTBL-1 interaction with biologically relevant oligosaccharides: The bars show the average of the relative fluorescent units (RFUs) for each glycan of the matrix at 50 μ g/mL of rTBL1-FITC. Oligosaccharide structures of the top binders are shown attached to their corresponding signal. The numbers that accompany each structure indicate the ascending order of binders, starting with one as the highest RFU% (percentage of relative fluorescent units with respect to the strongest binder) for each glycan, and are indicated in parenthesis. The order of the binders and the RFU% were estimated considering the results obtained at 5 and 50 μ g/mL of rTBL1-FITC.

rTBL-1 binding does not seem to be affected when the β 1-6 branch is elongated with additional units of lactosamine (glycans 1, 5, 10, and 13) or α 2-3 sialic acid (glycans 8 and 9). However, α 1-6 fucosylated cores positively influence recognition. Binders 4 and 14 represent a clear example of this behavior, given that both have an identical structure—although binder 4 is a fucosylated core. The absence of this residue in binder 14 decreases the fluorescent signal of rTBL-1 by 56% with respect to its non-fucosylated counterpart. This same phenomenon can be observed for the glycan pairs 2/7 (41% decrease) and 6/11 (26% decrease).

The thermodynamics of binding with *N*-acetylglucosamine and other mono- and disaccharides, such as fucose, galactose, mannose, lactose, and sucrose, was approached by ITC; however, no binding was detected for any of the analyzed ligands. This, together with the glycan array matrix, shows that rTBL-1 binding is highly specific and requires complex *N*-glycans to display affinity.

These results are of the utmost importance, considering that β 1-6 branched *N*-glycans have been described as one of the most common tumor-related glycosylation events [4,5,16]. The possible mechanisms by which increased β 1-6 branched *N*-antennas enhance cancer progression include the lattice formation via galectin binding to LacNAc, leading to prolonged growth-factor signaling by EGFR [7]. Therefore, we posit that the cytotoxic effect that TBLF differentially displays between healthy and malignant cells could be related to the recognition of β 1-6 branched *N*-antennas on EGFR. This would disrupt its function, either through internalization/degradation or through the impediment of ligand binding that transduces intracellular proliferation signals [4,5,16].

4. Conclusions

A single copy insertion of a TBL-1 coding sequence into the *P. pastoris* genome leads to the high-performance production of *r*TBL-1 lectin (316 mg/mL culture). The possibility of increasing yields via the insertion of multiple gene cassettes will be addressed. Modifications derived from the heterologous expression of *r*TBL-1 did not affect its folding or biological behavior. *r*TBL-1 is glycosylated, which does not interfere with binding to other glycans but does contribute to the structural maintenance of the tetramer, as noted in its resolved crystal structure. The absence or the elongation of the glycosylation, if produced in other heterologous expression systems, could affect the overall folding and functionality of the protein. *r*TBL-1 recognizes β 1-6 branched *N*-glycans, which are overrepresented in several types of cancer, including colon cancer. The fucosylated core enhances *r*TBL-1–ligand affinity by an unknown mechanism.

We found an efficient system for the production of a recombinant lectin with anticancer potential in adequate yield to proceed with *in vitro* cytotoxicity assays on cancer cells and to perform *in vivo* tests. Currently, we are working on the co-crystallization of the *r*TBL-1 with ligands to obtain structural insights into its specific recognition by cancer-type glycans, and we are evaluating the effect of punctual mutations on its cytotoxic activity. The study of its interactions with membrane molecules with β 1-6 branched *N*-antennas (like EGFR) is also being addressed.

Author Contributions: Conceptualization, D.M.-A, T.G.-G., and A.B.-L.; protein production D.M.-A., P.P., and M.C.; characterization, D.M.-A.; crystallography, D.M.-A. and A.V.; formal analysis, D.M.-A.; writing—original draft preparation, D.M.-A.; writing—review and editing, all authors; supervision, T.G.-G., A.B.-L., A.V., E.F., and J.A.G.; project administration, T.G.-G. and A.B.-L.; funding acquisition, T.G.-G., A.B.-L., and A.V. All authors read and agreed to the published version of the manuscript.

Funding: This project received funding from the National Council for Science and Technology (CONACYT), the Program for Strengthening Educational Quality (PFCE), a CONACYT-Ciencia Básica grant (CB-2014-01-241181), and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant (765581).

Acknowledgments: The authors would like to thank Adrian P. Brown for his support with the mass spectrometry analysis; Ricardo Cervantes for providing the TBLF reference sample; and F. Josué López Martínez, Elizabeth Mendiola-Olaya, Valérie Chazalet, and Emilie Gillon for their technical assistance. They would also like to offer thanks for access to the beamlines BM30A-FIP at the European Synchrotron Facilities, Grenoble, France where the initial tests were performed and Proxima 1 at SOLEIL Synchrotron, Saint Aubin, France, (proposal number 20170827) where the final data were collected. Lastly, thanks also to local contacts Jean-Luc Ferrer, Serena Sirigu, and Pierre Legrand for their assistance and technical support.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ferriz-Martínez, R.A.; Torres-Arteaga, I.C.; Blanco-Labra, A.; García-Gasca, T. The role of plant lectins in cancer treatment. In *New Approaches in the Treatment of Cancer*; Mejía-Vázquez, M.C., Navarro, S., Eds.; Nova Science Publishers, Inc.: New York, NY, USA, 2010; pp. 71–89. ISBN 978-1-61728-304-8.
2. De Mejía, E.G.; Prisecaru, V.I. Lectins as bioactive plant proteins: A potential in cancer treatment. *Crit. Rev. Food Sci. Nutr.* **2005**, *45*, 425–445. [[CrossRef](#)]
3. Dabelsteen, E. Cell surface carbohydrates as prognostic markers in human carcinomas. *J. Pathol.* **1996**, *179*, 358–369. [[CrossRef](#)]
4. Pinho, S.S.; Reis, C.A. Glycosylation in cancer: Mechanisms and clinical implications. *Nat. Rev. Cancer* **2015**, *15*, 540–555. [[CrossRef](#)]
5. Varki, A.; Kannagi, R.; Toole, B.; Stanley, P. Glycosylation Changes in Cancer. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., et al., Eds.; Cold Spring Harbor Laboratory Press: New York, NY, USA, 2017.
6. Pearce, O.M.T.; Läubli, H. Sialic acids in cancer biology and immunity. *Glycobiology* **2015**, *26*, 111–128. [[CrossRef](#)]

7. Muinelo-romay, L.; Villar-portela, S.; Cuevas, E.; Gil-martín, E.; Fernández-briera, A. Identification of a (1, 6) fucosylated proteins differentially expressed in human colorectal cancer. *BMC Cancer* **2011**, *11*, 508. [[CrossRef](#)]
8. Moriwaki, K.; Miyoshi, E. Fucosylation and gastrointestinal cancer. *World J. Hepatol.* **2010**, *2*, 151–161. [[CrossRef](#)]
9. Chugh, S.; Gnanapragassam, V.S.; Jain, M.; Rachagani, S.; Ponnusamy, M.P.; Batra, S.K. Pathobiological implications of mucin glycans in cancer: Sweet poison and novel targets. *Biochim. Biophys. Acta Rev. Cancer* **2015**, *1856*, 211–225. [[CrossRef](#)]
10. Kannagi, R.; Sakuma, K.; Miyazaki, K.; Lim, K.T.; Yusa, A.; Yin, J.; Izawa, M. Altered expression of glycan genes in cancers induced by epigenetic silencing and tumor hypoxia: Clues in the ongoing search for new tumor markers. *Cancer Sci.* **2010**, *101*, 586–593. [[CrossRef](#)]
11. Taniguchi, N.; Kizuka, Y. Glycans and cancer: Role of N-Glycans in cancer biomarker, progression and metastasis, and therapeutics. In *Advances in Cancer Research*; Kenneth, D.T., Fisher, P.B., Eds.; Academic Press Inc.: California, CA, USA, 2015; Volume 126, pp. 11–51.
12. Li, J.H.; Wang, Y.C.; Qin, C.D.; Yao, R.R.; Zhang, R.; Wang, Y.; Xie, X.Y.; Zhang, L.; Wang, Y.H.; Ren, Z.G. Over expression of hyaluronan promotes progression of HCC via CD44-mediated pyruvate kinase M2 nuclear translocation. *Am. J. Cancer Res.* **2016**, *6*, 509–521.
13. Li, P.; Xiang, T.; Li, H.; Li, Q.; Yang, B.; Huang, J.; Zhang, X.; Shi, Y.; Tan, J.; Ren, G. Hyaluronan synthase 2 overexpression is correlated with the tumorigenesis and metastasis of human breast cancer. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 12101–12114.
14. Mereiter, S.; Balmaña, M.; Campos, D.; Gomes, J.; Reis, C.A. Glycosylation in the era of cancer-targeted therapy: Where are we heading? *Cancer Cell* **2019**, *36*, 6–16. [[CrossRef](#)]
15. Kaszuba, K.; Grzybek, M.; Orłowski, A.; Danne, R.; Róg, T.; Simons, K.; Coskun, Ü.; Vattulainen, I. N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4334–4339. [[CrossRef](#)]
16. Pucci, M.; Venturi, G.; Malagolini, N.; Chiricolo, M. Glycosylation as a main regulator of growth and death factor receptors signaling. *Int. J. Mol. Sci.* **2018**, *19*, 580.
17. García-Gasca, T.; Hernandez-rivera, E.; Lopez-Martínez, J.; Casta, A.L.; Yllescas-gasca, L.; Rodriguez, A.J.; Mendiola-olaya, E.; Castro-guillen, L.; Blanco-labra, A. Effects of tepary bean (*Phaseolus acutifolius*) protease inhibitor and semipure lectin fractions on cancer cells. *Nutr. Cancer* **2012**, *64*, 1269–1278. [[CrossRef](#)]
18. Moreno-Celis, U.; López-Martínez, F.J.; Cervantes-Jiménez, R.; Ferríz-Martínez, R.A.; Blanco-Labra, A.; García-Gasca, T. Tepary bean (*Phaseolus acutifolius*) lectins induce apoptosis and cell arrest in G0/G1 by p53(ser46) phosphorylation in colon cancer cells. *Molecules* **2020**, *25*, 1021. [[CrossRef](#)]
19. Ferríz-Martínez, R.; Garcia-Gasca, K.; Torres-Arteaga, I.; Rodríguez-Mendez, A.J.; Guerrero-Carrillo, M. de J.; Moreno-Celis, U.; Ángeles-Zaragoza, M.V.; Blanco-Labra, A.; Gallegos-Corona, M.A.; Robles-Álvarez, J.P.; et al. Tolerability assessment of a lectin fraction from Tepary bean seeds (*Phaseolus acutifolius*) orally administered to rats. *Toxicol. Rep.* **2015**, *2*, 63–69. [[CrossRef](#)]
20. Moreno-Celis, U.; López-Martínez, J.; Blanco-Labra, A.; Cervantes-Jiménez, R.; Estrada-Martínez, L.E.; García-Pascalín, A.E.; De Jesús Guerrero-Carrillo, M.; Rodríguez-Méndez, A.J.; Mejía, C.; Ferríz-Martínez, R.A.; et al. *Phaseolus acutifolius* lectin fractions exhibit apoptotic effects on colon cancer: Preclinical studies using dimethylhydrazine or azoxi-methane as cancer induction agents. *Molecules* **2017**, *22*, 1670. [[CrossRef](#)]
21. Alatorre-Cruz, J.M.; Pita-López, W.; López-Reyes, R.G.; Ferríz-Martínez, R.A.; Cervantes-Jiménez, R.; de Jesús Guerrero Carrillo, M.; Vargas, P.J.A.; López-Herrera, G.; Rodríguez-Méndez, A.J.; Zamora-Arroyo, A.; et al. Effects of intragastrically-administered Tepary bean lectins on digestive and immune organs: Preclinical evaluation. *Toxicol. Rep.* **2018**, *5*, 56–64. [[CrossRef](#)]
22. Martínez-Alarcón, D.; Mora-Avilés, A.; Espinoza-Núñez, A.; Serrano Jamaica, L.M.; Cruz-Hernández, A.; Rodríguez-Torres, A.; Castro-Guillen, J.L.; Blanco-Labra, A.; García-Gasca, T. Rhizosecretion of a cisgenic lectin by genetic manipulation of Tepary bean plants (*Phaseolus acutifolius*). *J. Biotechnol. X* **2019**, *3*, 100013. [[CrossRef](#)]
23. Demain, A.L.; Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* **2009**, *27*, 297–306. [[CrossRef](#)]

24. Martínez-Alarcón, D.; Blanco-Labra, A.; García-Gasca, T. Expression of lectins in heterologous systems. *Int. J. Mol. Sci.* **2018**, *19*, 616. [[CrossRef](#)]
25. Cereghino, J.L.; Cregg, J.M. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* **2000**, *24*, 45–66. [[CrossRef](#)]
26. Bretthauer, R.K.; Castellino, F.J. Glycosylation of *Pichia pastoris*-derived proteins. *Biotechnol. Appl. Biochem.* **1999**, *30*, 193–200.
27. Raemaekers, J.M.; De Muro, L.; Gatehouse, J.A.; Fordham-Skelton, A.P. Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris*. Correct N-terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide. *Eur. J. Biochem.* **1999**, *265*, 394–403. [[CrossRef](#)]
28. Laemmli, U.K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **1970**, *227*, 680–685. [[CrossRef](#)]
29. Fitches, E.C.; Pyati, P.; King, G.F.; Gatehouse, J.A. Fusion to snowdrop lectin magnifies the oral activity of insecticidal ω -hexatoxin-Hv1a peptide by enabling its delivery to the central nervous system. *PLoS ONE* **2012**, *7*, e39389. [[CrossRef](#)]
30. Yang, S.; Pyati, P.; Fitches, E.; Gatehouse, J.A. A recombinant fusion protein containing a spider toxin specific for the insect voltage-gated sodium ion channel shows oral toxicity towards insects of different orders. *Insect Biochem. Mol. Biol.* **2014**, *47*, 1–11. [[CrossRef](#)]
31. Kabsch, W. XDS. *Acta Crystallogr.* **2010**, *66*, 125–132. [[CrossRef](#)]
32. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.W.; McCoy, A.; et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D* **2011**, *67*, 235–242. [[CrossRef](#)]
33. McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674. [[CrossRef](#)]
34. Hamelryck, T.W.; Poortmans, F.; Chrispeels, M.J.; Wyns, L.; Loris, R. The Crystallographic structure of phytohemagglutinin-L. *J. Biol. Chem.* **1996**, *271*, 20479–20485. [[CrossRef](#)]
35. Murshudov, G.N.; Nicholls, R.A. REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 355–367. [[CrossRef](#)] [[PubMed](#)]
36. Emsley, P.; Lohkamp, B. Features and development of Coot. *Acta Crystallogr. Sect. D* **2010**, *66*, 486–501. [[CrossRef](#)]
37. Cummings, R.D.; Darvill, A.G.; Etzler, M.E.; Hahn, M.G. Glycan-recognizing probes as tools. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Har, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., Eds.; Laboratory Press: Cold Spring Harbor NY, USA, 2015; pp. 611–625.
38. Hammarstrom, S.; Hammarstrom, M.L.; Sundblad, G.; Arnarp, J.; Lönngrén, J. Mitogenic leucoagglutinin from *Phaseolus vulgaris* binds to a pentasaccharide unit in N-acetyllactosamine-type glycoprotein glycans. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 1611–1615. [[CrossRef](#)] [[PubMed](#)]
39. Loris, R.; Hamelryck, T.; Bouckaert, J.; Wyns, L. Legume lectin structure. *BBA* **1998**, *1383*, 9–36. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

SUMMARY

Scedosporium apiospermum is an emerging opportunistic fungal pathogen responsible for life-threatening infections in humans. Due to its high mortality rate and low susceptibility to antifungals, there is an urgent need to develop new therapeutic strategies against it. Since microbial lectins are often involved in host recognition and adhesion, they constitute valuable drug targets and have aimed at for the development of antiadhesive therapy against other deadly pathogens. However, in the case of *S. apiospermum*, the development of such therapies is hindered because its lectins are unknown. Therefore, the main aim of this work was the identification and characterization of *S. apiospermum* lectins. This has been addressed by two different approaches: the first strategy comprises the *in-silico* prediction of putative lectins codified on its genome by data mining and their recombinant production, while the second strategy lies in the identification of native lectins from fungal protein extracts. These approaches have led to the identification of several proteins with carbohydrate-binding activity from this human pathogen. The preliminary analysis of those proteins reveals that some of them are potential pharmacological targets and others could find application as biotechnological tools.

A second objective of this thesis was the development of a new tool that allows the identification and purification of unknown lectins directly from crude protein extracts. This device also enables the study of lectin carbohydrate interactions without the need for fluorescent labeling. Therefore, it could greatly contribute to the future study of lectinome from other organisms.

Overall, this work represents a broad strategy to approach the study of *S. apiospermum* lectinome and our findings contribute to the understanding of the glycosylated surfaces recognition by this pathogen. We hope that the information gathered here will encourage the study of the proteins discovered to gain fundamental knowledge and for potential biomedical or biotechnological applications.

RESUMÉ

Scedosporium apiospermum est un pathogène fongique opportuniste émergent responsable d'infections potentiellement mortelles chez l'homme. En raison de son taux de mortalité élevé et de sa faible sensibilité aux antifongiques, il est urgent de développer de nouvelles stratégies thérapeutiques pour le combattre. Puisque les lectines microbiennes sont souvent impliquées dans la reconnaissance et l'adhésion à l'hôte, elles constituent des cibles pharmacologiques intéressantes et ont été ciblées pour le développement d'une thérapie antiadhésive contre d'autres pathogènes mortels. Cependant, dans le cas de *S. apiospermum*, le développement d'une telle thérapie est entravé car ses lectines sont inconnues. Par conséquent, l'objectif principal de ce travail a été l'identification et la caractérisation des lectines de *S. apiospermum*. Cette question a été abordée selon deux approches différentes; la première stratégie comprend la prédiction *in-silico* de lectines putatives encodées dans son génome par data mining tandis que la seconde réside dans l'identification de lectines natives à partir d'extraits de protéines fongiques. Ces approches ont conduit à l'identification de plusieurs lectines et protéines chez ce pathogène humain. L'analyse préliminaire de ces protéines révèle que certaines d'entre elles pourraient être des cibles pharmacologiques intéressantes tandis que d'autres pourraient trouver une application en tant qu'outil biotechnologique.

Un deuxième objectif de cette thèse a été le développement d'un nouvel outil permettant l'identification et la purification de lectines inconnues directement à partir d'extraits bruts de protéines. Ce dispositif permet également l'étude des interactions lectines-glucides sans avoir recours à un marquage fluorescent. Par conséquent, il pourrait grandement contribuer à l'étude future des lectinomes pour d'autres organismes.

Dans l'ensemble, ce travail représente une stratégie pour aborder l'étude du lectinome de *S. apiospermum* et nos résultats contribuent à la compréhension de la reconnaissance des surfaces glycosylées par ce pathogène. Nous espérons que les informations recueillies ici encourageront l'étude des protéines découvertes pour acquérir des nouvelles connaissances fondamentales et en vue de potentielles applications biomédicales ou biotechnologiques.