



**HAL**  
open science

# Leishmanian Galactofuranosyltransferases as promising versatile tools for therapeutic and chemoenzymatic approaches

Jihen Ati

► **To cite this version:**

Jihen Ati. Leishmanian Galactofuranosyltransferases as promising versatile tools for therapeutic and chemoenzymatic approaches. Agricultural sciences. Université d'Orléans, 2018. English. NNT : 2018ORLE2043 . tel-04416208

**HAL Id: tel-04416208**

**<https://theses.hal.science/tel-04416208>**

Submitted on 25 Jan 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.

**ÉCOLE DOCTORALE SANTE, SCIENCES BIOLOGIQUES ET CHIMIE DU VIVANT**

**Enzymology and Glycobiology team at ICOA**

**THÈSE présentée par/ THESIS presented by: JIHEN ATI**

Soutenue le/ defended : **11 Décembre 2018**

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline/ Spécialité : **Biologie – Biochimie**

**Leishmanian Galactofuranosyltransferases as  
promising versatile tools for therapeutic and  
chemoenzymatic approaches.**

**THÈSE dirigée par :**

**Richard DANIELLOU  
Pierre LAFITE**

Professeur- Université d'Orléans - Directeur de Thèse  
Maitre des conférences- Université d'Orléans –  
Co-encadrant de Thèse

**RAPPORTEURS :**

**Rob FIELD  
Christophe BIOT**

Professeur- John Innes Center  
Professeur- Université de Lille-1

**JURY :**

**Rob FIELD  
Christophe BIOT  
Florence ROBERT-GANGNEUX  
Richard DANIELLOU  
Pierre LAFITE**

Professeur- John Innes Centre  
Professeur- Université de Lille-1  
Professeur- Université de Rennes-1  
Professeur- Université d'Orléans  
Maitre des conférences- Université d'Orléans



# Acknowledgments

This doctoral work was carried out at the Institute of Organic and Analytical Chemistry (ICOA - UMR 7311) at the University of Orléans under the supervision of Professor Richard Daniellou and Dr. Pierre Lafite.

First, I would like to thank Professor Olivier Martin and Professor Pascal Bonnet for welcoming me into their institute.

Then, I would like to express my gratitude to Professor Rob Field from John Innes Center (UK) and Professor Cristophe Biot from the University of Lille-1 (Fr) for accepting to be reviewers of my doctoral studies. I would like also to acknowledge Professor Florence Robert-Grangneux from the University of Rennes-1 (Fr) for being a member of the Examination Committee of this thesis.

I joined the Glycobiology and Enzymology team in January 2015 for my 2<sup>nd</sup> year master degree internship and continued working on the same subject during my doctoral studies. These four years have been an intensive and rich scientific experience through which I discovered the world of enzymes and sugars!! But they were also an unforgettable human adventure full of tears of joy, deception, doubt, hope and “blackcat” moment. For that I would like to thank Richard, my director of thesis and Pierre, my co-supervisor for making this Glycobio team feels like a small family. Richard thank you for being supportive, making yourself available even when you not supposed to be. I am so grateful for showing me confidence in my work, for your patience, help, and for all the “VnB-moments”. Pierre, I would like to express my immense gratitude for being here when I needed help, tips, explanation, or just talk. I will never thank you enough for believing in me and the most important thing for making me believe in myself, and of course all the “Mcdomoments”. Richard and Pierre, thanks to you I gained self-confidence and acquired lot of knowledge that allowed me to conduct this project.

I would like to thank the other members of the Glycobio family; Philippe for all your help with *Bacteria* and baby *Leishmania* cultures, and thanks to who I discovered new instrumental music. Cedric Peyrot (CP), Laure Guillotin (LG) aka Jedi, the adopted member Matthieu Place (MP) for sharing with me a lot of funny moments, trips, WE and for your precious friendship. Mateja aka Padawan, thank you for being always enthusiastic and for cheering me up. To the students, Ophelie, Alicia, Pauline 1, Julie, Pauline 2 and special thanks to Laura.

To my colleague who became my friends, thank you; Yasmina for your joyful spirit, Melanie aka “July” for (my blabla N1) for your help and your presence, Helene and Eugénie (my blabla N2) for your support and Abdou for all the coffee-breaks.

To Emmanuelle, Ludovic, Blanka and Annabelle, thank you for your advices.

To Carine thank you for your help with my baby *Leishmania*.

I would like also to thank my friends who were here when I needed the most and made me feel better during the hard moments, and also for sharing with me all the good memories:

Here they are:

- To the squad: Diane, Mariska, Bouchra, Kissi, and special thanks to Manoun, thank you for your encouragement, our special coffee-time moments, music concerts rituals, and all your support.
- To Marraw and Bbti nessrine my two old friends-sisters thank you for your precious time that you spent on hearing me moan.
- To my volleyball-friends; FX, Marine, Adeline, Aurélie and Mathieu, thank you for all the funny adventures, for making me lough and also for the nicknames “Rebelle and Blackcat”.

I take this precious opportunity to express my gratitude to my family, Walidou, Gaelle, “les gnomes”, “Frère” and Romain, thank you for all what you have done to make this adventure possible.

And my last words go to my parents and specially to my Mom to who I say: Greatness in your Memory!



# Abbreviations

<b><math>\alpha</math>-Glc-1-P:</b> $\alpha$ -Glucose-1-Phosphate	<b>LPGs:</b> Lipophosphoglycans
<b>Cam :</b> Chloramphenicol	<b>mAGP:</b> Mycolyl-arabinogalactan-peptidoglycan
<b>CAZy :</b> Carbohydrate Active enZymes Database	<b>MCL:</b> Mucocutaneous leishmaniasis
<b>DCL:</b> Diffuse cutaneous leishmaniasis	<b>Mtb:</b> <i>Mycobacterium tuberculosis</i>
<b>ER:</b> Endoplasmic Reticulum	<b>NO:</b> Nitrogen oxide
<b>Galfase :</b> Galactofuranosidase	<b>NTD:</b> Neglected Tropical Diseases
<b>Gal/Ts:</b> Galactofuranosyltransferase	<b>PCR:</b> Polymerase Chain Reaction
<b>GalPUT:</b> Galactose-1-phosphate uridyl transferase	<b>pb:</b> Bases Pairs
<b>GHs:</b> Glycoside Hydrolases	<b>PDB:</b> Protein Data Bank
<b>GIPLs:</b> Glycoinositol Phospholipids	<b>Tris:</b> Tris(hydroxymethyl)aminomethane
<b>GK:</b> Galactose kinase	<b>UDP-Galf:</b> Uridine Diphosphate Galactofuranose
<b>GMR:</b> Galactose Mutarotase	<b>UDP-<math>\alpha</math>-Galp:</b> Uridine diphosphate- $\alpha$ -Galactopyranose
<b>GTs:</b> Glycosyltransferases	<b>UDP-D-Glcp:</b> UDP-D-Glucopyranose
<b>His-Tag:</b> <i>N</i> -terminal and/or <i>T</i> erminal of 6 Histidine sequence	<b>UGE:</b> UDP-Glc-4-epimerase
<b>HRMS:</b> High Resolution Mass Spectrometry	<b>UGM:</b> UDP-Galactopyranose Mutase
<b>HMBC:</b> Heteronuclear Multiple Bond Coherence	<b>UTP:</b> Uridine Triphosphate
<b>HIV:</b> Human Immunodeficiency Virus	<b>VL:</b> Visceral Leishmaniasis
<b>IMAC:</b> Immobilized Metal Affinity Chromatography	<b>WHO:</b> World Health Organization
<b>IPTG:</b> Isopropyl $\alpha$ -D-1-thiogalactopyranose	
<b>Kan:</b> Kanamycin	
<b><math>k_{cat}</math>:</b> Turn-over number	
<b><math>K_M</math>:</b> Michaëlis-Menten constant	
<b>LB:</b> Lysogeny Broth	
<b>LC-MS:</b> Liquid Chromatography-Mass Spectrometry	
<b>LPS:</b> Lipopolysaccharide	

# Table of contents

<b>Chapter I: General introduction</b> .....	19
<b>Chapter II: Bibliogrphay</b> .....	34
I. Glycoconjugates.....	36
A. Biological role of glycoconjugates.....	36
B. Structures.....	37
II. D-Galactofuranose.....	38
A. Therapeutic relevance of galactofuranose.....	38
1. In bacteria.....	38
2. In fungi.....	40
3. In protozoa.....	40
B. Biosynthesis of glycoconjugates containing galactofuranose.....	42
1. UDP-Galactofuranose transporter.....	43
2. UDP- Galactopyranose mutase.....	44
3. Galactofuranosidases .....	45
4. Galactofuranosyltransferases.....	45
III. Characterized galactofuranosyltransferase.....	48
A. GlfT1 and GlfT2 of <i>Mycobacterium tuberculosis</i> .....	48
B. GfsA of <i>Aspergillus</i> .....	51
C. WbbI, galactofuranosyltransferases of <i>E. coli</i> K-12.....	52
D. Summary.....	52
IV. Galactofuranose-containing glycoconjugates of Leishmania: therapeutic target for.....	
Leishmaniasis.....	53
A. Leishmaniasis as a neglected tropical disease.....	53
1. Causing agent, Leishmania.....	53
2. Life cycle.....	54
3. Clinical forms of leishmaniasis.....	55
4. Epidemiology.....	56
5. Diagnosis and treatment.....	58
B. Cell wall glycoconjugates as therapeutic targets.....	61
1. Lipophosphoglycans.....	63
2. Glycosylinositol phospholipids.....	65
3. Biosynthesis of glycoconjugates.....	66
V. Objective of thesis.....	69
References.....	74
<b>Chapter III: LPG1x production in prokaryotic host</b> .....	83
I. Recombinant protein production.....	87
A. <i>E. coli</i> as a bacterial expression system: advances and challenges.....	87
1. The well-suitable expression vectors.....	88
2. How to “extract” proteins from <i>E. coli</i> ?.....	90
3. Obtaining pure proteins.....	91
4. Inclusion bodies and solubility of proteins: Common problems related to recombinant protein production in <i>E. coli</i> .....	91
II. LPG1x production: A challenging goal.....	91
A. Genomic and structural study of four putative GalTs.....	91
B. LPG1 x cloning, overexpression and purification.....	93

1. pET vectors as heterologous expression system.....	93
2. Cloning .....	95
3. Overexpression of LPG1x .....	96
4. ATPase activity.....	104
5. Removal of Chaperonin GroEL.....	106
6. pMal vectors as heterologous expression system.....	110
III. Conclusion.....	116
References .....	118
<b>Chapter IV- Characterization of prokaryotic LPG1x .....</b>	<b>122</b>
I- The characterization of galactofuranosyltransferases is an ongoing challenge.....	125
A. UDP-Galactofuranose synthesis.....	125
1. Chemical approaches .....	125
2. Enzymatic approaches.....	127
B. Enzymatic assays of GT .....	128
1. Spectrophotometric assays .....	129
2. High performance liquid chromatography .....	130
II. Enzymatic characterization of LPG1x proteins .....	131
A. Coupled Spectrophotometric assay .....	131
1. Galactofuranosyltransferases activity of LPG1x proteins.....	131
2. Product characterization.....	135
B. Substrate specificity.....	137
1. Donor specificity .....	137
2. Acceptors specificity.....	142
3. Temperature and pH effect.....	144
III. Crystallization of LPG1x.....	145
IV. Conclusion .....	149
References.....	151
<b>Chapter V- Expression of eukaryotic LPG1x .....</b>	<b>154</b>
I. Recombinant proteins produced in eukaryotic cells.....	157
A. Glycosylation of proteins.....	157
B. <i>Leishmania tarentolae</i> as an expression host system.....	160
1. <i>Leishmania tarentolae</i> .....	160
2. Lexsy vector .....	161
II. Glycosylated recombinant GalTs of <i>Leishmania</i> .....	163
A. Cloning and transfection.....	163
1. Cloning.....	163
2. Transfection and selection.....	166
B. Protein expression and characterization.....	169
1. Protein Purification.....	169
2. Immunodetection.....	171
3. Glycosylation.....	173
4. Activity characterization.....	175
III. Conclusion.....	178
References.....	179

<b>Chapter VI- Conclusion and perspectives</b> .....	182
<b>Experimental part</b> .....	206
I. DNA Cloning.....	207
A. PCR and cloning strategies.....	207
B. Vectors digestion and transformation in DH5 $\alpha$ cells.....	211
C. Plasmid DNA purification, analysis and sequencing.....	212
II. Expression and purification of proteins.....	212
A. Prokaryotic proteins.....	212
1. Transformation in and overexpression of proteins.....	212
2. Lysis of cells.....	213
3. Protein purification.....	213
B. Eukaryotic proteins.....	214
1. Culture of <i>Leishmania</i> .....	214
2. <i>Leishmania</i> transfection.....	215
III. Analysis of recombinant proteins.....	215
A. Bradford protein assays.....	216
B. Tryptic digestion of proteins.....	216
C. Western Blot.....	216
1. Solutions.....	216
2. Transfer.....	217
3. Immunodetection.....	217
4. Detection.....	217
IV. Enzyme activity assays and kinetics parameters determination.....	217
A. Malachite green assay.....	217
B. HPLC detection.....	218
C. Spectrophotometric assays.....	218
D. Galactofuranosyltransferase reactions (scale-up) .....	218
E. Separation step and TLC analysis.....	218
F. Acetylation of sugars and analysis.....	219
V. Cristallogenesis of proteins.....	219
<b>Appendix</b> .....	221



# Table of Figures

## Chapter I: General introduction

<b>Figure I- 1:</b> A summary of carbohydrates and glycosidic bonds associated to peptides found in nature. ....	20
<b>Figure I- 2:</b> Identification and characterization of GHs and generation of thioligases through side directed mutagenesis strategy. ....	22
<b>Figure I- 3:</b> Galactofuranose-containing glycoconjugates biosynthesis pathway with the key enzymes.....	23

## Chapter II: Bibliography

<b>Figure II- 1:</b> Glycoproteins and glycolipids in cell membrane. ....	35
<b>Figure II-2:</b> Role of glycans in recognition and communication with other cells and microorganisms. ....	36
<b>Figure II- 3:</b> Furanose and pyranose conformations of D-galactose.....	37
<b>Figure II- 4:</b> Galactan moiety of the O-antigen of <i>Klebsiella pneumonia</i> .....	38
<b>Figure II- 5:</b> Structure of mAGP complex in Mtb.....	39
<b>Figure II- 6:</b> Galactomannan pattern as found in <i>Aspergillus fumigatus</i> .....	40
<b>Figure II- 7:</b> <i>Trypanosoma cruzi</i> mucin.....	40
<b>Figure II- 8:</b> Lipophosphoglycan core from <i>Leishmania</i> .....	41
<b>Figure II- 9:</b> The two known pathways that generate UDP-Galactofuranose.....	42
<b>Figure II- 10:</b> Transport of UDP-Galactofuranose in eukaryotes.....	43
<b>Figure II- 11:</b> Conversion of UDP- $\alpha$ -Galactopyranose to UDP- $\alpha$ -Galactofuranose by UGM.....	44
<b>Figure II- 12:</b> Schematic diagram of six stranded Rossmann fold.....	45
<b>Figure II- 13:</b> Structures of (C) GT-A-fold of diphosphate sugar transferase SpsA from <i>Bacillus subtilis</i> (PDB ID 1QGQ) and (D) GT-B-fold bacteriophage T4 $\beta$ -glucosyltransferase.....	46
<b>Figure II- 14:</b> Schematic representation of inverting catalytic mechanism that utilizes a direct displacement SN2-like reaction that results in an inverted anomeric configuration via a single oxocarbenium ion-like transition state.....	46
<b>Figure II- 15:</b> Pathway for the biosynthesis of mycobacterial arabinogalactan.....	48
<b>Figure II- 16:</b> Crystal structure of GlfT2 complexed with UDP.....	49
<b>Figure II- 17:</b> The structure of the catalytic site binding the UDP and Mn <sup>2+</sup> .....	50
<b>Figure II- 18:</b> Schematic in-vitro assay reaction of AFGfsA.....	51
<b>Figure II- 19:</b> Promastigote and amastigote form of <i>Leishmania</i> , the causative agent of leishmaniasis.....	53
<b>Figure II- 20:</b> The life cycle of <i>Leishmania</i> parasites, whether in the sand-fly vector or the human host.....	54
<b>Figure II- 21:</b> Statute of endemic cutaneous leishmaniasis worldwide 2015.....	56
<b>Figure II- 22:</b> Statute of endemic visceral leishmaniasis worldwide 2015.....	57
<b>Figure II- 23:</b> Chemical formula of two pentavalent antimonial compounds.....	58
<b>Figure II- 24:</b> Chemical formula of pentamidine.....	58
<b>Figure II- 25:</b> Chemical formula of amphotericin B.....	59
<b>Figure II- 26:</b> Chemical formula of miltefosine.....	60
<b>Figure II- 27:</b> Different glycoconjugates structures of <i>Leishmania</i> .....	61
<b>Figure II- 28:</b> Evolution of glycoconjugates during different life cycle stage of <i>L. donovani</i> parasites.....	62
<b>Figure II- 29:</b> Structure of the four domains of Lipophosphoglycan.....	63
<b>Figure II- 30:</b> Schematic representation of the three types of GIPLs identified in <i>Leishmania</i> ....	65

<b>Figure II- 31:</b> Biosynthesis of Leishmania phosphoglycans.....	67
<b>Figure II- 32:</b> Representation of LPG1, LPG1R, LPG1L and LPG1G.....	69
<b>Figure II- 33:</b> Evaluation of the expression of the four genes; <i>lpg1</i> , <i>lpg1R</i> , <i>lpg1L</i> and <i>lpg1G</i> over the time by <i>L. major</i> , <i>L. donovani</i> and <i>L. infantum</i> .....	70
<b>Figure II- 34:</b> Objectives of thesis.....	72

### Chapter III: LPG1x production in prokaryotic host

<b>Figure III- 1:</b> General steps required to produce recombinant proteins.....	88
<b>Figure III- 2:</b> Scheme representing the essential element present into expression vector.....	88
<b>Figure III- 3:</b> Induction mechanism without and with IPTG.....	90
<b>Figure III- 4:</b> IPTG structure.....	90
<b>Figure III- 5:</b> Partial amino acid sequence alignment of LPG1, LPG1L, LPG1R and LPG1G realized with CLC sequence viewer software.....	92
<b>Figure III-6:</b> Transmembrane domain prediction of LPG1, LPG1L, LPG1R and LPG1G. TM: Transmembrane region, predicted from THMM software, are localized in the N-terminus of proteins.....	93
<b>Figure III-7:</b> pET Scheme representing cloning product obtained with pET-28a(+) and pET-32a(+) and their respective recombinant proteins fused with histidine tag.....	94
<b>Figure III- 8:</b> Scheme representing the strategy used to remove the transmembrane domain from the recombinant proteins.....	95
<b>Figure III- 9:</b> Classical cloning principle.....	95
<b>Figure III- 10:</b> SDS-PAGE (12%) analysis of the overexpression of LPG1 (pET-28a(+)) under different conditions of the induction; 2,4 and 20 hours of induction after the addition of IPTG in the medium at 25, 30 and 37°C.....	97
<b>Figure III- 11:</b> SDS-PAGE (8%) analysis of the overexpression of LPG1 (pET-32a(+)) under different conditions of the induction; 2,4 and 20 hours of induction after the addition of IPTG in the medium at 25, 30 and 37°C.....	98
<b>Figure III- 12:</b> Interaction between the poly-histidine tag fused to the recombinant protein and two different immobilized metal affinity chromatography (IMAC) Ni-NTA and Co <sup>2+</sup> .....	100
<b>Figure III- 13:</b> Analysis of the overexpression of of LPG1(pET- pET28(a+)) and its purification of LPG1(pET- pET28a(+)) on Ni-NTA and Cobalt column on SDS-PAGE 12%.....	100
<b>Figure III- 14:</b> SDS-PAGE (12%) analysis of overexpression of LPG1G (pET-28a(+)) proteins after the induction at 25°C for 2 hours, and its purification on Ni-NTA column. Three different detergent were tested solubilize the recombinant proteins.....	102
<b>Figure III- 15:</b> SDS-PAGE (12%) analysis of overexpression of LPG1G (pET-28a(+)) proteins after the induction at 25°C for 2 hours, and its purification on Ni-NTA column followed by size exclusion chromatography.....	103
<b>Figure III- 16:</b> Electrophoresis of contaminant on SDS-PAGE (12%).....	104
<b>Figure III- 17:</b> GroEl-GroEs complex.....	105
<b>Figure III-18:</b> Protein folding cycle of <i>Escherichia coli</i> complex GroEL-GroES.....	106
<b>Figure III-19:</b> SDS-PAGE (12%) analysis of overexpression of LPG1G (pET-28a(+)) proteins after the induction at 25°C for 2 hours, and its purification after lysis strategy 2 on Ni-NTA column with ATP.....	107
<b>Figure III- 20:</b> SDS-PAGE (12%) analysis of overexpression of LPG1G (pET-28a(+)) proteins after the induction at 25°C for 2 hours using C43 (DE3) and BL21 (DE3) <i>E. coli</i> strains, and its purification after lysis strategy 2 on Ni-NTA column.....	108
<b>Figure III-21:</b> Representation of the strategy of cloning of <i>lpg1x</i> genes into pET-28a(+) expression vector.....	109
<b>Figure III- 22:</b> SDS-PAGE (12%) analysis of overexpression of LPG1G-2H (pET-28a) proteins after the induction at 25°C for 2 hours, and its purification after lysis alternative strategy on Ni-NTA column.....	110

<b>Figure III- 23:</b> pMAL Scheme representing Strategy of cloning and the expected recombinant proteins fused with MBP-Tag in C-terminus.....	111
<b>Figure III- 24:</b> SDS-PAGE (8%) analysis of the overexpression of LPG1-MBP under different conditions of the induction; 2,4 and 20 hours of induction after the addition of IPTG in the medium at 25 and 30°C.....	112
<b>Figure III- 25:</b> SDS-PAGE (8%) analysis of the purification of LPG1x-MBP by affinity column followed by SEC.....	113
<b>Figure III- 26:</b> Mass spectrometry analysis of LPG1G-MBP .....	114
<b>Figure III- 27:</b> Analysis of the removal of MBP-Tag with factor X from LPG1G-MBP proteins...115	
<b>Figure III- 28:</b> Summary of the strategies used to produce LPG1x proteins.....	117

## Chapter IV- Characterization of prokaryotic LPG1x

<b>Figure IV- 1:</b> Synthesis of UDP-Galf described by Tsvetkov and Nikolaev.....	126
<b>Figure IV- 2 :</b> Synthesis of UDP-Galf described Kiessling and co-workers starting from an activated 5'-N-methyl phosphorylimidazole .....	126
<b>Figure IV- 3:</b> Synthesis of UDP-Galf described Ferrières and co-workers starting from thiomidates.....	127
<b>Figure IV- 4:</b> UDP-Galp Mutase catalysed the conversion of UDP-Galp into UDP-Galf.....	127
<b>Figure IV-5:</b> Chemoenzymatic strategy to synthesize UDP- $\alpha$ -D-Galf described by Field and coworkers.....	128
<b>Figure IV- 6:</b> Coupled spectrophotometric assay to measure the activity of GalT by following the consumption of NADH at 340nm.....	129
<b>Figure IV- 7:</b> Malachite green assay principle.....	130
<b>Figure IV- 8:</b> LPGs and GIPLs structures in <i>Leishmania major</i> containing galactofuranose residue linked to mannose residue.....	131
<b>Figure IV-9:</b> Oxidation of NADH over time detected at 340nm in 200 $\mu$ l reaction mixture containing 1Mm UDP-Galf, 1Mm Me-Manp and 0.2 $\mu$ g of LPG1.....	132
<b>Figure IV- 10:</b> Michaelis-Menten plot that corresponds to UDP-Galf kinetics as donor substrate of LPG1.....	133
<b>Figure IV- 11:</b> Typical TLC obtained with the GalT reaction and UDP- $\alpha$ -D-Galf as donor.....	135
<b>Figure IV-12:</b> HRMS analysis of peracetylated disaccharide product after LPG1-catalyzed reaction using $\alpha$ -D-Methyl-mannoside as acceptor and UDP-Galf as sugar donor.....	136
<b>Figure IV- 13:</b> List of NDP-pyranoses tested with LPG1x GalTs.....	137
<b>Figure IV- 14:</b> Michaelis-Menten plot that corresponds to UDP-Galp kinetics as donor substrate of LPG1.....	138
<b>Figure IV-15:</b> Typical TLC obtained with the GalT reaction and NDP-Pyranose as donor.....	140
<b>Figure IV-16:</b> HRMS analysis of peracetylated disaccharide product after LPG1-catalyzed reaction using $\alpha$ -D-Methylmannoside as acceptor and UDP-Galp as sugar donor.....	141
<b>Figure IV- 17:</b> The hypothetic reaction of leishmanial GalTs with pNP-glycosides as acceptors, resulting the formation of pNP-disaccharide.....	143
<b>Figure IV- 18:</b> HPLC analysis of reaction mixture containing 0.1 $\mu$ M of LPG1, 1Mm of UDP-pyranoside and 1mM of pNP-Mannose.....	144
<b>Figure IV- 19:</b> Different step of protein crystallogenesis protocol.....	146
<b>Figure IV- 20:</b> Different state that protein could reach during crystallization process resulted by vapor diffusion in gas state.....	147
<b>Figure IV-21:</b> Two forms that proteins could adopt after reaching supersaturated state.....	147
<b>Figure IV-22:</b> Optimization strategy of crystallization condition using 20-42% of PEG and 160 – 200Mm of ammonium sulfate.....	148
<b>Figure IV- 23:</b> LPG1 crystals stained with methylene blue.....	148

## Chapter V- Expression of eukaryotic LPG1x

<b>Figure V- 1:</b> Different type of glycosylation that occur in eukaryotic species.....	158
<b>Figure V- 2:</b> N-glycosylation examples produced in E.R and Golgi apparatus of mammalian, insect and plant cells. ....	159
<b>Figure V- 3:</b> Glycosylation profiles of recombinant glycoproteins produced by yeast, mammalian, plant, insect and leishmanial cells. ....	160
<b>Figure V- 4:</b> pLEXSY-2 plasmid cartography. ....	161
<b>Figure V- 5:</b> Transfection and integration of expression cassette from LEXSY plasmid into the genome of <i>Leishmania tarentolae</i> by homologous recombination thanks to SSU regions.....	162
<b>Figure V- 6:</b> The two strategies of cloning.. ....	164
<b>Figure V- 7:</b> Mechanism of In-Fusion cloning reaction used to insert genes of interest into LEXSY plasmid. ....	165
<b>Figure V- 8:</b> Preparation of linearized expression cassette by enzymatic digestion of cloning constructions by Swal. ....	166
<b>Figure V- 9:</b> Culture state after selection with hygromycin antibiotic. ....	167
<b>Figure V- 10:</b> Microscopic image (X600) of lpg1x-transfected cells with lpg1G during the selection process with hygromycin. ....	167
<b>Figure V- 11:</b> Microscopic image (X 1000) of transfected <i>L. tarentoale</i> cells with lpg1G after several passage showing healthy promastigotes cells with lance-like shape. ....	168
<b>Figure V- 12:</b> Cryo -Electron microscopy picture showing the presence of transfected promastigotes with drop-like shape and flagellum.....	169
<b>Figure V- 13:</b> Evaluation of the expression and the purity of <i>L. major</i> GalT's after Ni-NTA elution step in 1-D 8% SDS-PAGE stained with Coomassie.....	170
<b>Figure V- 14:</b> Colorimetric detection.....	171
<b>Figure V- 15:</b> Reaction of phosphatase alkaline in the presence of substrate BCIP/NBT.....	171
<b>Figure V- 16:</b> Identification of LPG1G (67 kDa), LPG1L (62 kDa) and LPG1R (52 kDa) by western blot technique, using antibodies against to C-terminal poly-histidine tags fused to the proteins of interest.....	172
<b>Figure V- 17:</b> Predictions for <i>N</i> -Glycosylation sites in specific sequence in secreted and recombinant LPG1G amino acid sequence. ....	172
<b>Figure V- 18:</b> Predictions for <i>O</i> -Glycosylation sites in specific sequence in secreted and recombinant LPG1G amino acid sequence. ....	174
<b>Figure V- 19:</b> Removal of N-linked oligosaccharides by PNGase. ....	174
<b>Figure V- 20:</b> Evaluation of deglycosylation of LPG1G due to PNGase activity on 1-D 8% SDS-PAGE stained with Coomassie with standard mixture marker proteins. ....	175
<b>Figure V- 21:</b> Expected enzymatic reaction catalyzed by glycosylated LPG1G. ....	176
<b>Figure V- 22:</b> Michaelis-Menten plot that corresponds to UDP-Galp kinetics as donor substrate of glycosylated LPG1G.....	176

## Chapter VI- Conclusion and perspectives

<b>Figure VI- 1:</b> Summary of LPG1x production using pET and pMAL vectors. ....	184
<b>Figure VI- 2:</b> Summary of the substrate recognized by the four LPG1x produced by <i>E. coli</i> .....	186
<b>Figure VI- 3:</b> Summary of the glycosylated and non-glycosylated LPG1x produced by <i>Leishmania tarentolae</i> and their activity. ....	187
<b>Figure VI- 4:</b> Reported trisaccharide substrates for GlfT2. ....	188
<b>Figure VI- 5:</b> Structures of the synthetic acceptors tested for GlfT1.....	188
<b>Figure VI- 6:</b> Imminosugars tested as substrate with GlfT210.....	190

**Figure VI- 7:** Detection strategies of galactofuranose-containing glycoconjugates synthesized by Gal/Ts. ....190



# Table of Tables

## Chapter II: Bibliography

**Table II- 1:** Summary of available data about the acceptors, kinetics parameters, and crystal structures of the four GalT's; GlfT2, GlfT1, GfsA, and WbbI.....51

## Chapter III: LPG1x production in prokaryotic host

**Table III- 1:** Summary table of the homology of LPG1x proteins, GlfT, GlfT2, GfsA and WbbI.....92

**Table III- 2:** Size of native genes, the removed transmembrane regions and the molecular weight of native and recombinant proteins without TM .....96

**Table III- 3:** Chemical structures of different used detergents.....101

**Table III- 4:** The expected molecular weight of the recombinant proteins (RP) LPG1x fused with His-Tag in both *N*- and *C*- termini.....109

**Table III- 5:** Expected molecular weight of the recombinant proteins (RP) LPG1x fused with MBP-Tag *N*-termini.....111

**Table III- 6:** Coverage percent of digested peptide with LPG1G-MBP and LPG1R-MBP amino acid sequence.....114

## Chapter IV- Characterization of prokaryotic LPG1x

**Table IV- 1:** Kinetic parameters of leishmanial GalT's LPG1, LPG1G, LPG1L, LPG1R compared with mycobacterial GlfT2 for UDP- $\alpha$ -D-Galp. <sup>a</sup>: Me-Manp was used as the acceptor.....134

**Table IV- 2:** Kinetic parameters of leishmanial GalT's LPG1, LPG1R, LPG1L and LPG1G compared for different NDP-pyranosides donors. <sup>a</sup>: Me-Manp was used as the acceptor.....139

**Table IV- 3:** Concentration of LPG1x used for crystallogenes trials.....146

## Chapter V- Expression of eukaryotic LPG1x

**Table V-1:** Comparison between *Leishmania tarentolae* and *Leishmania major* sequenced genome.....161

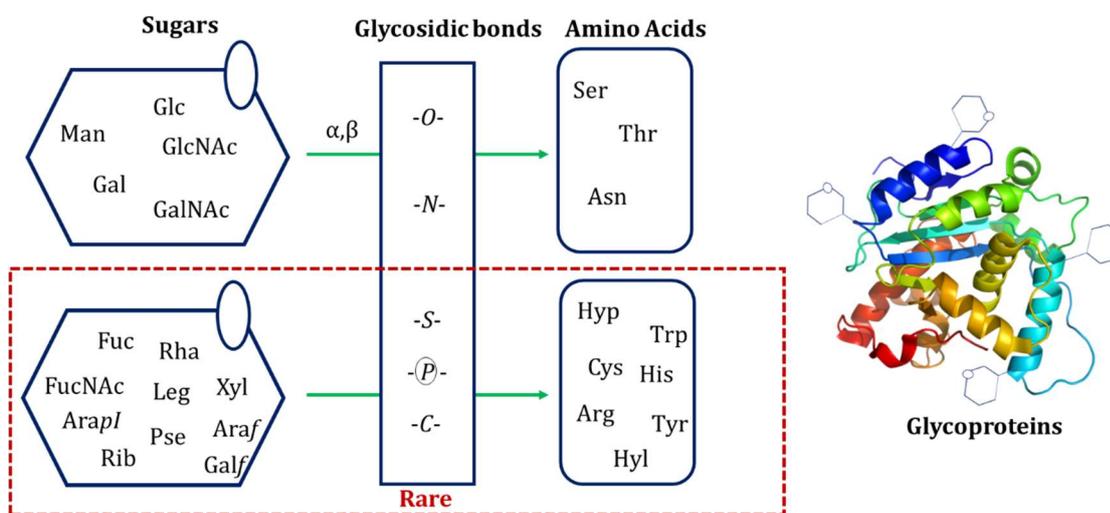
**Table V-2:** Type of cloning used for each gene and the expected PCR product length.....165

**Table V-3:** Kinetic parameters of eukaryotic GalT LPG1G compared with prokaryotic GlfT2 for UDP- $\alpha$ -D-Galp and Me-Manp was used as the acceptor.....177



# ***Chapter I: General introduction***

Carbohydrates play key roles in numerous biological process such as growth, development, function, structure, communication, and survival of cells from all kind of living species.<sup>1,2</sup> Glycans can also be attached to proteins. This mechanism is regulated during the post translational modification pathways that occurs in eukaryotic species. The glycosylation of proteins is highly important for their structure and activity. Indeed, an error in protein glycosylation can lead to the various diseases, such as autoimmune diseases and cancer.<sup>3,4</sup> Unlike proteins synthesis glycosylation is not template-driven and it cannot be easily predicted.<sup>5</sup> Two class of enzymes, glycosyltransferases (GTs) and glycoside hydrolases (GHs) are responsible for the biosynthesis of glycoconjugates. GTs transfer monosaccharides onto suitable acceptor, while GHs catalyse the hydrolysis of glycosidic linkages resulting in the release of carbohydrates.<sup>6</sup> Nowadays, available data about the different glycosides and their linkages represent only “the tip of iceberg”(Figure I-1). Indeed, there are rare sugars and glycosidic bonds that need to be studied. So, we need to expand our knowledge in order to be able to synthesize natural glycoconjugate-like compounds that can be used in numerous fields such as pharmaceuticals, cosmetics, food, materials and renewable resources.<sup>7,8</sup> Up to date, glycochemists have developed efficient tools to successfully synthesize carbohydrates derivatives. However, these means require long procedures including protection and deprotection steps.<sup>9</sup>



**Figure I- 1:** A summary of carbohydrates and glycosidic bonds associated to peptides found in nature. (Adapted from Pierre Lafite and Richard Daniellou, 2012)<sup>2</sup>

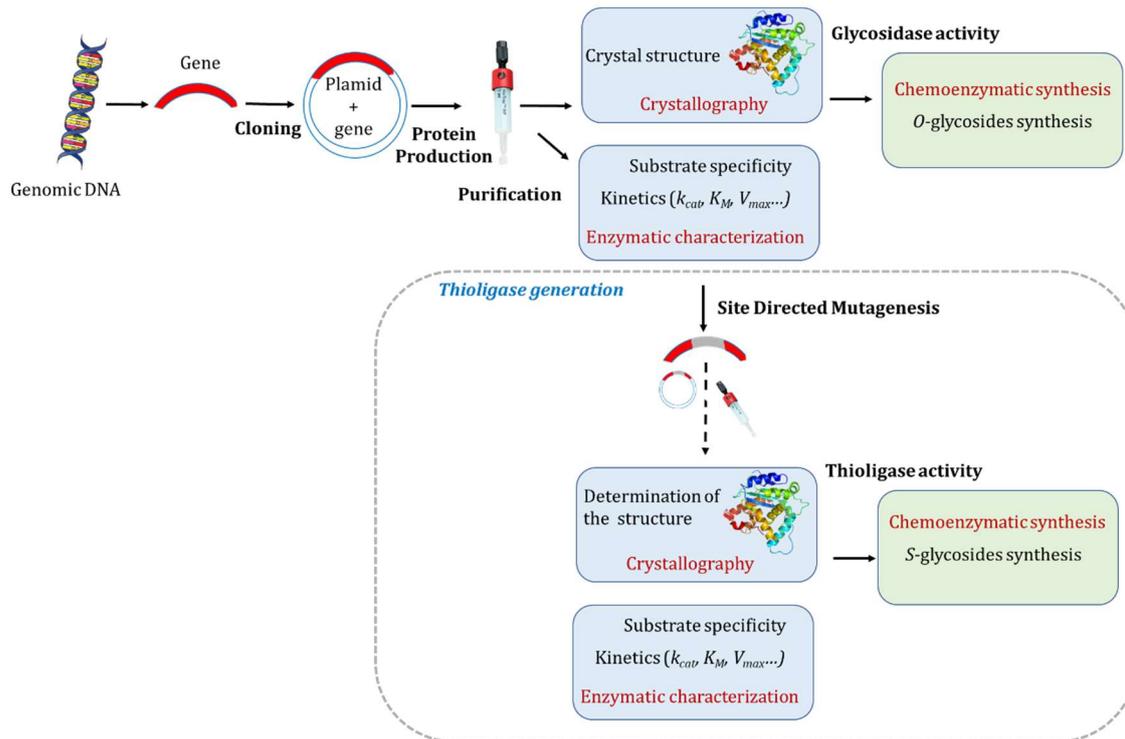
The glycobiochemistry and enzymology team at the Institute of Organic and Analytical Chemistry (ICOA UMR CNRS 7311), which is led by Professor Richard DANIELLOU, is interested in the study of rare glycosidic linkages and carbohydrates by discovering and characterizing new carbohydrates-active enzymes. These enzymes are used as biocatalysts that represent a powerful and eco-compatible tool for the synthesis of natural carbohydrates and their analogues. Indeed, the enzymes can display high specificity and regioselectivity toward substrates, in addition, they

work in aqueous environment.<sup>10</sup> These characteristics are highly required today in chemical industry, because they offer new solutions for chemical reactions to which there is no suitable or available chemical catalysts. Moreover, enzymatic reactions replace chemical processes that necessitate a long time and produce large amount of toxic wastes. By this, they follow the rules of the “green chemistry”, which include environmental-friendly bioprocesses.<sup>11</sup> In parallel, our knowledge on the enzymatic mechanisms allow the modification and the improvement of the original activities through a site directed mutagenesis method. Two main projects can be cited as examples:

- **Chemoenzymatic synthesis of *S*-glycosides**

One of the projects is based on the development of novel pathway to synthesize *S*-glycopeptides. *S*-glycosylation are highly unknown post-translational modifications.<sup>2</sup> They attracted a particular interest because of the reactivity of the sulfur atom that facilitates the specific sugar-amino acid coupling and also to the resistance of *S*-glycosidic bonds to chemical and enzymatic hydrolysis.<sup>12</sup> Thus, *S*-glycosidic linkages are considered as interesting element for the synthesis of therapeutic agents to treat a variety of pathological conditions, such as cancer and infectious diseases.<sup>13,14,15</sup> However, few data are available about *S*-glycosyltransferases that are responsible for the synthesis of *S*-glycoproteins, and our group used one of this enzymes, UGT74B1, to generate new *S*-glycosides.<sup>16</sup> Furthermore, the team got interested in other class of enzymes, mutated GHs, which are able to create thioglycosidic linkages.<sup>17,18</sup> First a wild type glycosidase, *DtGly* from thermophilic bacteria *Dictyoglomus thermophilum* was cloned, overexpressed and enzymatically characterized. It offered good transglycosylation activity. Then, mutants were created through site-directed mutagenesis and used to produce a range of *S*-glycoconjugates (Figure I-2).<sup>19</sup> Excellent yields of products were obtained without any protection and deprotection steps. The obtained *S*-glycolipids can exhibit hydrogelcation properties, and other act as “smart material” for cosmetics.<sup>20</sup> Thanks to directed evolution the team is able to modify and enhance enzyme properties and overcome their limits.<sup>21</sup>

## Glycoside hydrolase Identification

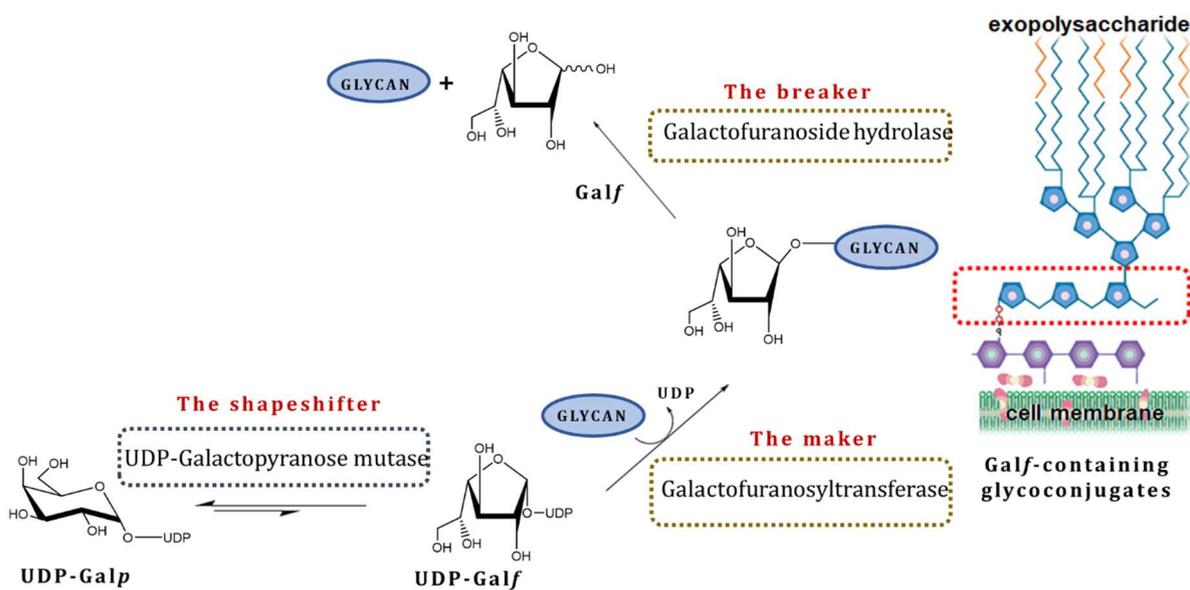


**Figure I- 2:** Identification and characterization of GHs and generation of thioligases through site directed mutagenesis strategy.

- **Galactofuranose metabolism in Leishmania**

Galactofuranose (Gal<sub>f</sub>) is a rare sugar that is found in many pathogenic species such as *Mycobacterium tuberculosis*, *Aspergillus* and *Leishmania*, but is absent in mammals.<sup>22,23</sup> The presence of Gal<sub>f</sub> residues are considered as antigenic epitopes highly immunogenic in mammals.<sup>24</sup> The galactofuranose became an interesting potential therapeutic target molecule to eliminate pathogenicity of some species. The role of Gal<sub>f</sub> was not completely elucidated, but it seemed to be involved in the virulence and infection of several microorganisms.<sup>25,26</sup> Thus, studying its biosynthesis pathway could be exploited to synthesize inhibitors and anti-virulent agents. The glycobiochemistry and enzymology team is interested in the metabolism of galactofuranose in *Leishmania*. This parasite is the causative agent of leishmaniasis, which is a set of diseases classified as a neglected tropical diseases (NTD) by the World Health Organization (WHO).<sup>27</sup> It is transmitted to humans by a bite of a *Phlebotome*.<sup>28</sup> It poses a real public health problem, globally, with 15 million people affected. In addition, *Leishmania* infect animals leading to the creation of a zoonotic reservoir.<sup>29,30</sup> The emergence of drug resistance parasite is considered as one of the greatest existing threats to human health.<sup>31,32</sup> Moreover, the available treatments are too expensive and have low efficiency.<sup>33,34</sup> It is then urgent to find new anti-leishmanial drugs with novel mechanisms.

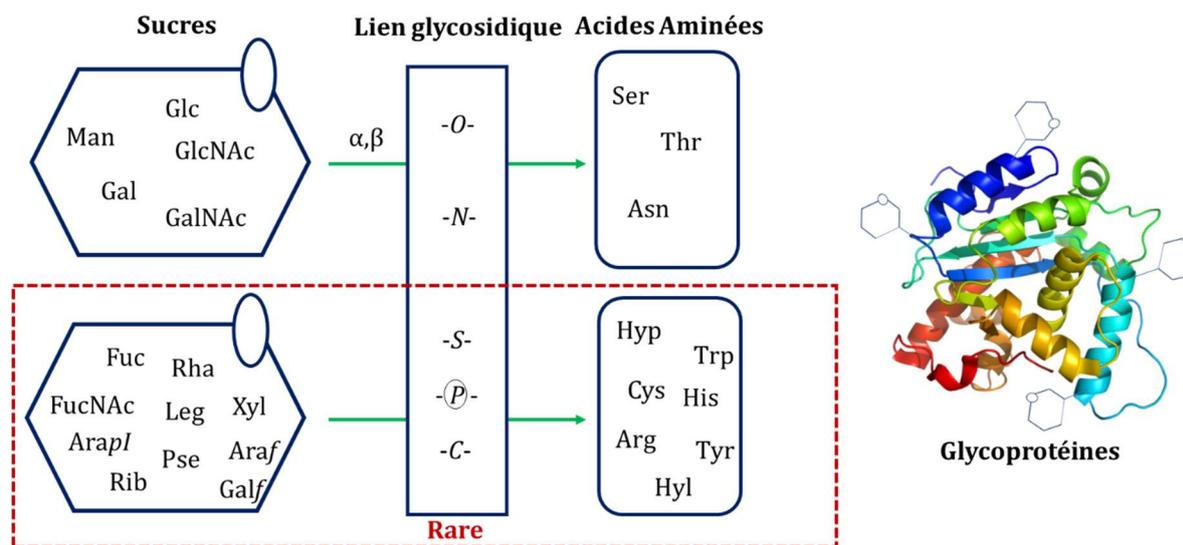
Galactofuranose-containing glycoconjugates are considered as therapeutic targets of leishmaniasis as well as other diseases such as tuberculosis, and aspergillosis.<sup>35,36</sup> It is then important to study their biosynthesis. Some key enzymes are involved in this process (Figure I-3). First, there is the “shapeshifter” or UDP-galactopyranose mutase that is responsible for the conversion of UDP-Galp to UDP-Galf.<sup>25</sup> Then, the “breaker” or furanoside hydrolase that is responsible for the degradation of galactofuranose from glycoconjugates.<sup>37,38</sup> Finally, there is the “maker” or galactofuranosyltransferase (Gal/Ts), enzymes that catalyzes the incorporation of Galf moieties onto polysaccharides.<sup>39</sup> The study of these enzymes is considered as essential step for understanding the biosynthesis of galactofuranose-containing glycoconjugates and revealing their biological role in the virulence of parasites. The enzymology and glycobiochemistry team is currently investigating these three enzymes. The PhD topic that will be described herein deals with four genes, *lpg1*, *lpg1*, *lpg1R* and *lp1G* that were identified in the genome of *Leishmania major* and were predicted to encode for four galactofuranosyltransferases.<sup>40</sup> We will present the strategies used to express and characterize these four putative Gal/Ts with the aim that this work will pave the way for the development of chemoenzymatic synthesis of furanosyl-containing glycoconjugates as well as the design of improved drugs against leishmaniasis.



**Figure I- 3:** Galactofuranose-containing glycoconjugates biosynthesis pathway with the key enzymes: UDP-Galactopyranose mutase (The shapeshifter), the galactofuranosyltransferase (The maker), and the galactofuranosyl hydrolase (The breaker).



Les sucres jouent un rôle clé dans de nombreux processus biologiques tels que la croissance, le développement, le fonctionnement, la structure, la communication et la survie des cellules de toutes sortes d'espèces vivantes.<sup>1,2</sup> Les glycanes peuvent également être attachés aux protéines ce qui est régulé pendant les étapes de modification post-traductionnelles qui se produisent chez les espèces eucaryotes. La glycosylation des protéines est très importante pour leur stabilité, leur structure ainsi que leur activité. Une erreur de glycosylation sur les protéines peut entraîner diverses pathologies, telles que des maladies auto-immunes et le cancer.<sup>3,4</sup> À l'inverse des protéines, la glycosylation ne peut pas être prédite à partir d'une matrice d'ADN.<sup>5</sup> Deux classes d'enzymes, les glycosyltransférases (GT) et les glycoside hydrolases (GH), sont responsables de la biosynthèse et du métabolisme des glycoconjugués. Les GTs transfèrent les monosaccharides sur un accepteur approprié, tandis que les GHs catalysent l'hydrolyse des liaisons glycosidiques entraînant la libération de glucides.<sup>6</sup> De nos jours, les données disponibles sur les différents glycosides et leurs liaisons ne représentent que «la partie émergée de l'iceberg ». En effet, plusieurs sucres rares et des liaisons glycosidiques sont peu connus (Figure I-1). Nous devons donc élargir nos connaissances afin de pouvoir synthétiser des composés naturels de type glycoconjugués qui peuvent être utilisés dans de nombreux domaines tels que les produits pharmaceutiques, les cosmétiques, les aliments, les matériaux et les ressources renouvelables.<sup>7,8</sup> À ce jour, les glycochimistes ont développé des outils efficaces pour synthétiser avec succès des dérivés de glucides. Cependant, ces méthodes nécessitent de longues procédures, incluant des étapes de protection et de déprotection.<sup>9</sup>



**Figure I-1** : Un résumé de sucres et liens glycosidiques associés aux peptides retrouvés dans la nature. (Adapté de Pierre Lafite and Richard Daniellou, 2012)<sup>2</sup>

L'équipe de glycobiologie et enzymologie à l'Institut de Chimie Organique et Analytique (ICOA UMR CNRS 7311), dirigée par le professeur Richard DANIELLOU, s'intéresse à l'étude des liaisons

glycosidiques et des sucres rares. Ils découvrent et caractérisent de nouvelles enzymes qui sont impliquées dans la biosynthèse et le métabolisme des glucides. Ces enzymes sont utilisées comme biocatalyseurs, ce qui représente un outil puissant et éco-compatible pour la synthèse des glucides naturels et de leurs analogues. En effet, les enzymes peuvent présenter une spécificité et une régiosélectivité élevées vis-à-vis de leur substrat. De plus, elles fonctionnent en milieu aqueux.<sup>10</sup> Ces caractéristiques sont aujourd'hui indispensables dans l'industrie, car elles offrent de nouvelles solutions pour les réactions chimiques. De plus, elles remplacent les processus chimiques qui nécessitent plusieurs étapes ainsi qu'un long temps réactionnel et qui requièrent l'utilisation de solvant toxiques. Ils suivent ainsi les règles de la « chimie verte », qui inclut des bioprocédés respectueux de l'environnement.<sup>11</sup> Parallèlement, nos connaissances sur les mécanismes enzymatiques ont permis de modifier et d'améliorer les activités d'origine grâce à une méthode de mutagenèse dirigée. Deux projets principaux peuvent être cités comme exemples :

- **Synthèse chimioenzymatique des S-glycoconjugués**

L'un des projets repose sur le développement de nouvelles voies de synthèse des S-glycopeptides. La S-glycosylation est une modification post-traductionnelle peu étudiée<sup>2</sup> qui a attiré un intérêt particulier en raison de la réactivité de l'atome de soufre qui assure le couplage spécifique sucre-acide aminé et également de la résistance des liaisons S-glycosidiques à l'hydrolyse chimique et enzymatique.<sup>12</sup> Ainsi, les liaisons S-glycosidiques sont considérées comme un élément intéressant pour la synthèse d'agents thérapeutiques pour traiter différentes pathologies, telles que le cancer et les maladies infectieuses.<sup>13,14,15</sup> Cependant, peu de données sont disponibles sur les S-glycosyltransférases responsables de la synthèse des S-glycoprotéines. L'équipe utilise une de ces enzymes, UGT74B1, qui génèrent de nouveaux S-glycosides.<sup>16</sup> De plus, l'équipe s'est intéressée à d'autres classes d'enzymes, les GH mutées, capables de créer des liaisons thioglycosidiques.<sup>17,18</sup> Tout d'abord, une glycosidase de type sauvage, DtGly provenant de bactéries thermophiles *Dictyoglomus thermophilum*, a été clonée, surexprimée et caractérisée d'un point de vue enzymologique. Cette enzyme offre une bonne activité de transglycosylation. Ensuite, des mutants ont été créés par mutagenèse dirigée et ont été utilisés pour produire des rares S-glycoconjugués (Figure I-2).<sup>19</sup> Des produits ont été obtenus avec d'excellents rendements et sans aucune étape de protection et de déprotection. Certains S-glycolipides ont présenté des propriétés d'hydrogélification, et d'autres agissent comme « matériel intelligent » pour les cosmétiques.<sup>20</sup> Grâce à l'évolution dirigée, il est possible de modifier et d'améliorer les propriétés des enzymes et de dépasser leurs limites.

Identification des Glycosidases

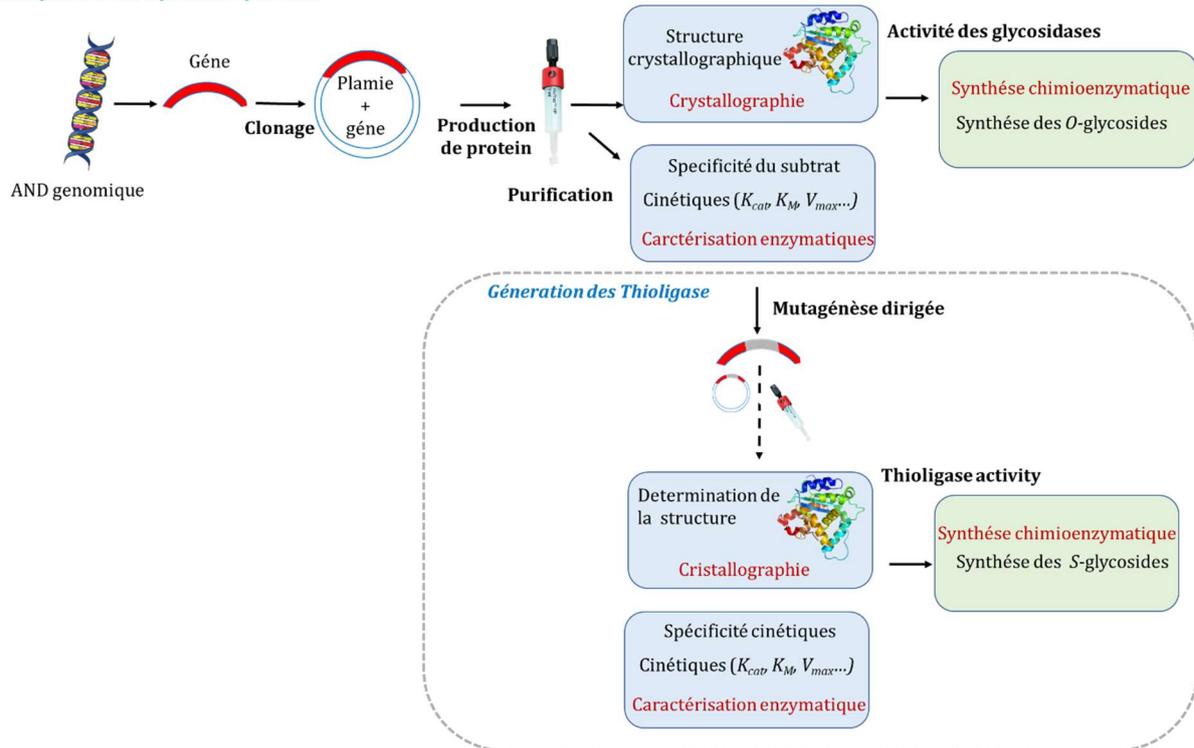


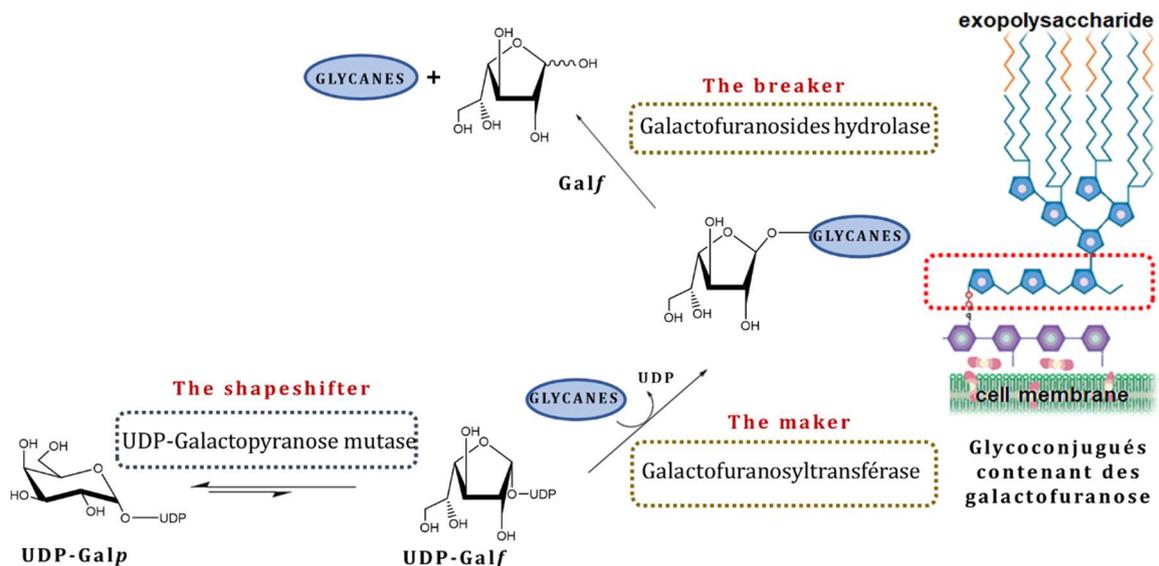
Figure I-2 : Identification et caractérisation des GHs et génération des thioligases selon la stratégie la mutagenèse dirigée.

• Le métabolisme du galactofuranose chez *Leishmania*

Le galactofuranose (Gal<sub>f</sub>) est un sucre rare retrouvé dans de nombreuses espèces pathogènes telles que *Mycobacterium tuberculosis*, *Aspergillus* et *Leishmania*. Cependant, il est absent chez les mammifères. La présence des résidus Gal<sub>f</sub> est considérée comme un épitope antigénique hautement immunogène chez les mammifères.<sup>22,23,24</sup> Le galactofuranose est ainsi devenu une molécule cible thérapeutique potentielle qui pourrait être utilisée pour altérer la pathogénicité de certaines espèces. Le rôle de Gal<sub>f</sub> n'a pas été complètement élucidé, mais il semble être impliqué dans la virulence et l'infection de plusieurs microorganismes.<sup>25,26</sup> Ainsi, l'étude de sa voie de biosynthèse pourrait être exploitée afin d'envisager la synthèse d'inhibiteurs et d'agents anti-virulents. L'équipe de glycobiochimie et enzymologie s'intéresse au métabolisme du galactofuranose chez *Leishmania*. Ce parasite est l'agent responsable de la leishmaniose, qui est un ensemble de maladies classées parmi les maladies tropicales négligées par l'Organisation mondiale de la santé (OMS).<sup>27</sup> Il est transmis aux humains par une piqûre de phlébotome.<sup>28</sup> Cela représente un véritable problème de santé publique à l'échelle mondiale, avec 15 millions de personnes touchées. De plus, *Leishmania* est capable d'infecter les animaux, créant ainsi un réservoir zoonotique.<sup>29,30</sup> L'émergence des parasites résistants aux antibiotiques est considérée comme l'une des plus grandes menaces pour la santé humaine.<sup>31,32</sup> De plus, les traitements

disponibles dans le marché sont trop coûteux et ont une faible efficacité.<sup>33,34</sup> Il est alors urgent de trouver de nouveaux anti-leishmanies dotés de mécanismes novateurs.

Les glycoconjugués contenant du galactofuranose sont considérés comme des cibles thérapeutiques de la leishmaniose ainsi que d'autres maladies telles que la tuberculose et l'aspergillose.<sup>35,36</sup> Il est alors important d'étudier leur biosynthèse. Certaines enzymes clés sont impliquées dans ce processus (Figure I-3). Premièrement, il y a le « shapeshifter » ou UDP-galactopyranose mutase qui est responsable de la conversion de UDP-Galp en UDP-Galf.<sup>25</sup> Ensuite, le « Breaker » ou furanoside hydrolase responsable de la dégradation du galactofuranose à partir des glycoconjugués.<sup>37,38</sup> Enfin, il existe le « maker » ou galactofuranosyltransférase (Gal $f$ Ts), enzymes qui catalysent l'incorporation de molécules de Galf sur les polysaccharides.<sup>39</sup> L'étude de ces enzymes est considérée comme une étape essentielle pour comprendre la biosynthèse des glycoconjugués contenant du galactofuranose et révéler leur rôle biologique dans la virulence des parasites. L'équipe enzymologie et glycobiologie étudie actuellement ces trois enzymes. Le sujet de thèse qui sera décrit ici concerne l'études de quatre gènes, *lpg1*, *lpg1*, *lpg1R* et *lp1G*, identifiés dans le génome de *Leishmania major* qui ont été prédits pour coder pour quatre galactofuranosyltransférases.<sup>40</sup> Nous présenterons les stratégies utilisées pour exprimer et caractériser ces potentiels Gal $f$ Ts. Ce travail a pour but d'ouvrir la voie au développement des méthodes chimio-enzymatiques pour la synthèse des glycoconjugués qui contiennent de galactofuranosides ainsi qu'à la conception de nouvelles molécules anti-leishmaniose.



**Figure I-3 :** Voie de synthèse de glycoconjugués contenant galactofuranose avec les enzymes clés : UDP-Galactopyranose mutase (The shapeshifter), la galactofuranosyltransférases (The maker), et la galactofuranosyl hydrolase (The breaker).

## Bibliography

1. Varki, A. Biological roles of oligosaccharides: All of the theories are correct. *Glycobiology* **3**, 97–130 (1993).
2. Lafite, P. & Daniellou, R. Rare and unusual glycosylation of peptides and proteins. *Nat. Prod. Rep.* **29**, 729–738 (2012).
3. Taniguchi, N. & Kizuka, Y. Glycans and cancer: Role of N-Glycans in cancer biomarker, progression and metastasis, and therapeutics. *Adv. Cancer Res.* **126**, 11–51 (2015).
4. Maverakis, E. *et al.* Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review. *J. Autoimmun.* **57**, 1–13 (2015).
5. Ajit Vark, iHudson H. F. and A. E. M. Overview of Glycoconjugate Analysis. *Curr .Protoc. Protein Sci.* **12.**, 1–16 (2010).
6. Hancock, S. M., Vaughan, M. D. & Withers, S. G. Engineering of glycosidases and glycosyltransferases. *Curr. Opin. Chem. Biol.* **10**, 509–519 (2006).
7. Solá, R. J. & Griebenow, K. A. I. Effects of Glycosylation on the Stability of Protein Pharmaceuticals. *Biochemistry* **98**, 1223–1245 (2010).
8. Ati, J., Lafite, P. & Daniellou, R. Enzymatic synthesis of glycosides: from natural O- and N-glycosides to rare C- and S-glycosides. *Beilstein J. Org. Chem.* **13**, 1857–1865 (2017).
9. Gandolfi-Donadío, L., Santos, M., De Lederkremer, R. M. & Gallo-Rodriguez, C. Synthesis of arabinofuranose branched galactofuran tetrasaccharides, constituents of mycobacterial arabinogalactan. *Org. Biomol. Chem.* **9**, 2085–2097 (2011).
10. Narancic, T., Davis, R., Nikodinovic-Runic, J. & O' Connor, K. E. Recent developments in biocatalysis beyond the laboratory. *Biotechnol. Lett.* **37**, 943–954 (2015).
11. Drepper, T. *et al.* Novel biocatalysts for white biotechnology. *Biotechnol. J.* **1**, 777–786 (2006).
12. Witczak, Z. J. & Culhane, J. M. Thiosugars: New perspectives regarding availability and potential biochemical and medicinal applications. *Appl. Microbiol. Biotechnol.* **69**, 237–244 (2005).
13. Agrawal, S. *et al.* Insulin and novel thioglycosides exert suppressive effect on human breast and colon carcinoma cells. *Oncotarget* **8**, 114173–114182 (2017).
14. Paul, B. & Korytnyk, W. S-, N-, and O-glycosyl derivatives of 2-acetamido-2-deoxy-D-glucose

- with hydrophobic aglycons as potential chemotherapeutic agents and N-acetyl- $\beta$ -D-glucosaminidase inhibitors. *Carbohydr. Res.* **126**, 27–43 (1984).
15. Blanc-Muesser, M. *et al.* Spacer-modified disaccharide and pseudo-trisaccharide methyl glycosides that mimic maltotriose, as competitive inhibitors for pancreatic alpha-amylase: a demonstration of the 'clustering effect'. *Carbohydr. Res.* **224**, 59–71 (1992).
  16. Marroun, S. *et al.* UGT74B1 from *Arabidopsis thaliana* as a versatile biocatalyst for the synthesis of desulfoglycosinolates. *Org. Biomol. Chem.* **14**, 6252–6261 (2016).
  17. Henrissat, B. Glycosidase families. *Biochem. Soc. Trans.* **26**, 153–156 (1998).
  18. Vasella, A., Davies, G. J. & Böhm, M. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **6**, 619–629 (2002).
  19. Guillotin, L., Cancellieri, P., Lafite, P., Landemarre, L. & Daniellou, R. Chemo-enzymatic synthesis of 3-O-( $\beta$ -D-glycopyranosyl)-sn-glycerols and their evaluation as preservative in cosmetics. *Pure Appl. Chem.* **89**, 1295–1304 (2017).
  20. Peyrot, C., Vivès, T., Legentil, L., Lemiègre, L. & Daniellou, R. Microwave-Assisted Reduction of Nitroarenes by Aminothiophenol/dithiotreitol. *ChemistrySelect* **2**, 5214–5217 (2017).
  21. Porter, J. L., Rusli, R. A. & Ollis, D. L. Directed Evolution of Enzymes for Industrial Biocatalysis. *ChemBioChem* **17**, 197–203 (2016).
  22. De Lederkremer, R. M. & Colli, W. Galactofuranose-containing glycoconjugates in trypanosomatids. *Glycobiology* **5**, 547–552 (1995).
  23. Houseknecht, J. B. & Lowary, T. L. Chemistry and biology of arabinofuranosyl- and galactofuranosyl-containing polysaccharides. *Curr. Opin. Chem. Biol.* **5**, 677–682 (2001).
  24. Marino, C., Rinflerch, A. & de Lederkremer, R. M. Galactofuranose antigens, a target for diagnosis of fungal infections in humans. *Futur. Sci. OA* **3**, 1–14 (2017).
  25. Kleczka, B. *et al.* Targeted gene deletion of *Leishmania major* UDP-galactopyranose mutase leads to attenuated virulence. *J. Biol. Chem.* **282**, 10498–10505 (2007).
  26. Pan, F. *et al.* Cell Wall Core Galactofuran Synthesis Is Essential for Growth of Mycobacteria. *J. Bacteriol.* **183**, 3991–3998 (2001).
  27. Reithinger, R. *et al.* Cutaneous leishmaniasis. *Lancet Infect. Dis.* **7**, 581–596 (2007).
  28. Torres-Guerrero, E., Quintanilla-Cedillo, M. R., Ruiz-Esmenjaud, J. & Arenas, R.

- Leishmaniasis: a review. *F1000Research* **6**, 750 (2017).
29. Akhoundi, M. *et al.* A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies. *PLoS Negl. Trop. Dis.* **10**, 1–40 (2016).
  30. Consuelo V. David, N. C. Cutaneous and mucocutaneous leishmaniasis. *Dermatol. Ther.* **22**, 491–502 (2009).
  31. Croft, S. L., Sundar, S. & Fairlamb, A. H. Drug Resistance in Leishmaniasis Drug Resistance in Leishmaniasis. *Society.* **19**, 111–126 (2006).
  32. Croft, S. L. Monitoring drug resistance in leishmaniasis. *Trop. Med. Int. Heal.* **6**, 899–905 (2001).
  33. Lamotte, S., Späth, G. F., Rachidi, N. & Prina, E. The enemy within: Targeting host-parasite interaction for antileishmanial drug discovery. *PLoS Negl. Trop. Dis.* **11**, e0005480 (2017).
  34. Banjara, M. R. *et al.* Visceral Leishmaniasis Clinical Management in Endemic Districts of India, Nepal, and Bangladesh. *J.Trop.Med* **2**, 1–8 (2012).
  35. Rose, N. L. *et al.* Expression, Purification, and Characterization of a Galactofuranosyltransferase Involved in Mycobacterium tuberculosis Arabinogalactan Biosynthesis. *J. Am. Chem. Soc.* **128**, 6721–6729 (2006).
  36. Komachi, Y. *et al.* gfsA encodes a novel galactofuranosyltransferase involved in biosynthesis of galactofuranose antigen of O-glycan in Aspergillus nidulans and Aspergillus fumigatus. *Mol. Microbiol.* **90**, 1054–1073 (2013).
  37. Ramli, N., Fujinaga, M., Tabuchi, M., Takegawa, K. & Iwahara, S. Isolation and Characterization of a Novel Endo- $\beta$ -galactofuranosidase from *Bacillus* sp. *Biosci. Biotechnol. Biochem.* **59**, 1856–1860 (1995).
  38. Wallis, G. L. F., Hemming, F. W. & Peberdy, J. F.  $\beta$ -Galactofuranoside glycoconjugates on conidia and conidiophores of Aspergillus niger. *FEMS Microbiol. Lett.* **201**, 21–27 (2001).
  39. Ryan, K. A., Garraway, L. A., Descoteaux, A., Turcot, S. J. & Beverley, S. M. Isolation of virulence genes directing surface glycosyl-phosphatidylinositol synthesis by functional complementation of Leishmania. *Proc. Natl. Acad. Sci. USA.* **90**, 8609–8613 (1993).
  40. Zhang, K., Barron, T., Turco, S. J. & Beverley, S. M. The LPG1 gene family of Leishmania major. *Mol. Biochem. Parasitol.* **136**, 11–23 (2004).

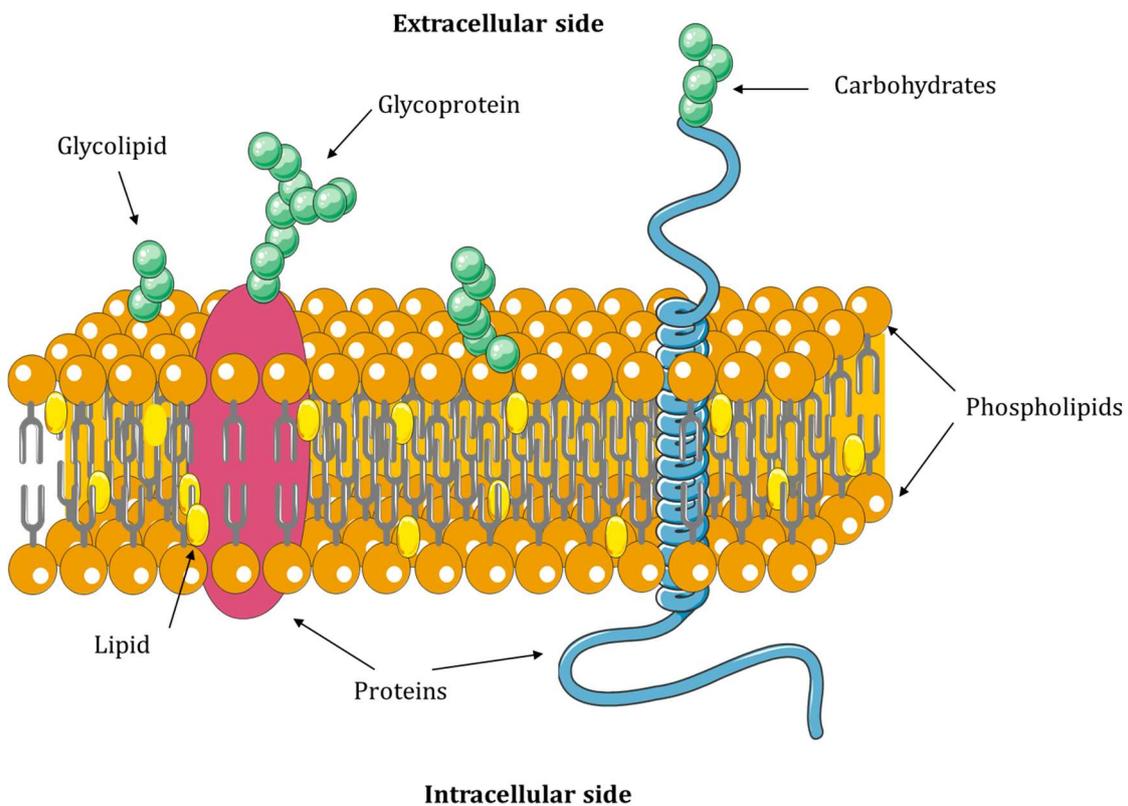
## ***Chapter II: Bibliography***

I.	Glycoconjugates .....	36
A.	Biological role of glycoconjugates .....	36
B.	Structures .....	37
II.	D-Galactofuranose .....	37
A.	Therapeutic relevance of galactofuranose.....	38
1.	In bacteria.....	38
1.	In fungi.....	39
3.	In protozoa .....	40
B.	Biosynthesis of glycoconjugates containing galactofuranose .....	41
1.	UDP-Galactofuranose transporter .....	42
2.	UDP-Galactopyranose mutase .....	43
3.	Galactofuranosidases.....	44
4.	Galactofuranosyltransferases .....	44
a)	Classification .....	44
b)	Structure.....	45
c)	Mechanism .....	46
III.	Characterized galactofuranosyltransferases .....	47
A.	GlfT1 and GlfT2 from <i>Mycobacterium tuberculosis</i> .....	47
B.	GfsA from <i>Aspergillus</i> .....	50
C.	WbbI from <i>E. coli</i> K-12.....	51
D.	Summary .....	51
IV.	Galactofuranose-containing glycoconjugates of <i>Leishmania</i> : therapeutic target for leishmaniasis.....	52
A.	Leishmaniasis as a neglected tropical disease .....	52
1.	Causing agent, <i>Leishmania</i> .....	52
2.	Life cycle .....	53
3.	Clinical forms of leishmaniasis .....	54
a)	Cutaneous leishmaniasis .....	54
b)	Diffuse cutaneous leishmaniasis.....	54
c)	Mucocutaneous leishmaniasis .....	55
d)	Visceral leishmaniasis .....	55
4.	Epidemiology .....	55
5.	Diagnosis and treatment .....	57
a)	Diagnosis .....	57
b)	Treatments .....	58
B.	Cell wall glycoconjugates as therapeutic targets.....	60

1. Lipophosphoglycans (LPGs) .....	62
a) Structure.....	62
b) Functions .....	63
(1) In vertebrate host .....	63
(2) In mammalian host.....	64
2 Glycosylinositol phospholipids (GIPLs).....	64
c) Structure.....	64
d) Functions .....	65
3. Biosynthesis of glycoconjugates .....	65
V. Objective of thesis.....	68
References .....	73

Carbohydrates are widely distributed in Nature.<sup>1</sup> They are known for many years as food sources or fuel cellular reactions.<sup>2</sup> In addition, they were overshadowed by other biopolymers such as oligonucleotides and proteins. Nevertheless, there was a situation reversal in the 1960's due to the work of Walkins. He stressed out the role of sugars in blood group antigenicity.<sup>3</sup> Indeed, human red-blood cells are covered by specific carbohydrates that help defining the blood type. These carbohydrates are part of the antigens that are recognized by the immune system. It is a small difference but it could be fatal in case of mismatch blood transfusion.<sup>4</sup> Since that discovery, chemists and biologists showed a remarkable interest for carbohydrates. These advancements gave rise to a new area called glycobiology, *i.e.* the study of carbohydrates in biological context.

It is important to highlight that carbohydrates can be covalently linked to proteins or lipids to generate respectively glycoproteins and glycolipids (Figure II-1). These species are part of the family of glycoconjugates. Nowadays, we know that glycoconjugates are extremely important and are involved in a vast array of biological processes.<sup>5</sup> Indeed, they are essential constituents of countless cells membrane and play crucial roles in cell-cell communication, signal transduction, pathogens infection and immune response. Thus, glycoconjugates are interesting therapeutic targets for drug design and diagnosis tool development.

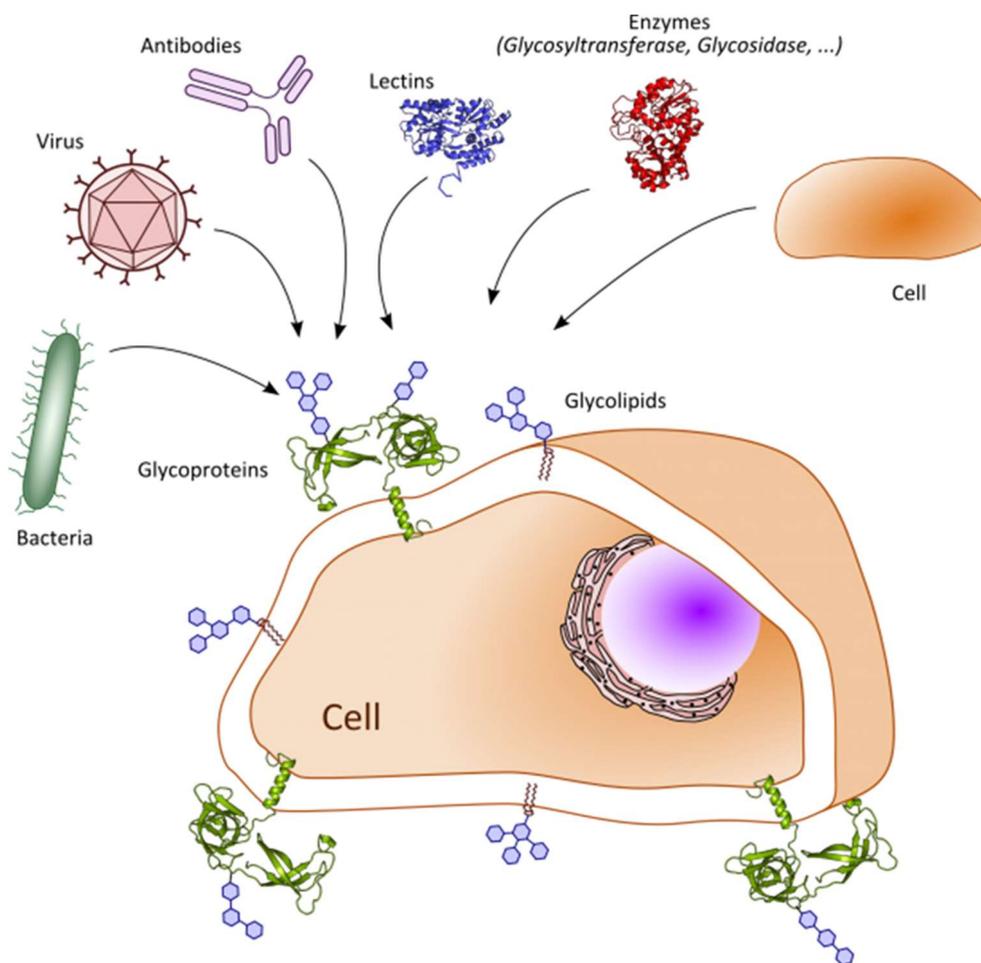


**Figure II- 1:**Glycoproteins and glycolipids in cell membrane

## I. Glycoconjugates

### A. Biological role of glycoconjugates

Cells surface are heavily decorated by diverse glycans. Therefore, it is not surprising that numerous biological events and human diseases conditions are linked to the presence, absence or modification of these glycoconjugates. The biological role of glycoconjugates span the spectrum from those that seems subtle, to those that appear to be crucial for cell growth and development<sup>6</sup> or survival and infection process of microorganisms that synthetize them<sup>7</sup> (Figure II-2). Moreover, they are critical for the stability and integrity of proteins.<sup>8</sup> Despite all efforts made to study them over the years, many of these glycoconjugates have not been assigned to any function.



**Figure II- 2:** Role of glycans in recognition and communication with other cells and microorganisms. (From Pierre Lafite and Richard Daniellou, 2011).<sup>9</sup>

Immune system is a perfect example to stress out the importance of glycoconjugates. Indeed, immune cells use glycan-binding proteins to recognize specific carbohydrates expressed by self or non-self (pathogen) cells.<sup>10</sup> In some cases, defects in mammalian glycosylation can influence self/non-self-recognition and then lead to autoimmune diseases.<sup>11,12</sup> Moreover, changes in cellular

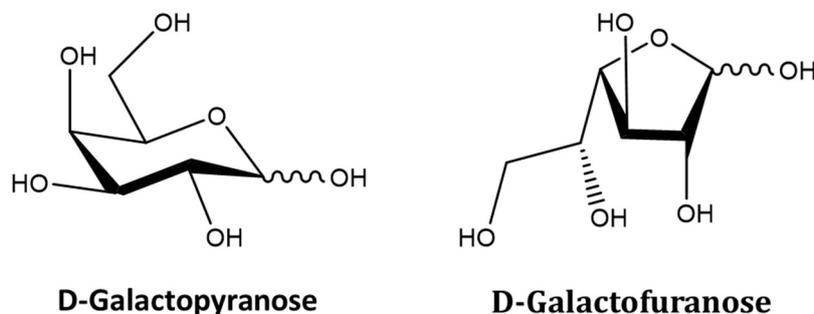
glycosylation signatures are often associated with cancer progression.<sup>13,14</sup> It is then important to study the structure of glycoconjugates and identify their functions.

## B. Structures

The study and the synthesis of glycoconjugates remain an ongoing challenge due to the remarkable structure diversity and complexity of carbohydrates. Indeed, each monosaccharide, which is the base unit of all carbohydrates, may be attached to another monosaccharide thanks to one of its several hydroxyl groups. In addition, the monosaccharide can have  $\alpha$  or  $\beta$  configurations and can adopt either the opened or one of the two ring forms: the pyranose form, which is six-membered ring and the furanose form, characterized by a five-membered ring. Moreover, the polysaccharide structures are not only linear, but can also be branched and their complexity is closely related to the monosaccharides number that constitutes them.

## II. D-Galactofuranose

For years, hexopyranosyl form was considered as the most important constituent of glycoconjugates. However, this last two decades showed that hexofuranosyl forms are found in various pathogenic and essentially non-pathogenic organisms such as bacteria<sup>15</sup>, protozoa<sup>16</sup>, fungi,<sup>17</sup> and plants.<sup>18</sup> Consequently, glycoconjugates-containing hexofuranoses have attracted interest of many glycoscientists for their potential biological applications.<sup>19</sup> Furanose is the thermodynamically disfavored form than pyranose form.<sup>20</sup> The most studied and widespread hexose that exists in furanosyl form is D-galactofuranose (D-Galf). It is found in countless glycoproteins and glycolipids of many human infectious microorganisms and is absent in mammalian cells.<sup>16,21</sup> D-galactose can exist in pyranose and furanose configurations (Figure II-3). However, only galactopyranose is a part of polysaccharide of the animal kingdom. Remarkably, galactofuranose has been identified several human infectious agent such as *Mycobacterium tuberculosis*, *Aspergillus*, and *Leishmania*.<sup>22,23</sup>

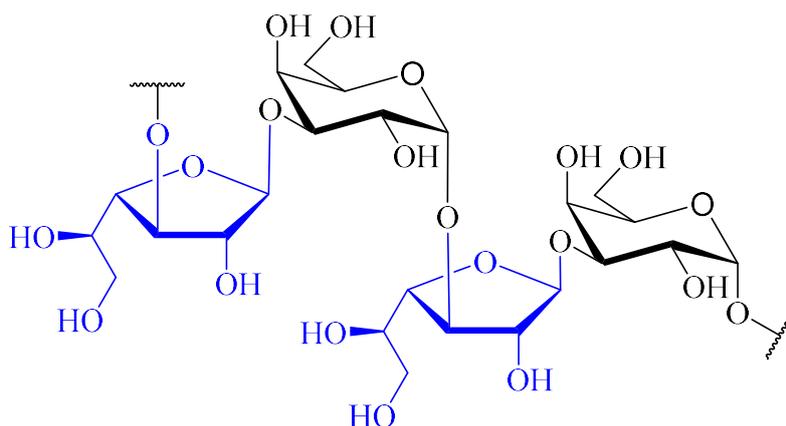


**Figure II- 3:** Furanose and pyranose conformations of D-galactose.

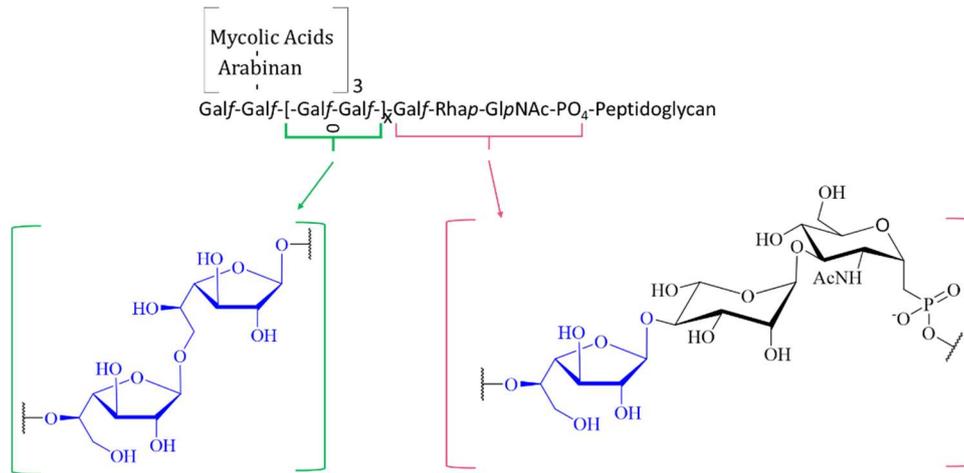
## A. Therapeutic relevance of galactofuranose

### 1. In bacteria

Both pathogenic Gram-negative and Gram-positive bacteria express oligosaccharides and glycoconjugates containing  $\alpha$ - or  $\beta$ -D-galactofuranose.<sup>24,25</sup> These structures are involved in several immunological reactions. First, there are different strains of *Escherichia coli* such as O124, and O164 and that cause extra-intestinal infections. They were reported to express D-Galf in their lipopolysaccharide (LPS) *O*-antigens. It can be part of the repeating units such as Galf- $\alpha$ (1-->4)GalpA<sup>26</sup>, Galf- $\beta$ (1-->3)GalNac<sup>27</sup> or branched by  $\beta$ -D-Glc at position-6.<sup>28</sup> Second, disaccharide Galf- $\beta$ (1-->3)Galp was identified in the *O*-antigen of *Klebsiella pneumonia* that has become important pathogens in nosocomial infections (Figure II-4).<sup>29</sup> Third, *Salmonella typhimium*, a Gram-negative bacterium that causes gastroenteritis in humans express in their LPS repeating unit Galf- $\alpha$ (1-->2)-L-Rha.<sup>30</sup> The *O*-antigens are an important virulence factors that influence the host-bacteria relationship at different levels by providing diverse functions such as resistance to complement action, which is a part of immune reaction, resistance to phagocytosis and tissue adhesion from the host.<sup>31,32</sup> Finally, there is the most studied and highly pathogenic *Mycobacterium tuberculosis* (Mtb) that causes tuberculosis disease and responsible for 1.4 million deaths worldwide. On its cell wall there is the mycolyl-arabinogalactan-peptidoglycan complex (mAGP) containing a galactan core of 30-35 alternating  $\beta$ (1-->5) and  $\beta$ (1-->6)Galf linkages. In addition there is Galf- $\beta$ (1-->4)-L-Rha unit that are connected to peptidoglycan (Figure II-5).<sup>32,33</sup> It was shown that galactofuranosyl residues are important for the viability and development of Mtb.<sup>34</sup>



**Figure II- 4:** Galactan moiety of the *O*-antigen of *Klebsiella pneumonia*

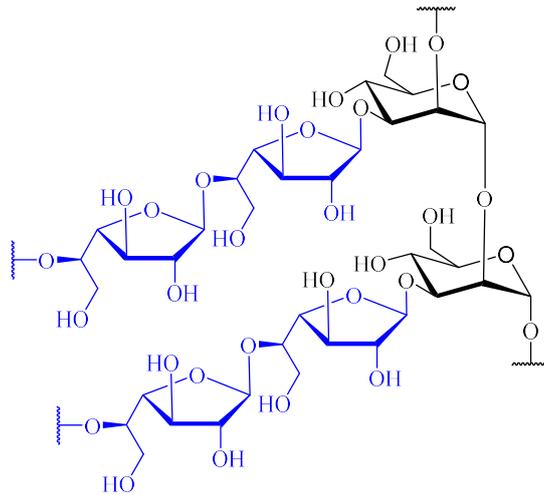


**Figure II- 5:** Structure of mAGP complex in Mtb.

During these last years, antibiotic resistant strains of Mtb have been spreading all over the world.<sup>35</sup> This is threatening the health of millions of people, especially immunocompromised patients. It is then urgent to find new drugs to fight the emergence of these resistant strains. Thus, galactofuranose-containing glycoconjugates have become interesting targets for designing new anti-tuberculosis drugs.

## 2. In fungi

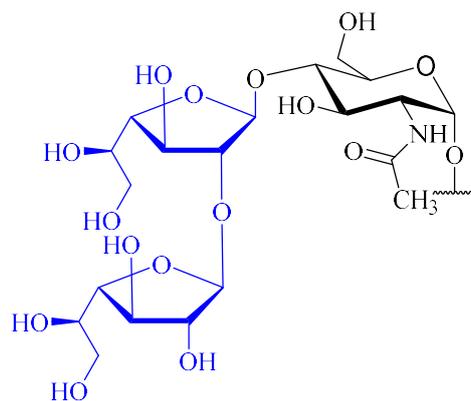
The most frequent case for human pathogen fungi is infection with *Aspergillus fumigatus* species. It causes aspergillosis diseases, an invasive fungal infection in immunosuppressed patients.<sup>36</sup> Some exopolysaccharides with immunological properties are produced during the growth of fungi. Both EPS from *Aspergillus* and *Penicillium* are constituted with a main core chain composed of (1-->2)- and (1-->6)-linked  $\alpha$ -D-mannopyransides branched at O-6 and O-3 with short side chain of Gal $\beta$ -(1-->5) (Figure II-6).<sup>37,17</sup> Another important D-Galf containing conjugates that has been described in fungi, is the glycosylinositol phosphorylceramide; It has been extracted from the mycopathogen *Paracoccidioides brasiliensis*; it is characterized by  $\alpha$ -D-Manp-(1-->3)-[ $\beta$ -D-Galf-(1-->-)]- $\alpha$ -D-Manp-(1-->2)-myo-inositol-(1-phosphoryl-1) ceramide.<sup>38</sup> In fungi, it has been shown that galactofuranose residues are immunodominant.<sup>39</sup> This makes them an interesting serological and taxonomic marker, in addition to their valuable use for the development of new diagnosis tools.



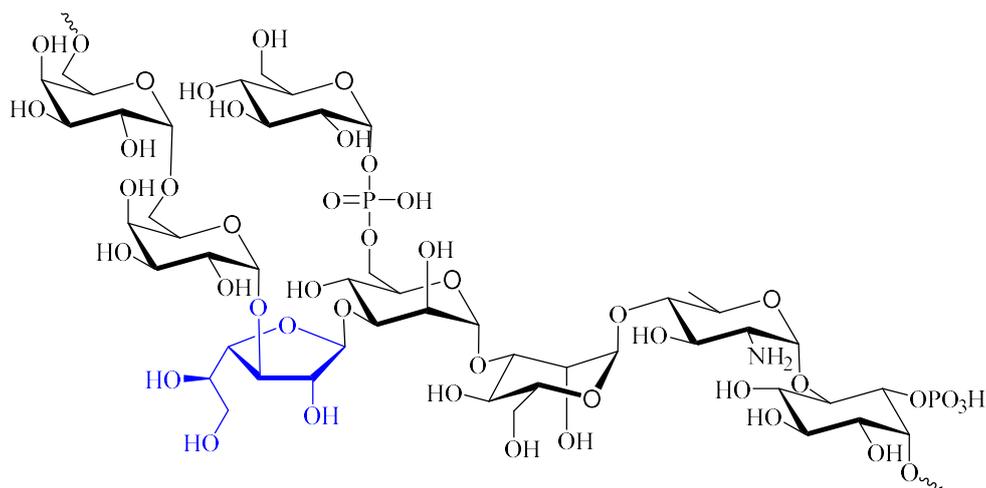
**Figure II- 6:** Galactomannan pattern as found in *Aspergillus fumigatus*.

### 3. In protozoa

Two important species of pathogenic parasite have been described for containing  $\beta$ -D-Galactofuranose in their glycoconjugates. First, *Trypanosoma cruzi*, the causative agent of Chagas Disease,<sup>40</sup> express glycoinositolphospholipids (GIPLs) with terminal Gal<sub>f</sub> unit linked to mannose, and have in their mucin core Gal<sub>f</sub>- $\beta$ (1-->4)Glc-Nac units (Figure II-7).<sup>41</sup> Second, *Leishmania spp.* responsible of leishmaniosis infection display on their surface GIPLs that contain external galactofuranose external unit attached to mannose at O-3 (Gal<sub>f</sub>- $\beta$ (1-->3)Man).<sup>42,43</sup> These parasites display also Lipophosphoglycans (LPGs) conjugates with internal Gal<sub>f</sub> unit also linked to mannose (Figure II-8). These glycoconjugates play an important role in the attachment of parasites in the vector insect midgut, and the infection process of immune cells of mammalian hosts. The presence of Gal<sub>f</sub> residues are considered as antigenic epitopes highly immunogenic in mammals.<sup>44</sup>



**Figure II- 7:** *Trypanosoma cruzi* mucin.



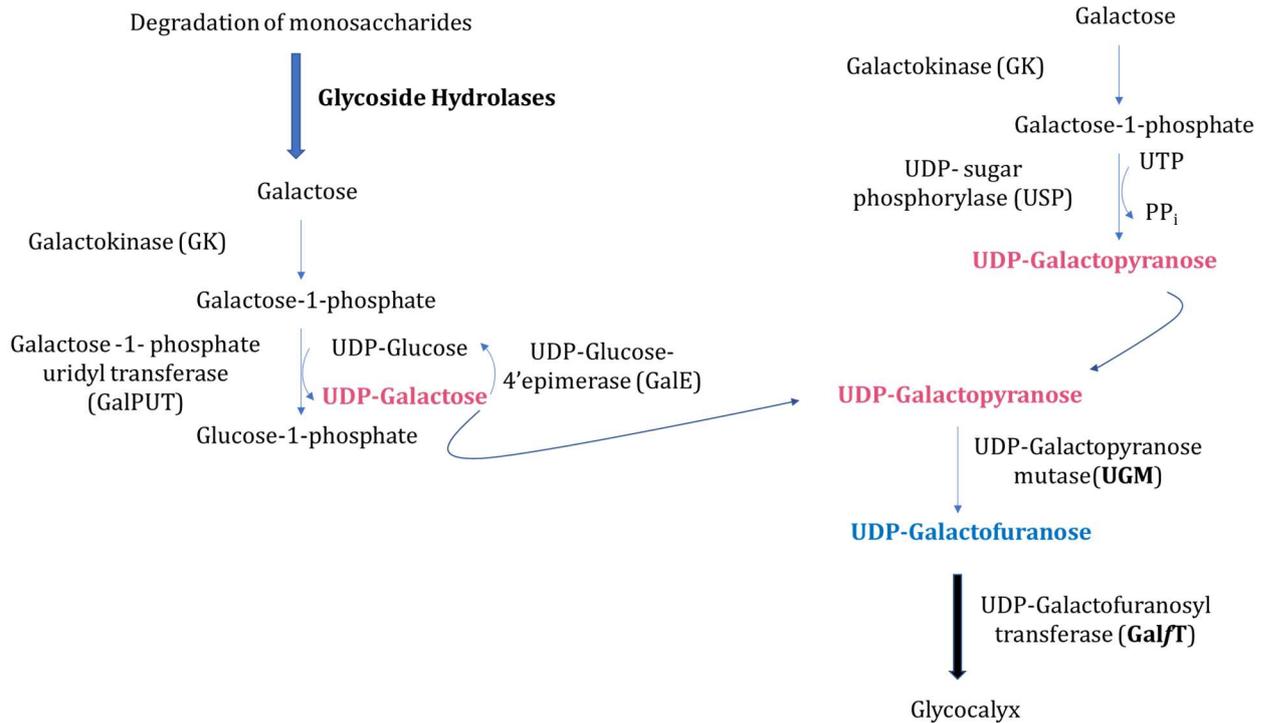
**Figure II- 8:** Lipophosphoglycan core from *Leishmania*.

Being a component of numerous polysaccharides and glycoconjugates in various human pathogens and absent in mammalian cells, galactofuranose became an interesting potential therapeutic target molecule to suppress or modulate pathogenicity of some species. The role of galactofuranose is not completely elucidated, but it seems to be involved in the virulence and infection of several microorganisms. Thus, studying its biosynthesis pathway could be exploited to synthesize inhibitors and anti-virulent agents. Moreover, this study could be useful to design a diagnostic tool to detect infectious organism containing galactofuranose.

### B. Biosynthesis of glycoconjugates containing galactofuranose

Galactofuranose needs to be activated as uridine diphosphate  $\alpha$ -galactofuranose (UDP-Galf) in order to be incorporated into glycans. UDP-Galf can be generated by two metabolic pathways (Figure II-9). The first one is the Galactose salvage pathway that consists on the degradation of monosaccharides from glycoconjugates within cells and transported back to the cytoplasm by specific transporters.<sup>45</sup> Most of the salvaged sugars are reused and activated once they reach the cytoplasm. Here salvaged  $\beta$ -Galactose is first converted to  $\alpha$ -Galactose by Galactose mutarotase (GMR).<sup>46</sup> This is followed by a direct phosphorylation reaction of  $\alpha$ -Galp at the anomeric carbon by Galactose kinase (GK) to form  $\alpha$ -Galactose-1-phosphate ( $\alpha$ -Galp-1-P).<sup>47</sup> Then, Gal-1-phosphate-uridylyltransferase (GalPUT) reacts with uridine triphosphate (UTP) to give Uridine diphosphate- $\alpha$ -Galactopyranose (UDP- $\alpha$ -Galp). This product is transformed to UDP-Galf by specific enzyme called UDP-galactopyranose mutase (UGM).<sup>48</sup>

The second one is due to *the novo* synthesis using the first step of glycolysis pathway.<sup>23</sup>  $\alpha$ -Glucopyranose is phosphorylated by hexokinase (HK) at C-6 to form  $\alpha$ -Glucose-6-Phosphate which is converted to  $\alpha$ -Glucose-1-Phosphate ( $\alpha$ -Glc-1-P) by Phosphoglucose mutase. Glucose-1-Phosphate-Uridyltransferase (GPUT) catalyses the production of UDP-  $\alpha$ -Glc $p$ . This is finally converted to UDP-  $\alpha$ -Gal $p$  by UDP-Glc-4-epimerase (UGE). UGM generate later UDP-Galf.<sup>49</sup>



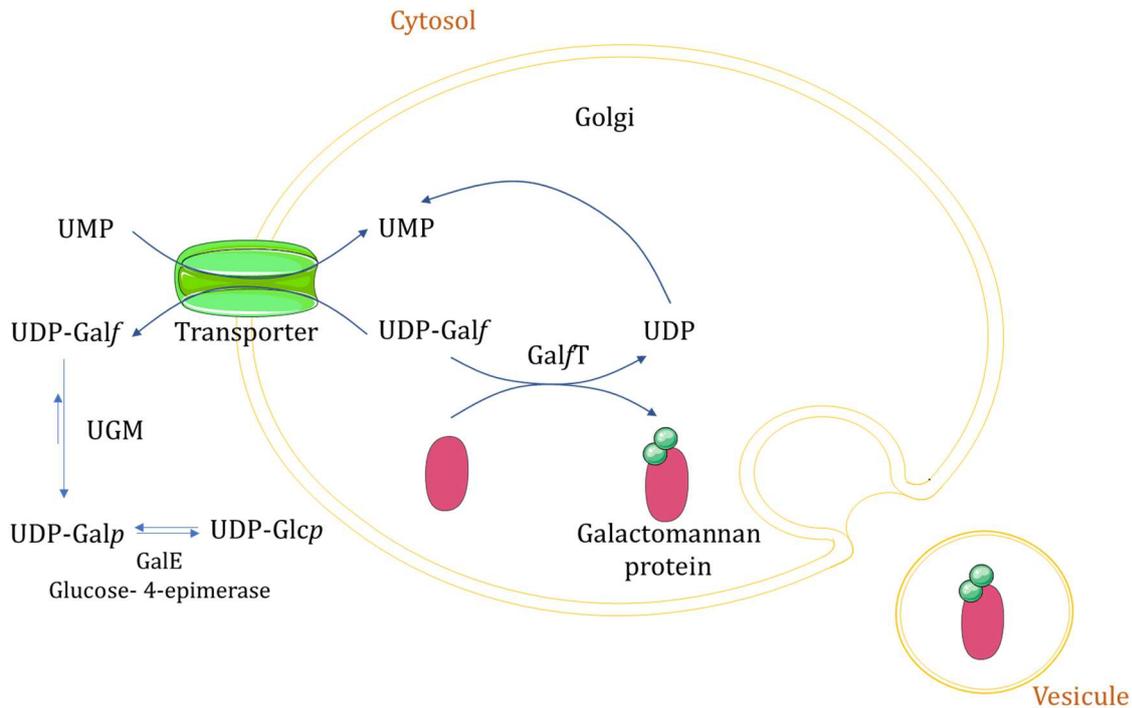
**Figure II- 9:** The two known pathways that generate UDP-Galactofuranose (adapter from Carla Marino et al., 2016)<sup>50</sup>

Specific transporters expressed only in eukaryotic species and three main enzymes, UDP-galactopyranose mutase, galactofuranosidase (*Gal*fase) and galactofuranosyltransferase (*Gal*fT), are needed for biosynthesis of UDP-*Gal*f and elongating galactans. Since *Gal*f is often associated with virulence in pathogenic species, besides its absence in mammalian glycoconjugates, these enzymes and transporters became interesting therapeutic targets to study.

### 1. UDP-Galactofuranose transporter

The synthesis of UDP-*Gal*f occurs in the cytosol.<sup>51</sup> Then the activated sugar-nucleotide is transported into the right compartment, which is the Golgi apparatus, for the incorporation of galactofuranose into glycan. Indeed, it was shown that in some eukaryotic species, the UDP-*Gal*f is synthesized in the cytosol but the glycan elongation occurs mostly in the Golgi apparatus (Figure II-10).<sup>52</sup> Thus, specific transporters must carry out the transfer of nucleotide sugars. Recently, putative transporters of UDP-*Gal*f were identified, such as *Glf*B in *Aspergillus fumigatus*, *Ug*TA in *Aspergillus nidulans*, *Ug*TA and *Ug*TB in *Aspergillus niger*.<sup>53,52</sup> Preliminary results indicate that the absence of these transporter alter the biosynthesis of *Gal*f-containing glycoconjugates. As an example, Jakob Engel et al. (2009) identified for the first time *glf*B gene that encodes for UDP-*Gal*f transporter in *Aspergillus fumigatus*. Gene deletion was correlated with the absence of galactofuranose into polysaccharide. They stressed the involvement of this transporter in *Gal*f

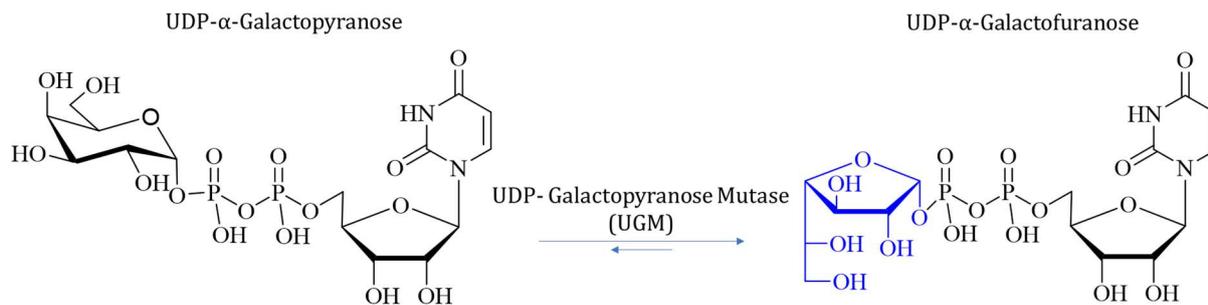
metabolism. Moreover, *glfB*<sup>-/-</sup> strains showed growth phenotype correlated with reduction of virulence.<sup>53</sup>



**Figure II- 10:** Transport of UDP-Galactofuranose in eukaryotes. (Adapted from Carla Marino et al., 2016).<sup>50</sup>

## 2. UDP-Galactopyranose mutase

UDP-Galactopyranose Mutase (UGM) is considered as the most studied enzymes involved in hexofuranosides biosynthesis at present. It is a flavoenzyme responsible for the conversion of UDP-Galp to UDP-Galf (Figure II-11). *In vitro*, this reaction is reversible but favors the formation of pyranose configuration because it is more stable. The gene encoding for UGM was first identified in *E. coli* in 1996.<sup>28,54</sup> It was followed by the discovery of other genes in *K. pneumoniae* and *M. tuberculosis*, *A. fumigatus*, *L. major* and *T. cruzi*.<sup>55,56,57,58</sup> Around twenty-nine structures of UGM from six different microorganisms have been registered in Protein Data Bank (PDB) up to day.<sup>59</sup> It is important to highlight that UGM is not expressed in mammals,<sup>22</sup> and it was shown that it is crucial for the survival of some infectious organisms. As an example, a knockout of UGM gene in *L. major* resulted in the absence of expression of LPG and the decrease of virulence of the parasite.<sup>51</sup> Also, mice infected with *L. major* clones lacking mutase gene showed an attenuated onset symptoms.<sup>51</sup> Several crystal structures of bacterial and *Aspergillus* UGMs alone or in complex with substrates or substrates analogues were published, allowing a better understanding of their mechanism, however no crystal of leishmanian UGM are available.<sup>60,61</sup>



**Figure II- 11:** Conversion of UDP- $\alpha$ -Galactopyranose to UDP-  $\alpha$ -Galactofuranose by UGM.

### 3. Galactofuranosidases

Galactofuranose-containing glycoconjugates are degraded by extracellular  $\beta$ -D-galactofuranosidases ( $\beta$ -D-Galfases). These class of enzymes are not well known. Some exo- and endo- galactofuranosidases activities have been detected in some species such as *Penicillium fellutanum*, *Helminthosporium sacchari* and *Aspergillus species*.<sup>62,63,64,65</sup> The presence of Galfase was recently detected in *T. cruzi* by using a polyclonal antibody raised against  $\beta$ -D-galactofuranosidase of *P. fellutanum*.<sup>66</sup> None of Galfase genes have been identified until recently by Takegaxa team. They expressed and characterized the first  $\beta$ -D-galactofuranosidases from *Streptomyces* Species.<sup>67</sup> However, no crystal data is available.

### 4. Galactofuranosyltransferases

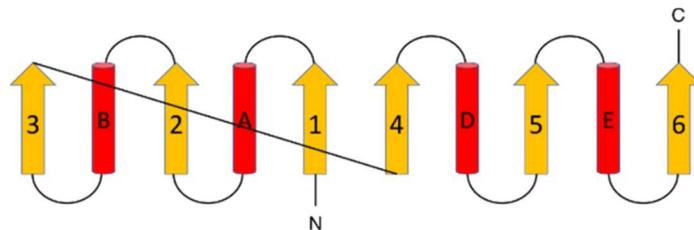
#### a) Classification

Galactofuranosyltransferases (GalTs) catalyze the transfer of specific sugar nucleotide donor onto a suitable acceptor for glycans and glycoconjugates biosynthesis.<sup>68,69</sup> They belong to the EC2.4.y.z family. The number EC is assigned to enzymes by International Union of Biochemistry and Molecular Biology (IUBMB) in order to classify them according to their specificity to some substrate or product. Indeed, “y” outlines the transferred sugar, and “z” describes the specificity of the acceptor. However, this means that enzymes should be classified once their activity and specificity are experimentally determined. It is not relevant to classify enzymes according to their substrate specificity because some enzymes from different species could catalyze the same reactions, or the formation of the same linkage. Furthermore, the same enzyme can promote the attachment to various acceptors. Likewise, identifying their substrate specificity does not give information about their structure and folding. So, for these reasons, another method to classify carbohydrate-active enzymes was developed by Henrissat. It is based on amino acid sequence similarities unrelatedly to their substrate specificity.<sup>69</sup> Then, enzymes which display homology sequencing, small common and conserved domains are classified in the same family, this similarity could provide an evidence that they are related by evolution. Up to 105 GT families are registered in CAZY (Carbohydrate active enzymes) data base, which contain four GT families that

correspond to galactofuranosyltransferases; GT2, GT4, GT31 and GT40. (<http://www.cazy.org/> / September 2018),

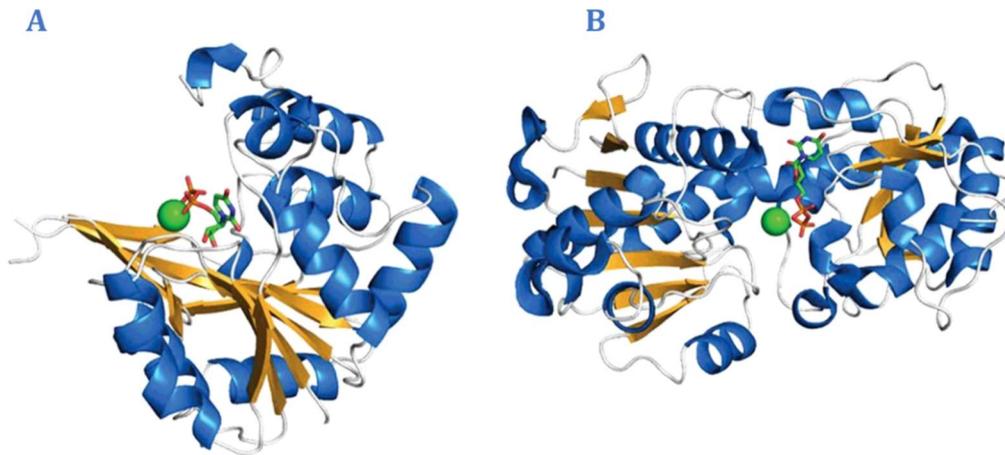
### b) Structure

Up to now, many sequence families have been classified, however, limited folds types have been observed. GalTs are expected to adopt GT-A fold, because they possess short sequence motif known as “DXD” (Asp- any residue -Asp).<sup>70</sup> They are expected to bear a carbohydrate binding site for sugar-nucleoside diphosphate and to be metal-ion dependent glycosyltransferases. Despite the fact that it can be seen and described as a single domain, GT-A fold topology consist on two abutting closely attached Rossmann-fold domains (Figure II-12). These domains are important elements for transfer reaction; they are responsible for recognition of the nucleotide and the binding of the acceptor.<sup>71</sup> The Rossmann-fold is a classical motif found in many binding nucleotides enzymes. It is composed of up to seven parallel beta sheets with  $\beta$ - $\alpha$ - $\beta$  supersecondary units. Indeed, the first three strands are associated with two  $\alpha$  helices, and this pattern is repeated once and inverted as it is showed in figure II-12.<sup>72</sup>



**Figure II- 12:** Schematic diagram of six stranded Rossmann fold.

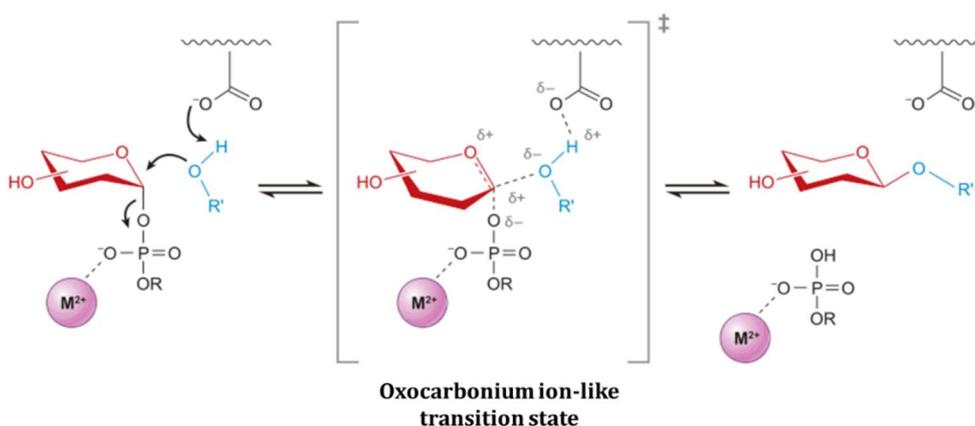
The second observed fold type is GT-B, characterized by two separated Rossmann domains that face each other. They are separated with a connecting linker region and a catalytic site located between the domains (Figure II-13).<sup>73</sup> GT-B glycosyltransferases do not have DXD motif and there is no evidence of a bound metal ion associated with catalysis. Recently, a third fold type GT-C was identified.<sup>74</sup>



**Figure II- 13:** Structures of (A) GT-A-fold of diphosphate sugar transferase SpsA from Bacillus subtilis (PDB ID 1QGQ) and (B) GT-B-fold bacteriophage T4  $\beta$ -glucosyltransferase (PDB ID 1JG7).

### c) Mechanisms

There are two possible mechanisms, which could be adopted by glycosyltransferases based on the inversion or retention of stereochemistry of anomeric carbon atom of the substrate donor.<sup>75</sup> Thanks to a  $S_N2$  reaction, the hydroxyl group of the acceptor, which is a good nucleophile, attacks the anomeric carbon from one side (from the top for example), and the nucleotide (UDP) is expelled from the other side (from the bottom), all in one-step. Here, UDP, the so-called leaving group, develops additional negative charge, and is electrostatically stabilized by the positive charge of metal ion bounded by the motif DXD. At the same time, side chain of aspartic acid or glutamic group of the catalytic site of the enzyme partially deprotonate the hydroxyl group of the acceptor. These reactions result in the formation of oxocarbenium ion-like transition state.<sup>75,76</sup> (Figure II-14)



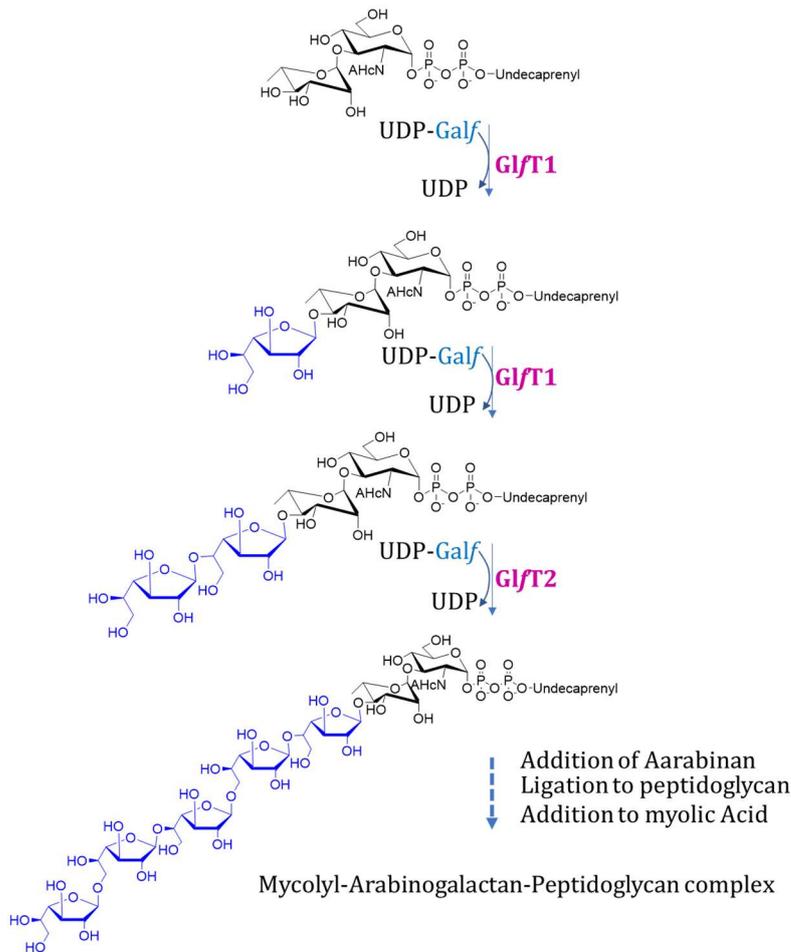
**Figure II- 14:** Schematic representation of inverting catalytic mechanism that utilizes a direct displacement  $S_N2$ -like reaction that results in an inverted anomeric configuration via a single oxocarbenium ion-like transition state. (From L.L. Lairson et al., 2008).

### III. Characterized galactofuranosyltransferases

Identification and characterization of functional galactofuranosyltransferases remain difficult. This is due to many factors. On one hand, obtaining high yield of pure proteins, which represents an essential step to characterize and determine their structures, is a difficult task.<sup>77</sup> This is due to their firm attachment to the cells membranes, in addition to their tendency to aggregate and loss of their activity.<sup>78</sup> On September 2018, there are 446370 entities for GTs that are registered on CAZy data base, however, only 2002 enzymes have been characterized and only 223 crystal structures were registered for 54 GT-families over 106 total GT families (<http://www.cazy.org/>). On the other hand, obtaining the donor UDP-Galf to evaluate the activity of GalfTs remains hard. Indeed, UDP-Galf is not commercially available, and its synthesis still complicated.

#### A. GlfT1 and GlfT2 from *Mycobacterium tuberculosis*

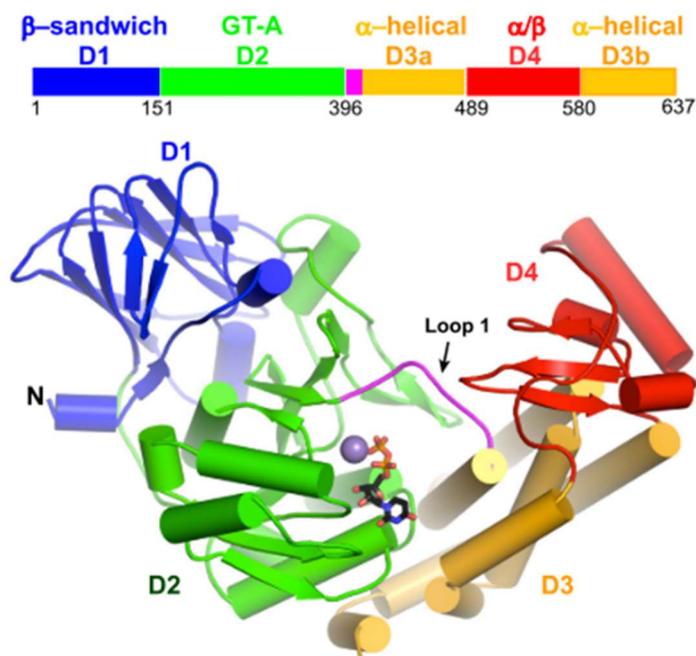
The analysis of the genome of *Mycobacterium tuberculosis* revealed the presence of two genes *Rv3809c* and *Rv3808c* that encode respectively for two galactofuranosyltransferases GlfT1 and GlfT2.<sup>79</sup> GlfT2 was the first to be discovered followed by GlfT1. Assays showed that both have bi-functional activities. Indeed, GlfT2 is able to catalyze the addition of Galf moiety from UDP-Galf donors using two synthetic acceptors;  $\beta$ -D-Galf-(1 $\rightarrow$ 5)- $\beta$ -D-Galf-O-C<sub>10:1</sub> and  $\beta$ -D-Galf-(1 $\rightarrow$ 6)- $\beta$ -D-Galf-O-C<sub>10:1</sub> resulting in two disaccharides;  $\beta$ -D-Galf-(1 $\rightarrow$ 6)- $\beta$ -D-Galf-(1 $\rightarrow$ 5)- $\beta$ -D-Galf-O-C<sub>10:1</sub>,  $\beta$ -D-Galf-(1 $\rightarrow$ 5)- $\beta$ -D-Galf-(1 $\rightarrow$ 6)- $\beta$ -D-Galf-O-C<sub>10:1</sub> and one tetra saccharide;  $\beta$ -D-Galf-(1 $\rightarrow$ 6)- $\beta$ -D-Galf-(1 $\rightarrow$ 5)- $\beta$ -D-Galf-(1 $\rightarrow$ 6)- $\beta$ -D-Galf-O-C<sub>10:1</sub>. GlfT2 is a polymerase responsible of the transfer of more than 20 Galf residues without interruption.<sup>68</sup> It is now well established that GlfT1 catalyze the two first transfers (1 $\rightarrow$ 4 and 1 $\rightarrow$ 5) Galf units on the  $\alpha$ -L-Rhap-(1 $\rightarrow$ 3)-Glc pNAc-phosphate lipidic acceptor.<sup>80</sup> GlfT1 and GlfT2 both belongs to GT2 family and are essential for the biosynthesis of mycobacterial mAGP. (Figure II-15).



**Figure II- 15:** Pathway for the biosynthesis of mycobacterial arabinogalactan. (Adapted from Carla Marino et al., 2016)<sup>50</sup>

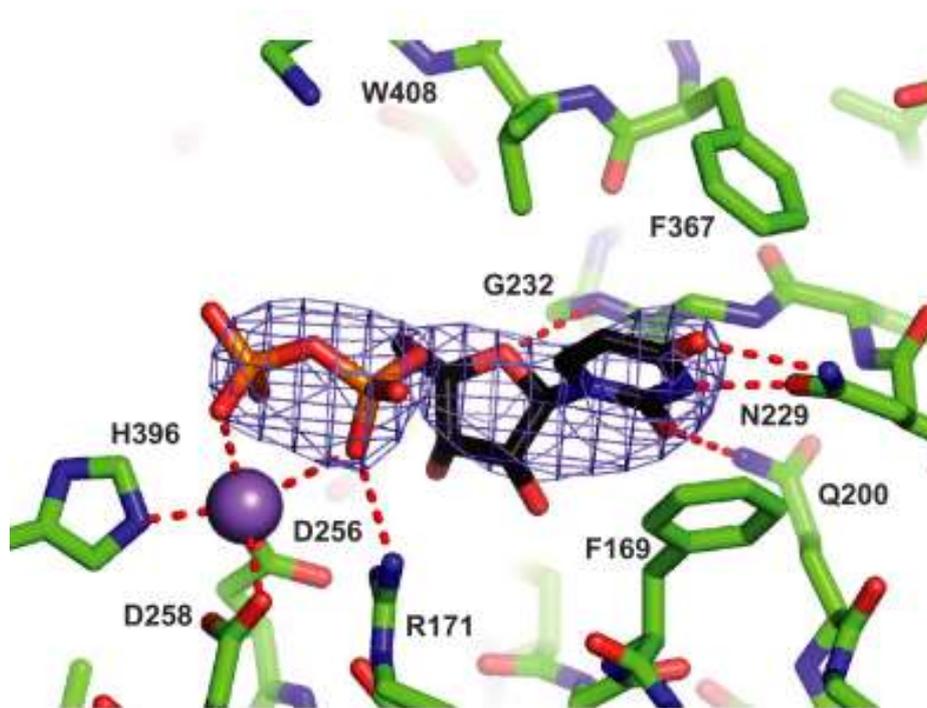
It must be noted that such a dual activity for glycosyltransferases is uncommon. The “one enzyme, one sugar linkage” was an old belief in glycobiology, that is why such bi-functionnal enzymes drawn a lot of attention and understanding their mechanism was very important and intriguing. It was thought that they probably have two active sites. Yet, an epitope mapping studies performed by the team of Lowary using STD NMR experiment showed that GlfT2 binds two competitive acceptors in different terminal linkages.<sup>81</sup> By this study, it was suggested then that GlfT2 has only one active site. Later, this hypothesis was confirmed by Kissling and co-workers.<sup>82</sup> Kinetic parameters are only available for GlfT2. Indeed, the parameter  $k_{cat}/K_m$ , that represents the substrate specificity of the enzymes was calculated with a value of  $\sim 1131 \text{ min}^{-1} \mu\text{M}^{-1}$  in the presence of UDP- $\alpha$ -D-Galf as donor and  $\beta$ -D-Galf-(1 $\rightarrow$ 5)-  $\beta$ -D-Galf-O-C10:1 as acceptor.<sup>83</sup>

The tetrameric structure of unliganded form of GlfT2 and the complexed form with the product UDP were also revealed.<sup>83</sup> Robert X. Weatley *et al.*, obtained similar structure of the two chains in both forms of GlfT2 (1.8 and 2.9 angström respectively). Each protein chain is composed of an *N*-terminal  $\beta$  sandwich, which precedes the central GT A family (GT A). Both regions are followed by an  $\alpha$  helical domain and *C*-terminal mixed  $\alpha$  and  $\beta$  domain (Figure II-16).<sup>83</sup>



**Figure II- 16:** Crystal structure of GlfT2 complexed with UDP. A ribbon diagram of the four domains, with  $Mn^{2+}$  presented by a grey sphere and UDP molecule presented in stick.<sup>83</sup>

The catalytic site and the binding site were also revealed, which are similar to other GT-A enzymes. The amino acid that interact with the metal ion  $Mn^{2+}$  to the active site were identified. As showed in Figure II-17, Asp-256 and Asp-258 from the DXD motif site and the His-396 present at the C-terminal end of the enzyme interact with  $Mn^{2+}$ . In the same way, Asn 229 and Gln-200 are responsible for donating and accepting hydrogen bonds with uracil base of the donor.

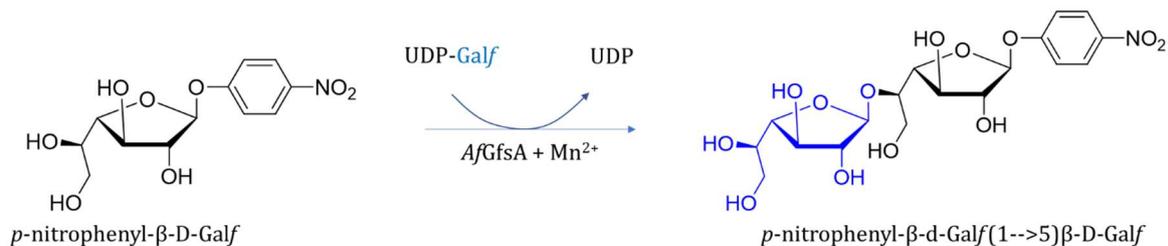


**Figure II- 17:** The structure of the catalytic site binding the UDP and  $Mn^{2+}$ .<sup>83</sup>

### B. GfsA from *Aspergillus*

The cell wall of fungi is covered by a variety of polysaccharides which are involved in several biological events.<sup>84</sup> Galactofuranose residues are frequently found in glycoproteins and in galactomannan of *Aspergillus*.<sup>85,86</sup>

The gene *gfsA* was recently identified, it encodes for a galactofuranosyltransferase, GfsA, which belongs to the GT31 family of the CAZy database. Mutant strains with *gfsA* gene revealed a sensitivity to antifungal treatments, thus, suggesting that it could be a therapeutic target to study.<sup>87</sup> To gain an in depth-understanding of the enzymatic functions of GfsA in *A. fumigatus* (*AfGfsA*), it has been cloned and expressed in *E. coli* and enzymatic assays to study its galactofuranosyltransferases activity were performed. UDP- $\alpha$ -D-galactofuranose was used as a sugar donor, and p-nitrophenyl- $\beta$ -D-galactofuranoside as an acceptor substrate (Figure II-18). Tests revealed that *AfGfsA* is responsible for the biosynthesis of  $\beta(1,5)$ -Galactofuranose in the galactofuran side chains of fungal-type galactomannan.<sup>88</sup> However, kinetic parameters and structural data are still not available.



**Figure II- 18:** Schematic *in-vitro* assay reaction of AfGfsA. The enzymatic transfer of D-Galf residue from UDP-Galf to the C-5 position of the Galf residue in the p-nitrophenyl-β-D-Galf via β-linkage.

### C. WbbI from *E. coli* K-12

Few information about *E. coli*- K12 GalT are available. *wbbI* gene was cloned and purified using *E. coli* BL21 (DE3). Preliminary results published by Corin Wing et al suggests that WbbI is a galactofuranosyltransferase and transfers Galf moiety on the acceptor *n*-octyl α-D-glucopyranoside from the donor UDP-Galf by a β-(1→6) linkage. Chemical structures were confirmed by HPLC. Further kinetics analysis are under ongoing investigations.<sup>89</sup>

### D. Summary

To sum up, these different GalTs are still poorly described enzymes despite the fact that they are crucial for the virulence and the pathogenicity of numerous microorganisms.

	Acceptor	* $k_{cat}/K_m$ ( $\text{min}^{-1}\text{mM}^{-1}$ )	Crystal structure
<b>GlfT2</b>	• β-D-Galf-(1→5)- β-D-Galf-O-C <sub>10:1</sub>	1131	Yes
	• β-D-Galf-(1→6)- β-D-Galf-O-C <sub>10:1</sub>	N.D	PDB: 4FIX
	• β-D-Galf-(1→6)- β-D-Galf	N.D	
	• Allyl 2,3,6-tri-O-benzoyl-5-hydroxy-β-D-galactofuranoside	N.D	
	• 12-d5-Phenoxy-dodec-2-enyl-β-D-galactofuranosyl-(1,6)-β-D-galactofuranoside	N.D	
	• 12-Phenoxy-dodec-2-enyl-β-D-galactofuranoside...	N.D	
<b>GlfT1</b>	α-L-Rhap-(1→3)-GlcPNAc	N.D	No
<b>GfsA</b>	p-nitrophenyl-β-D-galactofuranoside	N.D	No
<b>WbbI</b>	<i>n</i> -octyl α-D-glucopyranoside	N.D	No

**Table II- 1:** Summary of available data about the acceptors, kinetics parameters, and crystal structures of the four GalTs; GlfT2, GlfT1, GfsA, and WbbI.

As mentioned in the table II-1, their respective substrates of GlfT2, GlfT1, GfsA and WbbI are known, however, only GlfT2 kinetics were determined in the presence of UDP-Galf as donor and  $\beta$ -D-Galf-(1 $\rightarrow$ 5)- $\beta$ -D-Galf-O-C<sub>10:1</sub> as the acceptor. And, only structural data of GlfT2 are available. Moreover, GlfT2, GlfT1, GfsA and WbbI share less than 50% percent of amino acid sequences identity, which is considered as non- significant homology.

## IV. Galactofuranose-containing glycoconjugates of *Leishmania*: therapeutic target for leishmaniosis

### A. Leishmaniosis as a neglected inl disease

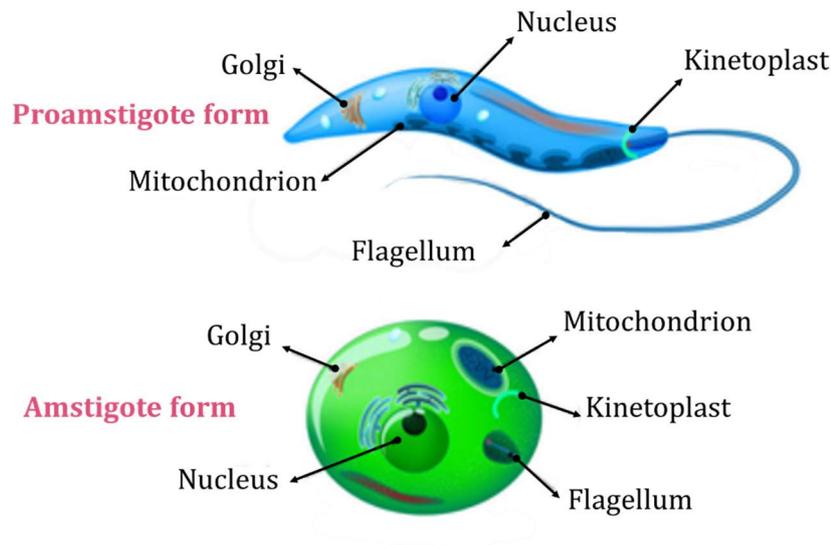
Leishmaniosis is a set of diseases classified as neglected tropical diseases (NTD) by the World Health Organization (WHO). It poses a real public health problem, globally, with 15 million people affected. It is one of the priority diseases of the WHO along with AIDS, malaria and tuberculosis.<sup>90</sup> The causal agent of leishmaniosis is a protozoan of the genus *Leishmania* and transmitted by a bite of a hematophagous arthropod mainly, *Phlebotome* and *Lutzomyia*. These parasitosis are characterized by a large spectrum of clinical manifestations that ranges from self-healing cutaneous lesions to fatal visceral infections. This depends on the species of the parasite and the immune system of the host.<sup>91,92</sup>

The occurrence of human leishmaniosis shows an important increase over the last decades due to various factors, including poverty, malnutrition, housing problems, hygiene, human migration, global warming and the emergence of drug-resistance in developing countries.<sup>93,94</sup>

#### 1. Causing agent, *Leishmania*

*Leishmania* are flagellated protozoa, belonging to the order Kinetoplastidae and the family Trypanosomatidae.<sup>95,96</sup> Four subgenera have been reported, *Leishmania*, *Viannia*, *Mundinia* and *Endotrypanum*. These parasites are transmitted to vertebrate hosts (humans, dogs, Lizard, Gerbils...) by a bite of an infected female sand-fly belonging to the genera of *Phlebotome* or *Lutzomyia*.<sup>97</sup> There are more than 20 different species of *Leishmania*, that can infect humans and causes a range of clinical manifestations; cutaneous (CL), mucocutaneous (ML) and visceral leishmaniosis (VL). It is important to highlight the fact that *Leishmania* infect animals leading to the creation of a zoonotic reservoir that can generate difficulties disease control.<sup>98,99,100</sup>

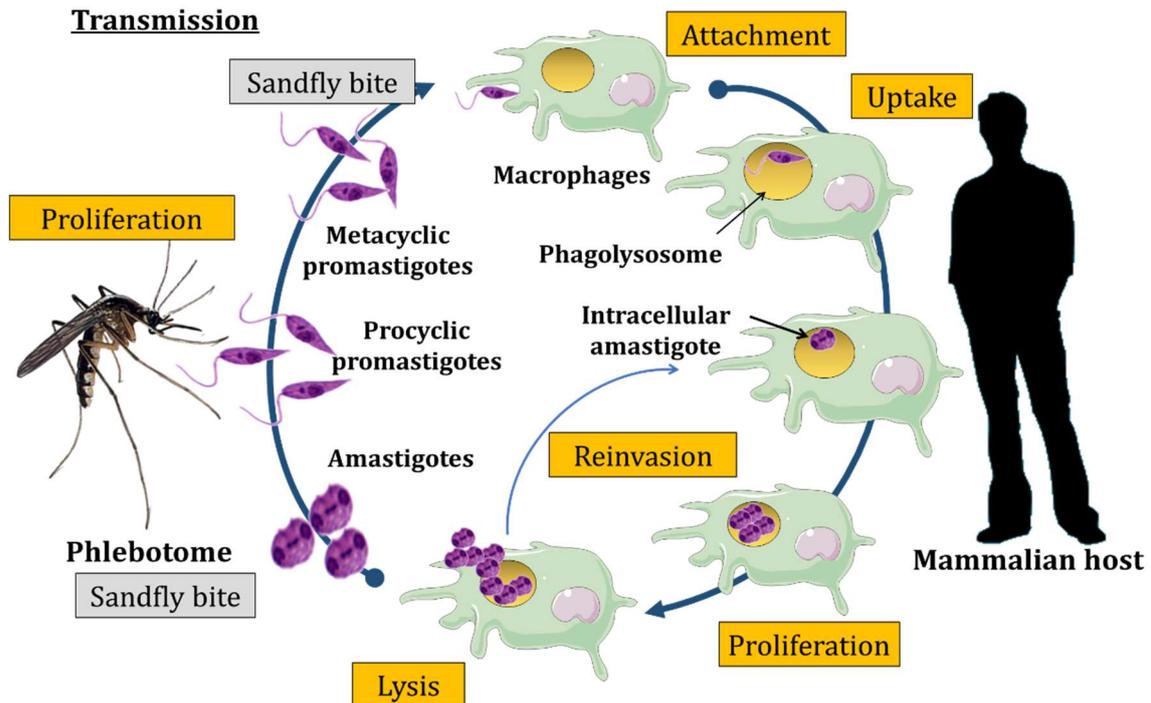
The parasites have two major body forms represented by the promastigote and the amastigote morphologies found respectively in the sand fly and the vertebrate hosts. The amastigote is a small spherical cell body of 2 to 6  $\mu$ m diameter, with a nucleus, a kinetoplast and an internal flagellum that is probably more focused on sensory functions. However, the promastigote is a long-bodied (15-25  $\mu$ m) and thin (2  $\mu$ m), with a central nucleus, a kinetoplast and a long free anterior flagellum that provides propulsive force.<sup>101</sup> (Figure II-19)



**Figure II- 19:** Promastigote and amastigote form of *Leishmania*, the causative agent of leishmaniasis.

## 2. Life cycle

Infected female sandflies deposit metacyclic promastigotes *Leishmania* into the skin of the vertebrate host during their blood meal. This leads to the recruitment of immune cells, macrophages and neutrophils cells (DC).<sup>102,103,104</sup> Since the life span of neutrophils is too short, macrophages are the final host cells. Metacyclic promastigotes are known as the mammalian infective stage and strictly non-dividing form. They are also highly motile cells, which explain the fact that parasite can migrate far from the bite site. Then, promastigotes differentiate into the non-motile amastigote form. These cells are smaller, hence the area over which the cells are exposed to the harsh environment. Amastigotes are able to divide by binary fission only if the macrophage is safe to ensure its survival. This causes the rupturing of the parasitophore vacuole membrane, and parasites continue to multiply. If the macrophages are dying, amastigotes are then released and infect other surrounding immune cells.<sup>105</sup> When sandfly feed on infected vertebrate, it ingests macrophages containing amastigotes. Few hours later, amastigotes differentiate into procyclic promastigotes under a combination of temperature, pH and the presence of stimuli. Procyclic promastigotes attach to epithelium of the sand fly to avoid being expelled during defecation. They then multiply by binary division and change into an infective metacyclic promastigotes. Further they escape and migrate towards the anterior end of the digestive tract. These parasites are transmitted to vertebrate host during the following blood meal (Figure II-20).<sup>106</sup>



**Figure II- 20:** The life cycle of leishmania parasites, whether in the sand-fly vector or the human host.

### 3. Clinical forms of leishmaniasis

Leishmaniasis has been a major public health problem in endemic areas. Four clinical forms of leishmaniasis have been identified, cutaneous leishmaniasis (CL), diffuse cutaneous leishmaniasis (DCL), mucocutaneous leishmaniasis (MCL) and visceral leishmaniasis (VL).

#### a) Cutaneous leishmaniasis

Weeks after the sand-fly bite, lesions and many dermal ulcerations appear at the site of bite mostly on the accessible and uncovered part of the body such as the face, arms and legs. At the beginning, a local increase of temperature and small papules appears that resemble to a bite of mosquitos. Then, it turns into a vesicle, then a pustule that grows and evolves to ulcerated forms. These lesions can last for months or years and cause handicap and disabilities. These ulcerations could heal naturally in 3–9 months in the case of *L. mexicana*, 2–6 months in the case of *L. major*, and 6–15 months if the agent is *L. braziliensis*, *L. tropica*, *L. panamensis*. However even with an adequate treatment, patient remains with indelible visible scars for the rest of their life. Also, prolonged contact with ulcerated area result in auto-inoculation and apparition of infected site distant from sand fly bite site.<sup>107,108,109</sup>

#### b) Diffuse cutaneous leishmaniasis

DCL is another type of cutaneous leishmaniasis caused by *L. Mexicana* complex. It is known as chronic leishmaniasis characterized by multiple non-ulcerative nodules spread from the site bite

to the rest of the body. DCL is difficult to treat because they are non-responsive to the common anti-leishmanial drugs and patient do not heal spontaneously.<sup>110,109</sup>

#### c) *Mucocutaneous leishmaniasis*

Mucocutaneous leishmaniasis or mucosal leishmaniasis involves mucous membranes in addition to the skin. It is often destructive form of the disease with mucosal inflammation and manifests days to years after cutaneous leishmaniasis. MCL is less responsive to treatment and is more disfiguring. In case of severe infections, this disease could cause death. MCL is typically a consequence of infection by New World Leishmania species such as *L. braziliensis*, *L. guyanensis*, *L. panamensis*, and *L. amazonensis*.<sup>111</sup>

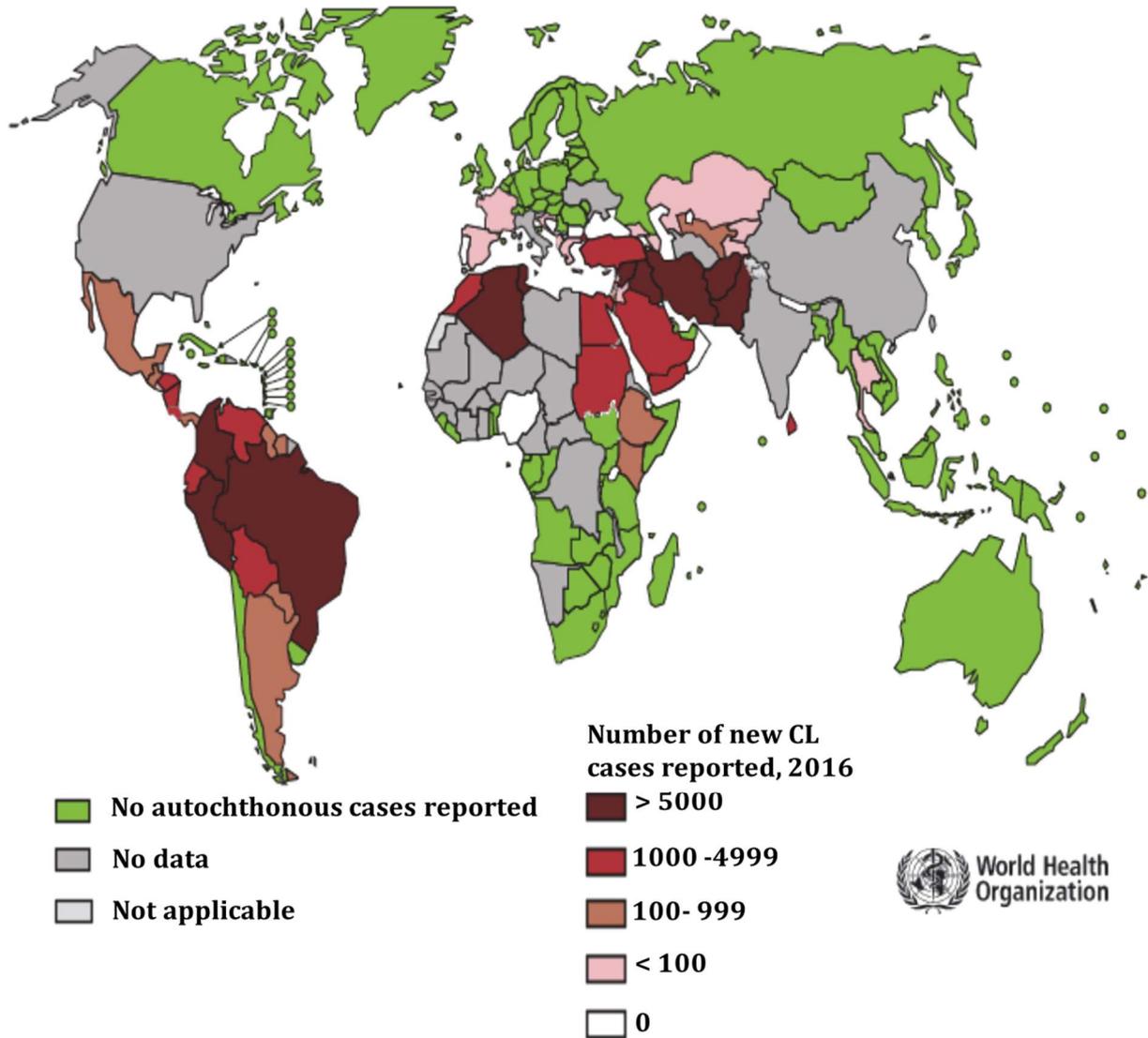
#### d) *Visceral leishmaniasis*

VL is the most severe and lethal clinical form. It is also known as kala-azar or black fever and affects mainly the poorest population. It is associated to malnutrition, poor housing and weak of immune system. The at-risk people are preschool children, immunocompromised and undernourished individuals and it is increased with patients who have AIDS. Patients are often asymptomatic for months to years before the activation of the infection. It is manifested by lymphadenopathy, hepatomegaly, splenomegaly, pallor, anemia, leukopenia, thrombocytopenia, fever, night sweats, weakness, anorexia, asthenia, cutaneous pigmentation, and weight loss, which can progress rapidly in weeks or months. In addition, affected children present characteristic chronic diarrhea and growth retardation. It is caused by *L. donovani*, *infantum*, *L. chagasi*, *L. amazonensis*, and *L. tropica*.<sup>112,91</sup>

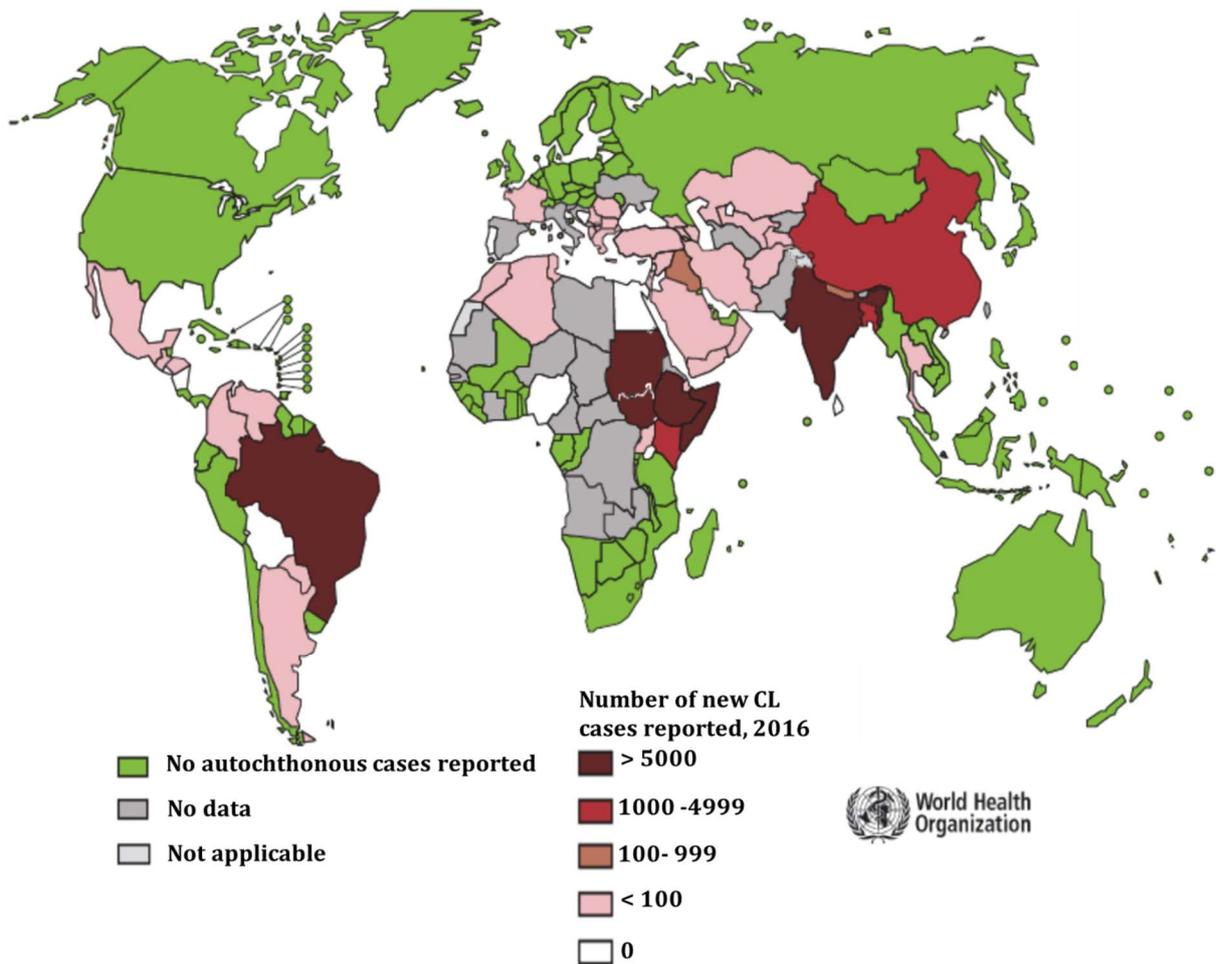
### 4. Epidemiology

Leishmaniasis is a widely distributed disease. It is found in 89 countries and 4 continents: Africa, North and South America, Asia and Europe (Figure II-21/22).<sup>113,114,98</sup> Globally, around 15 million people are infected in the world. Surveillance data estimate between 1.5 to 2 million new cases occur each year, 350 million are in risk of acquiring the disease and around 70 000 deaths. In 2014, WHO reported that 70% of visceral leishmaniosis cases are identified in Afghanistan, Algeria, Brazil, Columbia, Costa Rica, Ethiopia, Iran, Sudan and the Syrian Arab Republic. In addition, 90% of global cases of cutaneous leishmaniosis occurs in Bangladesh, Brazil, Ethiopia, Indi, South Sudan and Sudan. VL cases have been reported some countries in Europe mainly caused by *L. infantum* species. Most of registered cases are immunocompromised children and adult patents, particularly those suffering from the human immunodeficiency virus (HIV) but also transplanted solid organs or people receiving treatment immunosuppressant. Since the 1980s, LV is an emerging opportunistic disease in southwestern Europe (Portugal, Spain, France, Italy) where more than 1500 cases of HIV-Leishmania co-infections have been reported in addition to the asymptomatic carriage which are not counted.<sup>115</sup> In the south of France, the species

*Leishmania infantum* is endemic in Mediterranean climate regions. Its reservoir is the dog, and canine enzootic is important (about 30 000 annual cases). The main centers are those of the Pyrénées-Orientales, Cevennes, Provence, Côte d'Azur and Corsica. Human cases are limited in number (about 40 annual cases), mainly represented by LV, incidentally by cases of CL.<sup>114,116</sup>



**Figure II- 21:** Statute of endemic cutaneous leishmaniosis worldwide 2015 (Adapted from WHO).



**Figure II- 22:** Statute of endemic visceral leishmaniasis worldwide 2015 (Adapted from WHO).

## 5. Diagnosis and treatment

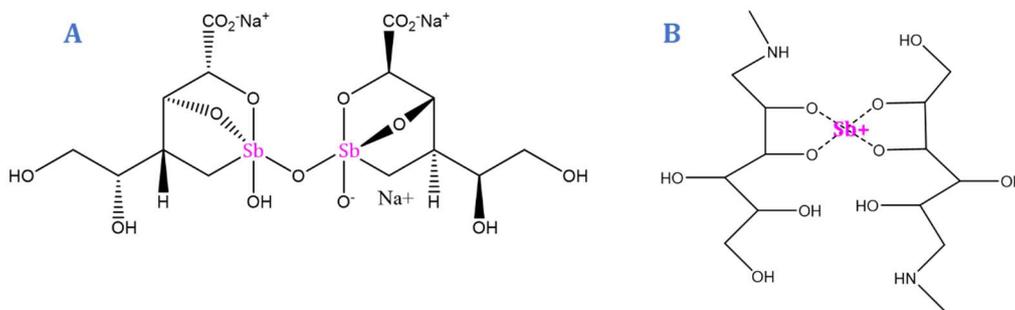
### a) Diagnosis

Different diagnosis methods and tools can be used according to the type of clinical form of the diseases. For example, *Leishmania* parasite DNA can be detected in skin lesion samples for cutaneous leishmaniasis, and bone marrow, liver biopsies or blood samples for visceral leishmaniasis. Polymerase chain reaction (PCR) is used, it targets ribosomal RNAs coding genes or kinetoplastic DNA genes.<sup>91,117</sup> This could also help identifying the species of parasite responsible of the infection which is very important.<sup>118</sup> Another method can also be used, it consists of the detection of circulating antibodies. However, it is not sufficient and cannot be very reliable.<sup>119,120,121</sup> Indeed, some patient who has only lesions skin could show negative serology results, but it does not mean that *Leishmania* infection is totally excluded. Moreover, immunocompromised patient will express a high titer of antibodies even after anti-leishmanial treatment. This is why serology test should be combined to other culture or molecular techniques in order to eliminate any cases of doubt.

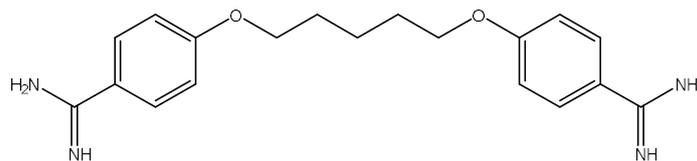
### b) Treatments

Different therapeutic approaches have been used to fight leishmaniasis. They depend on the species of the parasite and the medical state of patients (immunocompromised, malnourished...). The suitable treatment should be effective against the amastigote form of the parasite which infects mammalian hosts. Also, depending on the clinical form of leishmaniasis, it should be able to target either the parasite that causes skin lesions or visceral damages. Thus, the diffusion and degradation parameters of drugs should be well adapted. Up to date, various molecules are used to treat different leishmaniasis cases by acting at different levels.

On one hand, there are those which target the transduction machinery of parasites. As an example, pentavalent antimony which has been considered for six decades as the best effective treatment for all clinical forms of leishmaniasis (Figure II-23).<sup>122</sup> They act by selective inhibition of many parasite enzymes by inhibiting ATP synthesis and lead to the activation of endonucleases which are responsible for the defragmentation of parasite DNA.<sup>123</sup> However, in the last years, resistant *Leishmania* strains to this drug emerged in addition to the registration of some toxic side effects that could be fatal.<sup>123</sup> Pentamidine is another transduction inhibitor, which is commercialized under the name of pentacarinat® (Figure II-24). It works by inhibiting parasite DNA synthesis by blocking thymidine synthetase, binding RNA transfer, and disrupting mitochondrial activity. But because of the cost of this treatment and the fact that it increases the risk of developing insulin-dependent diabetes in addition to its poor effectiveness, its use has been limited.



**Figure II-23:** Chemical formula of two pentavalent antimonial compounds; (A) sodium stibogluconate and (B) meglumine antimoniate

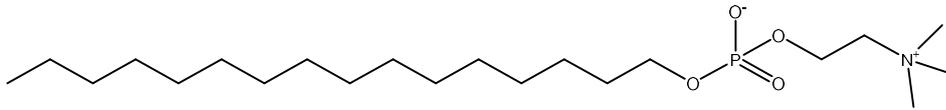


**Figure II- 24:** Chemical formula of pentamidine.



of miltefosine, ~100 euros for 10 mL, made the medicine exorbitant for the majority of patients.

131



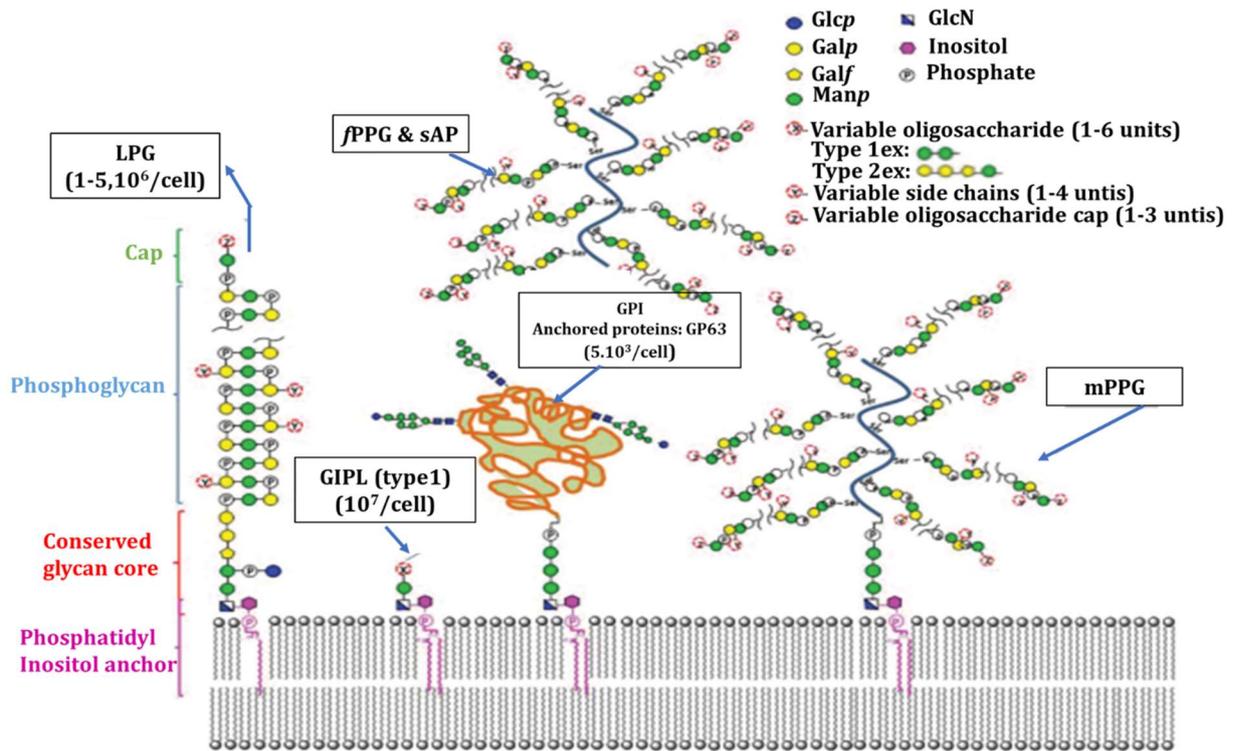
**Figure II- 26:** Chemical formula of miltefosine.

Some other molecules are currently used but are still facing the same problems, mainly the emergence of drug resistance which is one of the greatest existing threats to human health.<sup>132,133</sup> Also, these treatments are too expensive and not everyone can afford it, and we know that leishmaniasis affects the poorest population in the world and treatments are not easily found in tropical areas. Several trials to develop a vaccine for prevention of leishmaniasis were carried out but none have been successful at the moment, except for canine leishmaniasis. Indeed, this vaccine is based on *Leishmania* surface protein GP63 fragments. Another means of prevention is also used. It is directed against vectors, mainly it consists of the use of insecticide and mosquito nets. However, insecticides can have a harmful effect on human health. In addition, the phlebotomes are very small which allows them to pass through the meshes of the mosquito nets.

There is then a desperate need to find new therapeutic approaches. There is one old but efficient strategy consisting in targeting the cell wall biosynthesis which remains powerful. Today, we know the importance of glycoconjugates expressed in the surface of pathogens, so targeting the enzymes which are involved in their biosynthesis is an interesting strategy.

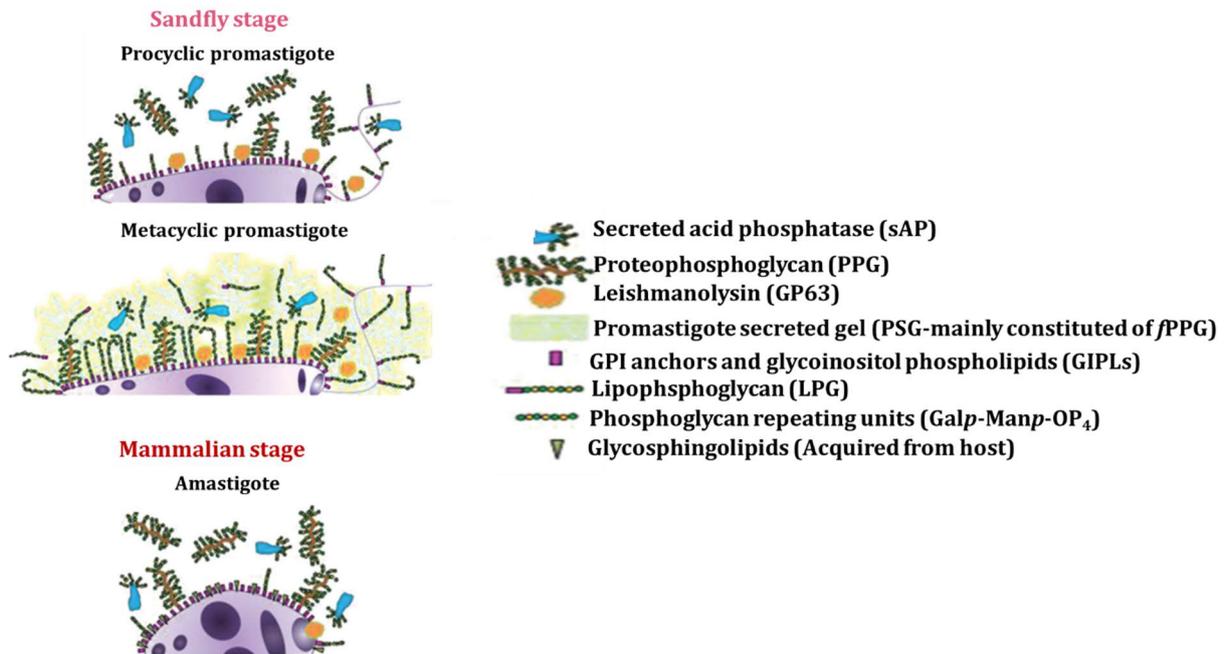
## B. Cell wall glycoconjugates as therapeutic targets

*Leishmania* pass through different stages during their life cycle, which are characterized by modifications of their morphology and the structure of their glycocalyx. These parasite exhibit on their surface a group of complex glycoconjugates which play a main role in their survival. These glycoconjugates are composed of oligosaccharides anchored to the cell membrane by phosphatidylinositol (PI). Different types of these glycoconjugates have been identified and showed in Figure II-27. Among these glycoconjugates, there are glycoinositol phospholipids (GIPLs), which are small structures composed of few glycosidic units.<sup>134</sup> In addition, lipophosphoglycans (LPGs) structures have been identified.<sup>135</sup> They are composed of phosphorylated polysaccharides. There are also, glycosylated proteins called proteophosphoglycans (PPGs) with similar patterns identified in LPG structures.<sup>136</sup>



**Figure II- 27:** Different glycoconjugates structures of *Leishmania*. (From Yari Cabezas *et al* ., 2015).<sup>135</sup>

It is important to highlight that the size of glycoconjugates that coat the glycocalyx display different size according the parasite species and the stage of development.<sup>137</sup> Indeed, the coat can reach 6nm for the promastigote forms and until 12 nm for the metacyclic forms.<sup>138</sup> However, it decreases during the amastigote stage.<sup>139</sup> (Figure II-28)

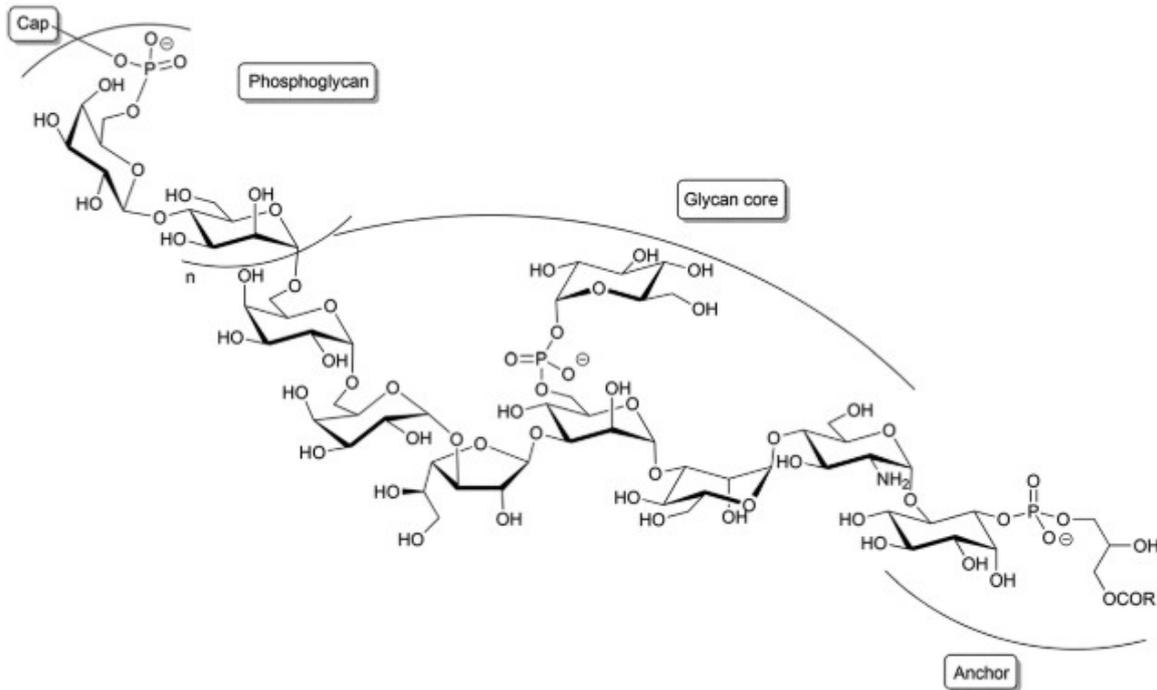


**Figure II- 28:** Evolution of glycoconjugates during different life cycle stage of *L. donovani* parasites.<sup>135</sup>

1. Lipophosphoglycans (LPGs)

a) Structure

Lipophosphoglycan of *Leishmania* is the major component of the promastigote stage and it covers the whole body even the flagellum. It is composed of four domains; a GPI anchor, a glycan core, a linear phosphoglycan chain (PG) and a terminating oligosaccharide cap. The first three domains are identical in all *Leishmania* species. However, the last one, the oligosaccharide cap and the side substituents of the linear PG display a heterogeneity depending of the species of the parasite and their life cycle stages. The GPI anchor site consists of phospholipid derivative 1-*O*-alkyl-2-lyso-phosphatidylinositol with a single saturated C<sub>24-26</sub> aliphatic chain.<sup>42</sup> The glycan core is an heptasaccharide with a non-glycosylated glucosamine linked to inositol and two mannosides followed by a galactofuranosides and two galactopyranosides moieties. The linear conserved phosphoglycan is formed by repetitive disaccharide Gal4-β(1,4)Manα1-PO<sub>4</sub>). The number of repeating units in the PG chain allows a variation of the thickness of the glycocalyx according to the parasite stage. Indeed, the glycocalyx is thicker for metacyclic form than procyclic form.<sup>140,141</sup> The oligosaccharide cap consists on galactopyranose and mannose assembled as Manα(1-->2)Man for *L. major* promastigotes or as Galβ(1-->4)Manα(1-->2)Man for *L. donovani* and *L. mexicana* promastigotes<sup>142,143,144</sup> (Figure II-29).



**Figure II- 29:** Structure of the four domains of Lipophosphoglycan; GPI anchor, a glycan core, linear phosphoglycan chain (PG) and terminating oligosaccharide cap (Yari Cabezas *et al.* , 2015).<sup>135</sup>

Three types of LPG have been identified according to the nature of side chain of the linear phosphoglycan. The first type of LPG is expressed on *L. donovani* characterized by the absence of substituent. LPG type 2 belongs to *L. Mexicana*, *L. tropica* and *L. major*, it contains branched galactopyranose. The last type of LPG is exclusively found in *L. athiopica* and is 35% mannosylated in the position 2 of mannose.<sup>145</sup>

### b) Functions

Lipophosphoglycans are involved in several process during the life cycle of parasite either in the vector or in the mammalian hosts.

#### (1) In vertebrate host

It is important to highlight the fact that each species of *Leishmania* is transmitted by specific species of sand-fly. This is partly due to the composition of LPGs. The carbohydrates that compose these glycoconjugates act as ligand for specific lectins expressed in the epithelium of vectors.<sup>146,147</sup> In this case there is no suitable receptor of the carbohydrate part of LPG in the stomach epithelium, the binding of the parasite cannot take place.<sup>148</sup> So, this interaction allows parasite to cling and avoid their elimination. These same LPGs play a major role in the detachment of *Leishmania*, allowing their migration to salivary glands of the sand-fly.<sup>149,150</sup> This is related to the metacyclogenesis which causes a thickening of the glycocalyx and thus prevents their recognition by lectins.

## (2) In mammalian host

During the infection of the mammalian host, *Leishmania* use its glycoconjugates to encounter the immune cells and ensuring its survival, some examples can be cited:

First, the presence of promastigotes in the dermis leads to the activation of the classical complement pathway, which is a part of the immune system.<sup>147</sup> It is based on a cascade of cleavage reactions of specific proteins leading to the final product, a protein named C5b-C9. This protein generates pores on the cell membrane of parasite inducing its lysing and facilitating its phagocytosis by macrophages. In this case, LPGs act as a shield which protect the parasite by preventing the attachment of the complex C5b-C9 on the cell membrane.<sup>151</sup>

Second, the macrophages activate nitric oxide synthases, enzymes that catalyze the synthesis of nitrogen oxide (NO). This endows macrophages with cytostatic or cytotoxic activity parasites. It was shown that the linear conserved repetitive units of phosphoglycan of *L. major* interfere with the expression of NOS (Nitrogen oxide synthase) mRNA and the decreasing of production of NO.<sup>141</sup>

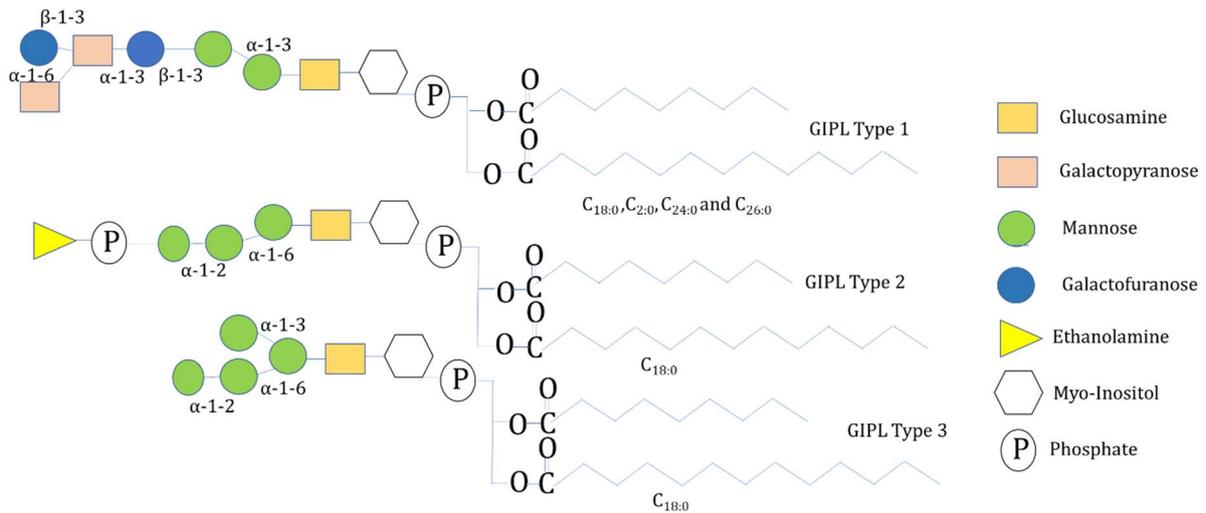
Third, after their uptake by macrophages, *Leishmania* parasite are conducted into phagosomes that differentiate into phagolysosomes. This will lead to the destruction of the pathogens by a combination of oxidative-processes such as the acidification of pH. Leishmanian promastigote are able to slow down the maturation of phagolysosome thanks to their LPGs ensuring then their differentiation into amastigotes before their degradation.<sup>152</sup> Thus, they can resist to harsh environment. Some experimental assays were made with lipophosphoglycan-free *Leishmania donovani* showed a normal maturation of phagolysosome. However, macrophages infected with wild type *L. donovani* were characterized by a retardation of phagolysosome formation.

## 2 Glycosylinositol phospholipids (GIPLs)

### c) Structure

GIPLs are small glycolipids expressed on the surface of both *Leishmania* forms but they are more abundant in the amastigote form. GIPLs are widely polymorphic, but with a basic conserved structure of  $\text{Man}\alpha(1-4)\text{GlcN}$  linked to the lipidic portion. Three types of glycosylinositol phospholipids were identified according to the structure of the glycan part.

The first types contain  $\text{Man}\alpha(1\rightarrow6)\text{Man}\alpha(1\rightarrow4)\text{GlcN}\alpha(1\rightarrow6)\text{-PI}$ . The second types distinguish by their  $\text{Man}\alpha(1\rightarrow3)\text{Man}\alpha(1\rightarrow4)\text{GlcN}\alpha(1\rightarrow6)\text{-PI}$  glycan structure which is similar to LPG anchor. Finally, the third are characterized by a branched structure part  $\text{Man}\alpha 1,6(\text{Man}\alpha 1\rightarrow3)\text{Man}\alpha(1\rightarrow4)\text{GlcN}\alpha(1\rightarrow6)\text{-PI}$ <sup>141</sup> (Figure II-30).



**Figure II- 30:** Schematic representation of the three types of GPIs identified in *Leishmania* adapted from De Asis *et al.*,2012.<sup>141</sup>

d) *Functions*

LPG is down-regulated in amastigotes, whose surface is dominated instead by the abundant GPIs, which suggested that these play critical roles in amastigote survival and virulence. In contrast to LPG, there are no available studies involving GPIs and sand fly interactions and this is an open field to be explored. The role of GPIs is not very clear. Recent results demonstrated that GPIs are recognized by Toll-Like Receptor 4 and this alter the resistance to the infection with *T. cruzi*.<sup>153</sup> Moreover, it was reported that GPIs of *L. major* inhibit the synthesis of NO which has been shown to play a causal role in the killing of the protozoan parasite.<sup>153</sup>

3. *Biosynthesis of glycoconjugates*

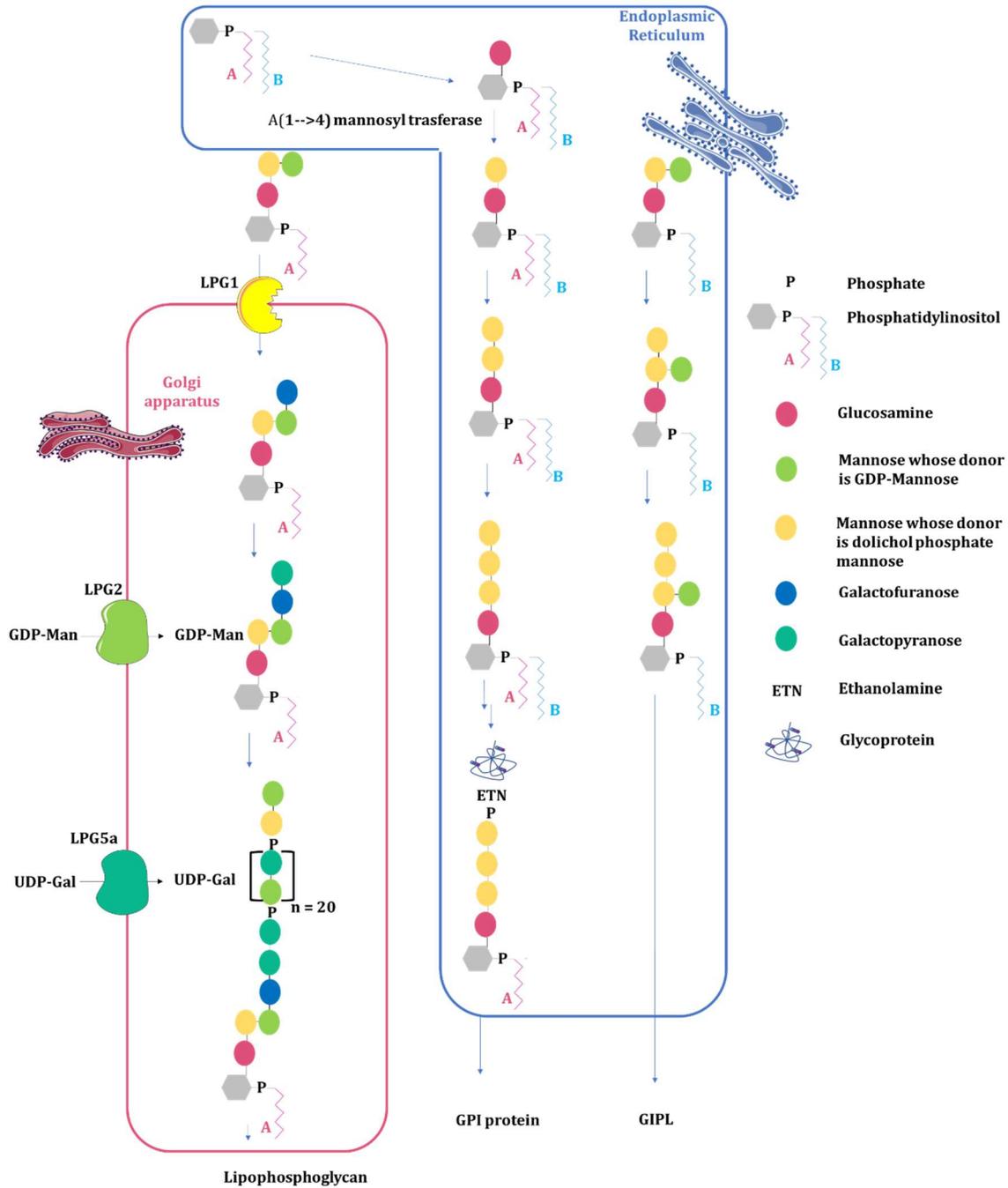
The biosynthesis of lipophosphoglycans is based on the consecutive addition of monosaccharides to the phosphatidylinositol that occur in the endoplasmic reticulum and the Golgi apparatus (Figure II-31). The addition of PI molecules is carried out by different enzymes, which are located in different subcompartments of the ER. This could explain the difference of the PI molecules chain length used for the synthesis of protein GPI (C24:0/ C26:0) anchors and GPIs (C18:0).<sup>148</sup>

*Leishmania* GPI are similar to those expressed by *Trypanosoma brucei*, so it is supposed that their biosynthesis should be similar also. It was then suggested that the synthesis of carbohydrate part of LPG, GIPL and GPI anchor of *Leishmania* starts from a common precursor, Man $\alpha$ (1 $\rightarrow$ 4)GlcN-phosphatidylinositol.<sup>148</sup> Furguson *et al.*, studied the enzyme,  $\alpha$ (1 $\rightarrow$ 4) mannosyl transferase, that is responsible of the addition of the first mannose residue onto the glycan part of *T. brucei*. They determined its donor, the dolichol phosphate-mannose. But they had to stop further experiments because of the difficulties of producing this GT. Indeed, the obtained recombinant proteins were not soluble.<sup>148,154</sup> The donor dolichol phosphate-mannose is synthesized by a dolichol phosphate-mannose synthase (DPMS) using GDP-Man and dolichol phosphate.<sup>154</sup> Thus, trials to knock out the

gene *lmdpms* coding for DPMS in *Leishmania Mexicana* were carried out. They resulted in the death of parasite cells. This validated the importance of GPI-containing molecules for the survival of *Leishmania*.<sup>154</sup>

The carbohydrate part of the GPI anchor and LPG contains branched mannoses.<sup>155</sup> The third residue of LPG core is galactofuranose, which is transferred by a specific type II transmembrane enzyme, galactofuranosyltransferases. One gene, *lpg1*, has been identified and supposed to encode for a putative Gal $\beta$ T. Attempts to knock out *lpg1* gene resulted the absence of LPG biosynthesis.<sup>136,156</sup>

The LPGs are characterized by the presence of repetitive unit regions that are assembled mostly in the Golgi apparatus. GDP-Mannose and UDP-Galactospyranose transporters as well as three mannosyl-phosphate transferases, and  $\beta(1\rightarrow4)$  galactosyl transferase were identified as essential for the assembly of this region.<sup>134</sup> Mannosyl-phosphate transferase that transfer Man $\alpha(1\rightarrow)PO_4$  to the Gal $\alpha1,6Gal$  was isolated from *L. donovani*, however, it was poorly studied.<sup>157</sup> Here, the mannose and galactose residues come from GDP-Man and UDP-Galp donors, which are transferred into the Golgi apparatus thanks to membrane transporters. *lpg2* was identified as leishmanial gene that encodes for multi-protein transporter. It was demonstrated that it is able to transfer GDP-Mannose, GDP-Arabinose and GDP-Fucose. Mutant *L major* with deletion of *lpg2* gene are lacking carbohydrate chain and LPG.<sup>158</sup> Moreover, *lpg5a* and *lpg5b* genes were also identified as encoding for UDP-Galp transporters. Knock out of *lpg5a* genes in *L. major* strains caused the inhibition of LPG biosynthesis, however, the deletion of *lpg5b* gene showed resulted modifications in the proteoglycan carbohydrate chains biosynthesis.<sup>134,159</sup>



**Figure II-31:** Biosynthesis of *Leishmania* phosphoglycans (Adapted from Shemyakin and Ovchinnikov, 2009). GPI anchors and phospholipids are synthesized by consecutive joining of monosaccharides to phosphatidylinositol with long C24:0/C26:0 (A) or short C18:0 (B) alkyl chains. The carbohydrates part of all types of phospholipids contain similar regions. The GPI proteins and GIPLs are synthesized in the endoplasmic reticulum, while LPGs are synthesized in the Golgi apparatus.

## V. Objective of thesis

Nowadays, leishmaniasis is classified as a major health problem for humans diseases.<sup>91,94,98</sup> The causative agent, *Leishmania*, has the remarkable ability to avoid destruction in hostile environments it encounters throughout its life cycle. As mentioned above, current treatments are expensive, have low efficiency and can be toxic. Also, resistant strains to the existing drugs have emerged and developing human vaccine is not imminent.<sup>160</sup> In spite of new and ongoing efforts, the desperate need for anti-leishmanial with novel mechanisms of action remains. It is then urgent to find new therapeutic targets.

*Leishmania* parasite can survive and then proliferate thanks to glycoconjugates expressed on its surface such as LPGs and GIPLs.<sup>16,161</sup> The biosynthesis of these glycoconjugates has been studied, and some relevant genes coding for nucleotide sugar transporters and transferases such as mannosyl-phosphate transferases and  $\beta(1\rightarrow4)$  galactosyl transferase have been identified<sup>134</sup>. However, other essential enzymes remain uncharacterized.

These polysaccharides contain in their structures rare sugar; galactofuranose, which is found in many pathogenic species but absent from mammalian cells. Then GalF-biosynthetic pathways have raised much interest as targets for drug development to fight parasites and microbial infections.<sup>16,22</sup> Targeting the biosynthesis of cell wall galactofuranose-containing glycoconjugates of leishmania has been already envisioned by scientists. Indeed, UDP-galactopyranose mutase of *Leishmania major*, enzyme that catalyze the transformation of UDP-galactopyranose to UDP-galactofuranose which acts as the precursor for  $\beta$ -GalF synthesis, has already been isolated and characterized.<sup>162</sup> Some compounds have been synthesized and tested as inhibitors for these enzymes. This represents some promising results.

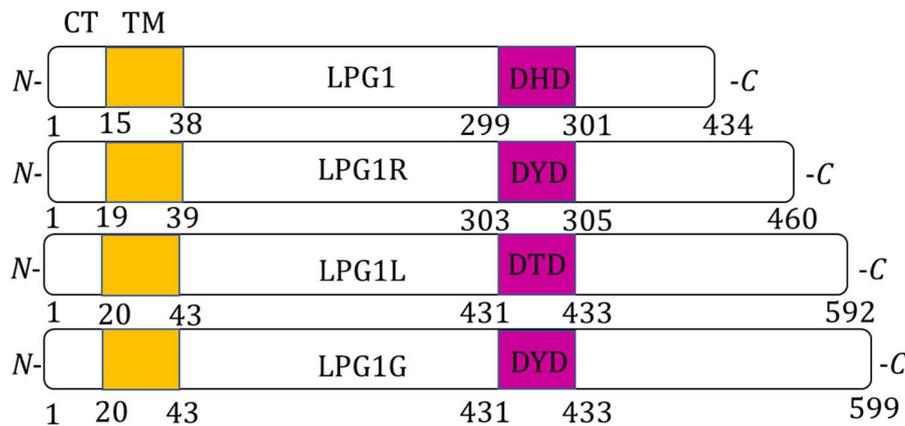
There are other important enzymes, which are involved in the biosynthesis of galactofuranose-containing polysaccharides, the galactofuranosyltransferases. They ensure the transfer of GalF moiety from an activated donor, UDP-GalF to an acceptor. This represents a key step for the biosynthesis of GalF-containing conjugates such as LPGs and GIPLs. This class of enzymes is poorly known. Few have been identified and only mycobacterial GalFT has been fully characterized.

<sup>163,81,87,83</sup>

In 1993, Salvatore J. Turco and co-workers identified for the first time *lpg1* gene (LmjF.25.0010). When they knocked it out, preliminary results showed that *Leishmania* mutant were deficient in the addition of galactofuranose addition in the glycan core of LPGs structures.<sup>156</sup> They suggested that this gene could be involved in the synthesis of the donor, UDP-GalF, the transfer of GalF moieties or other regulatory steps. Analysis of amino acid sequence of LPG1 favored the hypothesis of potential transferase. Indeed, LPG1 contain small region that reveals similarity with

mannose-binding proteins. This could be essential for binding the precursor of LPG carrying some mannosyl residues. moreover, putative signal anchor was revealed thanks to hydrophilicity analysis predicting specific protein topology that consists of a short cytoplasmic tail in the N-terminal of the protein, membrane anchor sequence and a long luminal domain containing four N-linked glycosylation sites. This arrangement is widely described for many glycosyltransferases. Consequently, it was presumed that *lpg1* encodes for a putative galactofuranosyl transferase, which is responsible for the addition of galactofuranose residues in the Endoplasmic Reticulum of *Leishmania*.<sup>156</sup>

Lately, Stephen M. Beverley and co-workers identified three related gene to *lpg1* in *Leishmania major*; *lpg1L* (LmjF.26.0550), *lpg1R* (LmjF.33.0300) and *lpg1G* that exists in identical three copies (LmjF.32.3990).<sup>164</sup> The predicted open frames for *lpg1*, *lpg1L*, *lpg1R* and *lpg1G* were predicted to encode for type-II-transmembrane proteins with short cytoplasmic tails (~20 amino acids) at the N-terminal part and a single transmembrane domain. Moreover, the four putative proteins seem to contain metal-binding DXD catalytic motif. (Figure II-32). This was correlated with Salvatore J. Turco and co-workers' observations.



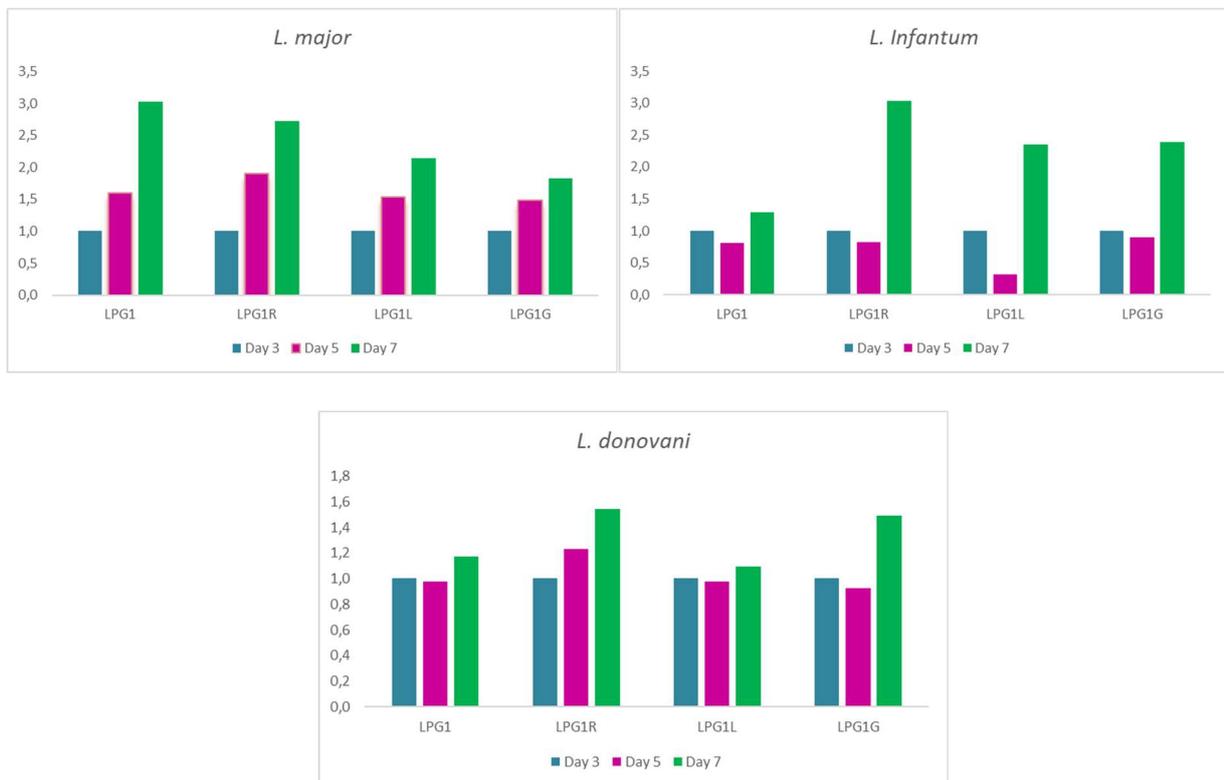
**Figure II- 32:** Representation of LPG1, LPG1R, LPG1L and LPG1G proteins adapted from Zhang et al.,2004.<sup>164</sup>. The numbers indicate amino acid positions. DXD conserved domain are in purple. CT refers to the cytoplasmic tails and TM for Transmembrane Domain.

Cellular localization experiments were carried out using green fluorescent protein (GFP) that was joined to the C-termini of LPG1 and LPG1R. They were detected in the Golgi apparatus, where most of glycosyltransferases are. These four proteins have never been expressed or purified before, only knock out gene experiment were done. Different *Leishmania* strains were generated in order to study the effect of LPG1x family on the biosynthesis of glycoconjugates LPGs and GIPLs. Mutant with a single knockout of the gene *lpg1L*- or *lpg1R*- displayed a similar glycans than the wild type (WT). Same results were obtained for double genes knockouts *lpg1L*-/*lpg1R*-. However, for triple knockout mutants *lpg1L*-/*lpg1R*-/*lpg1*-, only LPG was affected as much for single mutant *lpg1*-.<sup>165</sup>

This suggested that LPG1, LPG1R and LPG1L are not required for LPGs nor GIPLs biosynthesis. Only *lpg1* gene is essential for the synthesis of LPGs.<sup>165,164</sup> Up to date, there is no data on LPG1G activity nor expression.

Due to the importance of galactofuranose-containing glycoconjugates in the interaction of the parasite with its host, these putative galactofuranosyltransferases can become potential therapeutic targets for leishmaniosis disease.

Relative quantification of the expression of *lpg1x* genes of three different *Leishmania* species; *Leishmania major*, *Leishmania donovani* and *Leishmania infantum* were previously performed in the group by a master student Thibault RATTIER in collaboration with Pr. Florence Robert-Gangneux group. The expression of the genes was evaluated at different periods of the culture; after three, five and seven days. In addition, it was normalized with an invariant endogenous control called also reference gene coding for  $\beta$ -actin. This preliminary RT-qPCR experiments (Figure II-33) revealed that although the function of the four putative genes remain putative, their respective mRNA is expressed by the three parasites species. Moreover, the level of expression of each gene increases along with the parasite growth.



**Figure II- 33:** Evaluation of the expression of the four genes; *lpg1*, *lpg1R*, *lpg1L* and *lpg1G* over the time by *L. major*, *L. donovani* and *L. infantum*.

Thus, the objective of this thesis is to express and characterize these four putative galactofuranosyltransferases (Figure II- 34). The purpose is to discover whether or not these four

genes encode for galactofuranosyltransferases. The aim is to identify their substrates and their kinetics. Both prokaryotic and eukaryotic hosts were used to produce glycosylated and non-glycosylated proteins, in order to evaluate the effect of the glycosylation on the solubility and the activity of leishmanian LPG1x enzymes.

The first chapter will present the cloning, overexpression and purification steps to obtain pure and soluble recombinant LPG1x proteins. In this part, *E. coli* was used to produce non-glycosylated proteins. *E. coli* are well-known bacteria and widely by the scientific community to produce recombinant proteins, however, they lack the post-translational mechanism, like protein glycosylation.<sup>166</sup> In addition, crystallogensis trials to get protein crystals, an important step to determine the structure of LGP1x enzymes, will be described.

The second chapter presents the first enzymatic characterization of the four prokaryotic recombinant LPG1x. Herein, kinetic parameters of LPG1, LPG1L LPG1R and LPG1G were determined, presenting unexpected enzymatic activities and specificities.

The third chapter describes the production of galactofuranosyltransferases using *Leishmania* Expression System (LEXSy).<sup>167</sup> Since LPG1 is located in the membrane of Golgi apparatus of *Leishmania*, we decided to use *Leishmania tarentolae*, a non-pathogenic parasite for human, to produce glycosylated proteins. Indeed, as an eukaryotic host, it has the appropriate post-translational mechanism to ensure glycosylation and proper folding of our proteins of interest.<sup>168</sup> Moreover, preliminary assays to evaluate the activity of these proteins were performed and kinetic parameters were calculated.

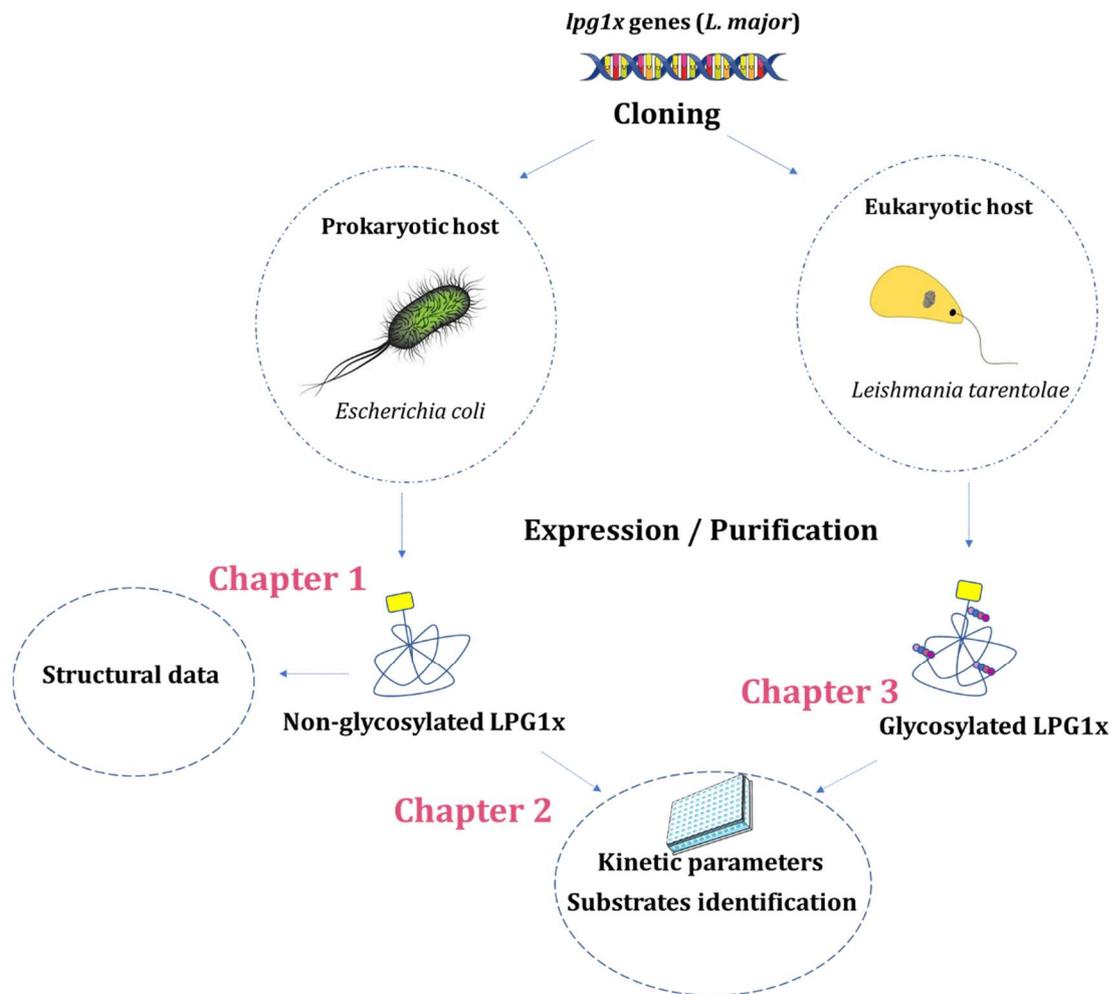


Figure II- 34: Objectives outline of this thesis manuscript.

## References

1. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495 (2014).
2. Jéquier, E. Carbohydrates as a source of energy. *Am. J. Clin. Nutr.* **59**, 682–685 (2018).
3. Watkins, W. The ABO blood group system: historical background. *Transfus. Med.* **11**, 243–265 (2001).
4. Hult, A. K., Dykes, J. H., Storry, J. R. & Olsson, M. L. A and B antigen levels acquired by group O donor-derived erythrocytes following ABO-non-identical transfusion or minor ABO-incompatible haematopoietic stem cell transplantation. *Transfus. Med.* **27**, 181–191 (2017).
5. Galan, M. C., Benito-Alifonso, D. & Watt, G. M. Carbohydrate chemistry in drug discovery. *Org. Biomol. Chem.* **9**, 3598 (2011).
6. Hospital, G., Street, T., Gard, T., Hoge, E. A. & Kerr, C. Glycan Engineering for Cell and Developmental Biology. *Cell Chem. Biol.* **23**, 108–121 (2016).
7. Unione, L. *et al.* Glycans in Infectious Diseases. A Molecular Recognition Perspective. *Curr. Med. Chem.* **24**, 4057–4080 (2017).
8. Solá, R. J. & Griebenow, K. A. I. Effects of Glycosylation on the Stability of Protein Pharmaceuticals. *Biochemistry* **98**, 1223–1245 (2010).
9. Lafite, P. & Daniellou, R. Rare and unusual glycosylation of peptides and proteins. *Nat. Prod. Rep.* **29**, 729–738 (2012).
10. Rabinovich, G. A., van Kooyk, Y. & Cobb, B. A. Glycobiology of immune responses. *Ann. N. Y. Acad. Sci.* **1253**, 1–15 (2012).
11. Wakeland, E. K., Liu, K., Graham, R. R. & Behrens, T. W. Delineating the genetic basis of systemic lupus erythematosus. *Immunity* **15**, 397–408 (2001).
12. Maverakis, E. *et al.* Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review. *J. Autoimmun.* **57**, 1–13 (2015).
13. Taniguchi, N. & Kizuka, Y. Glycans and cancer: Role of N-Glycans in cancer biomarker, progression and metastasis, and therapeutics. *Adv. Cancer Res.* **126**, 11–51 (2015).
14. Ohtsubo, K. & Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126**, 855–867 (2006).
15. Brennan, P. J. & Nikaido, H. The Envelope of Mycobacteria. *Annu. Rev. Biochem.* **64**, 29–63 (1995).
16. De Lederkremer, R. M. & Colli, W. Galactofuranose-containing glycoconjugates in trypanosomatids. *Glycobiology* **5**, 547–552 (1995).
17. Latgé, J.-P., Beauvais, A. & Chamilos, G. The cell wall of the human fungal pathogen *aspergillus fumigatus*: biosynthesis, organization, immune response, and virulence. *Annu. Rev. Microbiol.* **71**, 99–116 (2017).
18. Cherney, A. H., Kadunce, N. T. & Reisman, S. E. Enantioselective and Enantiospecific Transition-Metal-Catalyzed Cross-Coupling Reactions of Organometallic Reagents to Construct C-C Bonds. *Chem. Rev.* **115**, 9587–9652 (2015).

19. Lowary, T. L. Synthesis and conformational analysis of arabinofuranosides, galactofuranosides and fructofuranosides. *Curr. Opin. Chem. Biol.* **7**, 749–756 (2003).
20. Peltier, P., Euzen, R., Daniellou, R., Nugier-Chauvin, C. & Ferrières, V. Recent knowledge and innovations related to hexofuranosides: structure, synthesis and applications. *Carbohydr. Res.* **343**, 1897–1923 (2008).
21. Houseknecht, J. B. & Lowary, T. L. Chemistry and biology of arabinofuranosyl- and galactofuranosyl-containing polysaccharides. *Curr. Opin. Chem. Biol.* **5**, 677–682 (2001).
22. Tefsen, B., Ram, A. F. J., Van Die, I. & Routier, F. H. Galactofuranose in eukaryotes: Aspects of biosynthesis and functional impact. *Glycobiology* **22**, 456–469 (2012).
23. Oppenheimer, M., Valenciano, A. L. & Sobrado, P. Biosynthesis of Galactofuranose in Kinetoplastids: Novel Therapeutic Targets for Treating Leishmaniasis and Chagas' Disease. *Enzyme Res.* **10**, 1–13 (2011).
24. Marino, C., Lima, C., Mariño, K. & De Lederkremer, R. M. Synthesis of a derivative of alpha-D-Glcp(1→2)-D-Galf suitable for further glycosylation and of alpha-D-Glcp(1→2)-D-Gal, a disaccharide fragment obtained from varianose. *Beilstein J. Org. Chem.* **8**, 2142–2148 (2012).
25. Chlubnova, I. *et al.* Specific and non-specific enzymes for furanosyl-containing conjugates: Biosynthesis, metabolism, and chemo-enzymatic synthesis. *Carbohydr. Res.* **356**, 44–61 (2012).
26. Svensson, M. V., Weintraub, A. & Widmalm, G. Structural studies of the O-antigenic polysaccharide from Escherichia coli O177. *Carbohydr. Res.* **346**, 2300–2303 (2011).
27. Stenutz, R., Weintraub, A. & Widmalm, G. The structures of Escherichia coli O-polysaccharide antigens. *FEMS Microbiol. Rev.* **30**, 382–403 (2006).
28. Stevenson, G. *et al.* K-12 and the sequence of its rfb gene cluster. Structure of the O Antigen of Escherichia coli K-12 and the Sequence of Its rfb Gene Cluster. *J. Bacteriol.* **176**, 4144–4156 (1994).
29. Whitfield, C., Richards, J. C., Perry, M. B., Clarke, B. R. & MacLean, L. L. Expression of two structurally distinct D-galactan O antigens in the lipopolysaccharide of Klebsiella pneumoniae serotype O1. *J. Bacteriol.* **173**, 1420–1431 (1991).
30. Hoffman, J., Lindberg, B., Glowacka, M., Derylo, M. & Lorkiewicz, Z. Structural studies of the lipopolysaccharide from Salmonella typhimurium 902 (Collb drd2). *Eur. J. Biochem.* **105**, 103–107 (1980).
31. Murray, G. L., Attridge, S. R. & Morona, R. Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of Salmonella enterica serovar typhimurium with macrophages and complement. *J. Bacteriol.* **188**, 2735–2739 (2006).
32. O'Donoghue, E. J. *et al.* Lipopolysaccharide structure impacts the entry kinetics of bacterial outer membrane vesicles into host cells. *PLoS Pathog.* **13**, 1–18 (2017).
33. Brennan, P. J. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. *Tuberculosis* **83**, 91–97 (2003).
34. Pan, F. *et al.* Cell Wall Core Galactofuran Synthesis Is Essential for Growth of Mycobacteria. *J. Bacteriol.* **183**, 3991–3998 (2001).
35. Koul, A., Arnoult, E., Lounis, N., Guillemont, J. & Andries, K. The challenge of new drug discovery for tuberculosis. *Nature* **469**, 483–490 (2011).

36. Kosmidis, C. & Denning, D. W. The clinical spectrum of pulmonary aspergillosis. *Thorax* **70**, 270–277 (2015).
37. Leitão, E. A. *et al.*  $\beta$ -Galactofuranose-containing O-linked oligosaccharides present in the cell wall peptidogalactomannan of *Aspergillus fumigatus* contain immunodominant epitopes. *Glycobiology* **13**, 681–692 (2003).
38. Levery, S. B., Toledo, M. S., Straus, A. H. & Takahashi, H. K. Structure elucidation of sphingolipids from the mycopathogen *Paracoccidioides brasiliensis*: An immunodominant  $\beta$ -galactofuranose residue is carried by a novel glycosylinositol phosphorylceramide antigen. *Biochemistry* **37**, 8764–8775 (1998).
39. Fu, M., Zhang, G. & Ning, J. First synthesis of the immunodominant  $\beta$ -galactofuranose-containing tetrasaccharide present in the cell wall of *Aspergillus fumigatus*. *Carbohydr. Res.* **340**, 25–30 (2005).
40. Rassi, A., Rassi, A. & Marin-Neto, J. A. Chagas disease. *Seminars* **375**, 1388–1402 (2010).
41. Jones, C., Todeschini, A. R., Agrellos, O. A., Previato, J. O. & Mendonça-Previato, L. Heterogeneity in the biosynthesis of mucin O-glycans from *Trypanosoma cruzi* Tulahuen strain with the expression of novel galactofuranosyl-containing oligosaccharides. *Biochemistry* **43**, 11889–11897 (2004).
42. Turco, S. J. & Descoteaux, A. The lipophosphoglycan of *Leishmania* parasites. *Annu. Rev. Microbiol* **46**, 65–94 (1992).
43. Mcconville, J., Homan, W. & Gal, R. Structures of the Glycoinositolphospholipids *Leishmania* terminal. *J. Biol. Chem.* **265**, 7385–7394 (1990).
44. Marino, C., Rinflerch, A. & de Lederkremer, R. M. Galactofuranose antigens, a target for diagnosis of fungal infections in humans. *Futur. Sci. OA* **3**, 1–14 (2017).
45. Christopher J. Thibodeaux, Charles E. Melançon, and H. L. Natural Product Sugar Biosynthesis and Enzymatic Glycodiversification. *Angew. Chemie - Int. Ed.* **47**, 9814–9859 (2009).
46. Thoden, J. B., Kim, J., Raushel, F. M. & Holden, H. M. The catalytic mechanism of galactose mutarotase. *Protein Sci.* **12**, 1051–1059 (2003).
47. DEY, P. M. Galactokinase of *Vicia faba* seeds. *Eur. J. Biochem.* **136**, 155–159 (1983).
48. Holden, H. M., Rayment, I. & Thoden, J. B. Structure and Function of Enzymes of the Leloir Pathway for Galactose Metabolism. *J. Biol. Chem.* **278**, 43885–43888 (2003).
49. Wagstaff, B. A. *et al.* Enzymatic synthesis of nucleobase-modified UDP-sugars: Scope and limitations. *Carbohydr. Res.* **404**, 17–25 (2015).
50. Marino, C. & Lederkremer, R. M. De. Galactose configurations in nature with emphasis on the biosynthesis of galactofuranose in glycans Galactose configurations in Nature with emphasis on the biosynthesis of Galactofuranose in glycans. *Enzyme Res.* **5**, 1–28 (2016).
51. Kleczka, B. *et al.* Targeted gene deletion of *Leishmania major* UDP-galactopyranose mutase leads to attenuated virulence. *J. Biol. Chem.* **282**, 10498–10505 (2007).
52. Joohae Park, Boris Tefsen, Marc J. Heemskerk, Ellen L. Lagendijk, Cees A. M. J. J. van den Hondel, I. van D. and A. F. J. R. Identification and functional analysis of two Golgi-localized UDP-galactofuranose transporters with overlapping functions in *Aspergillus niger*. *BMC Microbiol.* **15**, 1–11 (2015).
53. Engel, J., Schmalhorst, P. S., Dörk-Bousset, T., Ferrières, V. & Routier, F. H. A single UDP-

- galactofuranose transporter is required for galactofuranosylation in *Aspergillus fumigatus*. *J. Biol. Chem.* **284**, 33859–33868 (2009).
54. Nassau, P. M. *et al.* Galactofuranose biosynthesis in *Escherichia coli* K-12 : identification and cloning of UDP-galactopyranose mutase . Galactofuranose Biosynthesis in *Escherichia coli* K-12 : Identification and Cloning of UDP-Galactopyranose Mutase. *J. Bacteriol.* **178**, 1047–1052 (1996).
  55. Kö, R., Brisson, J.-R. & Whitfield, C. UDP-galactofuranose Precursor Required for Formation of the Lipopolysaccharide O Antigen of *Klebsiella pneumoniae* Serotype O1 Is Synthesized by the Product of the rfbD KPO1 Gene. *J. Bacteriol.* **272**, 4121–4128 (1997).
  56. Scherman, M. *et al.* Biosynthetic origin of mycobacterial cell wall arabinosyl residues . Biosynthetic Origin of Mycobacterial Cell Wall Arabinosyl Residues. **177**, 7125–7130 (1995).
  57. Beverley, S. M. *et al.* Eukaryotic UDP-galactopyranose mutase (GLF Gene) in microbial and metazoal pathogens. *Eukaryot. Cell* **4**, 1147–1154 (2005).
  58. Bakker, H., Kleczka, B., Gerardy-Schahn, R. & Routier, F. H. Identification and partial characterization of two eukaryotic UDP-galactopyranose mutases. *Biol. Chem.* **386**, 657–661 (2005).
  59. John J. Tanner, Leonardo Boechi, J. Andrew McCammon, and Pablo Sobrado John J. Tanner, Leonardo Boechi, J. Andrew McCammon, and P. S. Structure, mechanism, and dynamics of UDP-galactopyranose mutase. *Arch. Biochem. Biophys.* **6**, 247–253 (2009).
  60. Van Straaten, K. E. *et al.* Structural basis of ligand binding to UDP-galactopyranose mutase from *Mycobacterium tuberculosis* using substrate and tetrafluorinated substrate analogues. *J. Am. Chem. Soc.* **137**, 1230–1244 (2015).
  61. Boechi, L. *et al.* Substrate-dependent dynamics of UDP-galactopyranose mutase: Implications for drug design. *Protein Sci.* **22**, 1490–1501 (2013).
  62. Lettera, P. S. Fungus cultures Assay of helminthosporoside. *plant Sci. Lett.* **30**, 145–154 (1983).
  63. Miletti, L. C. *et al.* Immobilized 4-aminophenyl 1-thio-beta-D-galactofuranoside as a matrix for affinity purification of an exo-beta-D-galactofuranosidase. *Carbohydr. Res.* **320**, 176–182 (1999).
  64. Ramli, N., Fujinaga, M., Tabuchi, M., Takegawa, K. & Iwahara, S. Isolation and Characterization of a Novel Endo-  $\beta$  -galactofuranosidase from *Bacillus* sp. *Biosci. Biotechnol. Biochem.* **59**, 1856–1860 (1995).
  65. Wallis, G. L. F., Hemming, F. W. & Peberdy, J. F.  $\beta$ -Galactofuranoside glycoconjugates on conidia and conidiophores of *Aspergillus niger*. *FEMS Microbiol. Lett.* **201**, 21–27 (2001).
  66. Miletti, L. C. *et al.* Evidence for exo  $\beta$ -D-galactofuranosidase in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **127**, 85–88 (2003).
  67. Matsunaga, E. *et al.* Identification and characterization of a novel galactofuranose-specific  $\beta$ -D-galactofuranosidase from *Streptomyces* species. *PLoS One* **10**, 1–16 (2015).
  68. Kremer, L. *et al.* Galactan biosynthesis in *Mycobacterium tuberculosis*: Identification of a bifunctional UDP-galactofuranosyltransferase. *J. Biol. Chem.* **276**, 26430–26440 (2001).
  69. Mikuová, K. *et al.* Biosynthesis of the galactan component of the mycobacterial cell wall. *J. Biol. Chem.* **275**, 33890–33897 (2000).

70. Imberty, A., Joziassse, H. & Geremia, A. Review Sequence-Function Relationships of Prokaryotic and Eukaryotic Relationships of Galactosyltransferases. *Biochem. J.* **1009**, 1000–1009 (1998).
71. Ünligil, U. M. & Rini, J. M. Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.* **10**, 510–517 (2000).
72. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
73. Park, S. H. *et al.* Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. *Biotechnol. Bioeng.* **102**, 988–994 (2009).
74. Liu, J. & Mushegian, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.* **12**, 1418–1431 (2003).
75. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).
76. Tvaroška, I. Molecular modeling insights into the catalytic mechanism of the retaining galactosyltransferase LgtC. *Carbohydr. Res.* **339**, 1007–1014 (2004).
77. Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J. & Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **16**, 29–37 (2006).
78. Anttinen, H. & Kivirikko, K. I. Affinity chromatography of collagen glucosyltransferase on a UDP-glucose derivative coupled to agarose. *BBA - Enzymol.* **429**, 750–758 (1976).
79. Rose, N. L. *et al.* Expression , Purification , and Characterization of a Galactofuranosyltransferase Involved in Mycobacterium tuberculosis Arabinogalactan Biosynthesis. *J. Am. Chem. Soc.* **128**, 6721–6729 (2006).
80. Belánová, M. *et al.* Galactosyl transferases in mycobacterial cell wall synthesis. *J. Bacteriol.* **190**, 1141–1145 (2008).
81. Szczepina, M. G., Zheng, B., Completo, G. C., Lowary, T. L. & Pinto, B. M. STD-NMR Studies Suggest that Two Acceptor Substrates for GlfT2 , a Bifunctional Galactofuranosyltransferase Required for the Biosynthesis of Mycobacterium tuberculosis Arabinogalactan , Compete for the Same Binding Site. *ChemBioChem* **10**, 2052–2059 (2009).
82. May, J. F., Splain, R. A., Brotschi, C. & Kiessling, L. L. A tethering mechanism for length control in a processive carbohydrate polymerization. *Proc. Natl. Acad. Sci.* **106**, 11851–11856 (2009).
83. Wheatley, R. W., Zheng, R. B., Richards, M. R., Lowary, T. L. & Ng, K. K. S. Tetrameric Structure of the GlfT2 Galactofuranosyltransferase Reveals a Scaffold for the Assembly of Mycobacterial. *J. Biol. Chem.* **287**, 28132–28143 (2012).
84. de Groot, P. W. J. *et al.* Comprehensive genomic analysis of cell wall genes in *Aspergillus nidulans*. *Fungal Genet. Biol.* **46**, S72–S81 (2009).
85. Morelle, W. *et al.* Galactomannoproteins of *Aspergillus fumigatus*. *Eukaryot. Cell* **4**, 1308–1316 (2005).
86. Schmalhorst, P. S. *et al.* Contribution of galactofuranose to the virulence of the opportunistic pathogen *Aspergillus fumigatus*. *Eukaryot. Cell* **7**, 1268–1277 (2008).
87. Komachi, Y. *et al.* gfsA encodes a novel galactofuranosyltransferase involved in biosynthesis of galactofuranose antigen of O-glycan in *Aspergillus nidulans* and *Aspergillus*

- fumigatus. *Mol. Microbiol.* **90**, 1054–1073 (2013).
88. Katafuchi, Y. *et al.* GfsA is a  $\beta$ 1,5-galactofuranosyltransferase involved in the biosynthesis of the galactofuran side chain of fungal-type galactomannan in *Aspergillus fumigatus*. *Glycobiology* **27**, 568–581 (2017).
  89. Wing, C., Errey, J. C., Mukhopadhyay, B., Blanchard, S. & Field, R. A. Expression and initial characterization of WbbI, a putative D-Gal f : a - D -Glc b -1, 6-galactofuranosyltransferase from *Escherichia coli* K-12. *Org. Biomol. Chem.* **21**, 3945–3950 (2006).
  90. Handler, M. Z., Patel, P. A., Kapila, R., Al-Qubati, Y. & Schwartz, R. A. Cutaneous and mucocutaneous leishmaniasis: Clinical perspectives. *J. Am. Acad. Dermatol.* **73**, 897–908 (2015).
  91. Torres-Guerrero, E., Quintanilla-Cedillo, M. R., Ruiz-Esmenjaud, J. & Arenas, R. Leishmaniasis: a review. *F1000Research* **6**, 750 (2017).
  92. Andrade-Narváez, F. J., Vargas-González, A., Canto-Lara, S. B. & Damián-Centeno, A. G. Clinical picture of cutaneous leishmaniasis due to *Leishmania (Leishmania) mexicana* in the Yucatan peninsula, Mexico. *Memórias do Inst. Oswaldo Cruz* **96**, 163–7 (2001).
  93. Alvar, J., Yactayo, S. & Bern, C. Leishmaniasis and poverty. *Trends Parasitol.* **22**, 552–557 (2006).
  94. Murray, H. W., Berman, J. D., Davies, C. R. & Saravia, N. G. Advances in leishmaniasis. *Lancet Infect. Dis.* **366**, 1561–1577 (2005).
  95. Ross, M. R. Further notes on leishman's bodies. *Br. Med. J.* **2**, 1401 (1903).
  96. Reithinger, R. *et al.* Cutaneous leishmaniasis. *Lancet Infect. Dis.* **7**, 581–596 (2007).
  97. Borghi, S. M. *et al.* Leishmania infection: painful or painless? *Parasitol. Res.* **116**, 465–475 (2017).
  98. Akhoundi, M. *et al.* A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies. *PLoS Negl. Trop. Dis.* **10**, 1–40 (2016).
  99. Consuelo V. David, N. C. Cutaneous and mucocutaneous leishmaniasis. *Dermatol. Ther.* **22**, 491–502 (2009).
  100. Dantas-torres, F. *et al.* Canine leishmaniosis in the Old and New Worlds: unveiled similarities and differences. *Trends Parasitol.* **28**, 531–538 (2012).
  101. Moradin, N., Descoteaux, A. & Beverley, S. M. Leishmania promastigotes: building a safe niche within macrophages. *Cell. Infect. microbiology* **2**, 1–7 (2012).
  102. Brandonisio, O., Spinelli, R. & Pepe, M. Dendritic cells in Leishmania infection. *Microbes Infect.* **6**, 1402–1409 (2004).
  103. Feijó, D., Tibúrcio, R., Ampuero, M., Brodskyn, C. & Tavares, N. Dendritic cells and leishmania infection: Adding layers of complexity to a complex disease. *J. Immunol. Res.* **2016**, (2016).
  104. Liu, D. & Uzonna, J. E. The early interaction of Leishmania with macrophages and dendritic cells and its influence on the host immune response. *Front. Cell. Infect. Microbiol.* **2**, 1–8 (2012).
  105. Descoteaux, A. & Turco, S. J. Glycoconjugates in Leishmania infectivity. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1455**, 341–352 (1999).
  106. Bates, P. A. Transmission of Leishmania metacyclic promastigotes by phlebotomine sand

- flies. *Int. J. Parasitol.* **37**, 1097–1106 (2007).
107. Vries, H. J. C. De, Reedijk, S. H. & Schallig, H. D. F. H. Cutaneous Leishmaniasis : Recent Developments in Diagnosis and Management. *Am. J. Clin. Dermatol.* **16**, 99–109 (2015).
  108. Dowlati, Y. Cutaneous leishmaniasis: Clinical aspect. *Clin. Dermatol.* **14**, 425–431 (1996).
  109. Weigle, K. & Saravia, N. G. Natural history, clinical evolution, and the host-parasite interaction in new world cutaneous leishmaniasis. *Clin. Dermatol.* **14**, 433–450 (1996).
  110. Griensven, J. Van, Gadisa, E., Aseffa, A. & Hailu, A. Treatment of Cutaneous Leishmaniasis Caused by *Leishmania aethiopica* : A Systematic Review. *PLoS neglected Trop. Dis. Trop. Dis.* **10**, 1–21 (2016).
  111. Consuelo V. David, N. C. Cutaneous and mucocutaneous leishmaniasis. *Dermatol. Ther.* **22**, 491–502 (2009).
  112. Piscopo, T. V & Mallia Azzopardi, C. Leishmaniasis. *Postgrad. Med. J.* **83**, 649–657 (2007).
  113. Reithinger, R., Dujardin, J. & Louzir, H. Cutaneous leishmaniasis. *Lancet Infect. Dis.* **7**, 581–596 (2007).
  114. Alvar, J. *et al.* Leishmaniasis worldwide and global estimates of its incidence. *PLoS One* **7**, 1–12 (2012).
  115. Lindoso, J. A. L., Cunha, M. A., Queiroz, I. T. & Moreira, C. H. V. Leishmaniasis–HIV coinfection: Current challenges. *HIV/AIDS - Res. Palliat. Care* **8**, 147–156 (2016).
  116. Marty, P. *et al.* A century of leishmaniasis in Alpes-Maritimes , France. *Ann. Trop. Med. Parasitol.* **101**, 563–574 (2007).
  117. Neitzke-Abreu, H. C. *et al.* Detection of DNA from *Leishmania* (Viannia): Accuracy of Polymerase Chain Reaction for the Diagnosis of Cutaneous Leishmaniasis. *PLoS One* **8**, 1–7 (2013).
  118. Cardoso, G. *et al.* Development and validation of PCR-based assays for diagnosis of American cutaneous leishmaniasis and identification of the parasite species. *Mem. Inst. Oswaldo Cruz* **107**, 664–674 (2012).
  119. Raghu SINHA, Sunil K. ARORA, Usha DATTA, and S. S. Detection of *Leishmania* Antigen in Kala Azar Patients Using Monoclonal Antibodies Abstract Twenty-one monoclonal antibodies were produced against promastigote ing from 42-116 kDa were selected as ' capture antibodies ' and compared with kala azar patients. *Microbiol. Immunol.* **36**, 391–400 (1992).
  120. Pedral-Sampaio, G. *et al.* Detection of IgG Anti-leishmania antigen by flow cytometry as a diagnostic test for cutaneous leishmaniasis. *PLoS One* **11**, 1–9 (2016).
  121. Célia Maria Ferreira Gontijo; Melo Norma Maria. Laboratory Diagnosis of Visceral Leishmaniasis Laboratory Diagnosis of Visceral Leishmaniasis. *Clin. Diagn. Lab. Immunol.* **9**, 951–958 (2002).
  122. Baiocco, P., Colotti, G., Franceschini, S. & Ilari, A. Molecular Basis of Antimony Treatment in Leishmaniasis. *J. Med. Chem* **52**, 2603–2612 (2009).
  123. Wyllie, S., Cunningham, M. L. & Fairlamb, A. H. Dual action of antimonial drugs on thiol redox metabolism in the human pathogen *Leishmania donovani*. *J. Biol. Chem.* **279**, 39925–39932 (2004).
  124. Palma, E. *et al.* Antileishmanial Activity of Amphotericin B-loaded-PLGA Nanoparticles: An Overview. *Materials (Basel).* **11**, 1167 (2018).

125. Bern, C. *et al.* Reviews Of Anti-infective Agents: Liposomal Amphotericin B for the Treatment of Visceral Leishmaniasis. *Clin. Infect. Dis.* **43**, 917–924 (2006).
126. B. DE KRUIJFF, W. J. GERRITSEN, A. OERLEMANS, R. A. D. and L. L. M. V. D. Polyen antibiotic-sterol interactions in membranes of *Alcholplasma laidlawii* cells and lecithin liposomes. *Biochem. Biophys. acta* **339**, 30–43 (1974).
127. Croft, S. L. & Engel, J. Miltefosine discovery of the antileishmanial activity of phospholipid derivatives. *Trans. R. Soc. Trop. Med. Hyg* **100**, 4–8 (2006).
128. Sindermann, H. & Engel, J. Development of miltefosine as an oral treatment for leishmaniasis. *Trans. R. Soc. Trop. Med. Hyg* **100**, 2–5 (2006).
129. Vakil, N. H., Fujinami, N. & Shah, P. J. Pharmacotherapy for leishmaniasis in the United States: Focus on miltefosine. *Pharmacotherapy* **35**, 536–545 (2015).
130. Sunyoto, T., Potet, J. & Boelaert, M. Why miltefosine a life-saving drug for leishmaniasis is unavailable to people who need it the most. *BMJ Glob. Heal.* **3**, 1–10 (2018).
131. Banjara, M. R. *et al.* Visceral Leishmaniasis Clinical Management in Endemic Districts of India, Nepal, and Bangladesh. *J.Trop.Med* **2**, 1–8 (2012).
132. Croft, S. L., Sundar, S. & Fairlamb, A. H. Drug Resistance in Leishmaniasis Drug Resistance in Leishmaniasis. *Clin. Microbiol. Rev.* **19**, 111–126 (2006).
133. Croft, S. L. Monitoring drug resistance in leishmaniasis. *Trop. Med. Int. Heal.* **6**, 899–905 (2001).
134. Novozhilova, N. M. & Bovin, N. V. Structure, functions, and biosynthesis of glycoconjugates of *Leishmania* spp. cell surface. *Biochem.* **75**, 686–694 (2010).
135. Cabezas, Y. *et al.* *Leishmania* cell wall as a potent target for antiparasitic drugs. A focus on the glycoconjugates. *Org. Biomol. Chem.* **13**, 8393–8404 (2015).
136. Garraway, L. A., Turco, S. J., Beverley, S. M. & Spa, G. F. The role ( s ) of lipophosphoglycan ( LPG ) in the establishment of *Leishmania* major infections in mammalian hosts. *PNAS* **100**, 9536–9541 (2003).
137. Teixeira, D. E. *et al.* The Cell Biology of *Leishmania*: How to Teach Using Animations. *PLoS Pathog.* **9**, 7–11 (2013).
138. Proudfoot, L., Schneider, P., Ferguson, M. a & McConville, M. J. Biosynthesis of the glycolipid anchor of lipophosphoglycan and the structurally related glycoinositolphospholipids from *Leishmania* major. *Biochem. J.* **308**, 45–55 (1995).
139. Pimenta, P. F. P., Saraiva, E. M. B. & Sacks, D. L. The comparative fine structure and surface glycoconjugate expression of three life stages of *Leishmania* major. *Exp. Parasitol.* **72**, 191–204 (1991).
140. Moody, S. F., Handman, E., McConville, M. J. & Bacic, A. The structure of *Leishmania* major amastigote lipophosphoglycan. *J. Biol. Chem.* **268**, 18457–18466 (1993).
141. De Assis, R. R., Ibraim, I. C., Nogueira, P. M., Soares, R. P. & Turco, S. J. Glycoconjugates in New World species of *Leishmania*: Polymorphisms in lipophosphoglycan and glycoinositolphospholipids and interaction with hosts. *Biochim. Biophys. Acta - Gen. Subj.* **1820**, 1354–1365 (2012).
142. McConville, M. J., Thomas-Oates, J. E., Ferguson, M. A. & Homans, S. W. Structure of the lipophosphoglycan from *Leishmania* major. *J Biol Chem* **265**, 19611–19623 (1990).
143. Guha-Niyogi, A., Sullivan, D. R. & Turco, S. J. Glycoconjugate structures of parasitic protozoa.

- Glycobiology* **11**, 45–59 (2001).
144. Descoteaux, A. & Turco, S. J. Functional aspects of the *Leishmania donovani* lipophosphoglycan during macrophage infection. *Microbes Infect.* **4**, 975–981 (2002).
  145. McConville MJ, Schnur LF, Jaffe C, S. P. *et al.* Structure of *Leishmania* lipophosphoglycan: inter- and intra-specific polymorphism in Old World species. *Biochem* **310**, 807–818 (1995).
  146. Kamhawi, S. *et al.* A role for insect galectins in parasite survival. *Cell* **119**, 329–341 (2004).
  147. Dillon, R. J. *et al.* Analysis of ESTs from *Lutzomyia longipalpis* sand flies and their contribution toward understanding the insect-parasite relationship. *Genomics* **88**, 831–840 (2006).
  148. Lodge, R. & Descoteaux, A. Modulation of phagolysosome biogenesis by the lipophosphoglycan of *Leishmania*. *Clin. Immunol.* **114**, 256–265 (2005).
  149. Pimenta, P. F. P. *et al.* Evidence that the vectorial competence of phlebotomine sand flies for different species of *Leishmania* is controlled by structural polymorphisms in the surface lipophosphoglycan. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9155–9159 (1994).
  150. Kamhawi, S. Phlebotomine sand flies and *Leishmania* parasites: friends or foes? *Trends Parasitol.* **22**, 439–445 (2006).
  151. Olivier, M., Gregory, D. J. & Forget, G. Subversion Mechanisms by Which *Leishmania* Parasites Can Escape the Host Immune Response : a Signaling Point of View Subversion Mechanisms by Which *Leishmania* Parasites Can Escape the Host Immune Response : a Signaling Point of View. *Clin. Microbiol. Rev.* **18**, 293–305 (2005).
  152. Dermine, J. F., Scianimanico, S., Privé, C., Descoteaux, A. & Desjardins, M. *Leishmania* promastigotes require lipophosphoglycan to actively modulate the fusion properties of phagosomes at an early step of phagocytosis. *Cell. Microbiol.* **2**, 115–126 (2000).
  153. Oliveira, A.-C. *et al.* Expression of Functional TLR4 Confers Proinflammatory Responsiveness to *Trypanosoma cruzi* Glycoinositolphospholipids and Higher Resistance to Infection with *T. cruzi*. *J. Immunol.* **173**, 5688–5696 (2004).
  154. Menon, A. K., Mayor, S. & Schwarz, R. T. Biosynthesis of glycosyl-phosphatidylinositol lipids in *Trypanosoma brucei*: involvement of mannosyl-phosphoryldolichol as the mannose donor. *EMBO J.* **9**, 4249–58 (1990).
  155. Ha, D. S., Schwarz, J. K., Turco, S. J. & Beverley, S. M. Use of the green fluorescent protein as a marker in transfected *Leishmania*. *Mol. Biochem. Parasitol.* **77**, 57–64 (1996).
  156. Ryan, K. A., Garraway, L. A., Descoteaux, A., Turcot, S. J. & Beverley, S. M. Isolation of virulence genes directing surface glycosyl- phosphatidylinositol synthesis by functional complementation of *Leishmania*. *Proc. Natl. Acad. Sci. USA.* **90**, 8609–8613 (1993).
  157. Brown, G. M. *et al.* Synthetic phospho-oligosaccharide fragments of lipophosphoglycan as acceptors for *Leishmania major* alpha-D-mannosylphosphate transferase. *Eur. J. Biochem.* **242**, 410–416 (1996).
  158. Spath, G. F., Garraway, L. A., Turco, S. J. & Beverley, S. M. The role(s) of lipophosphoglycan (LPG) in the establishment of *Leishmania major* infections in mammalian hosts. *PNAS* **100**, 9536–9541 (2003).
  159. Capul, A. A., Barron, T., Dobson, D. E., Turco, S. J. & Beverley, S. M. Two functionally divergent UDP-Gal nucleotide sugar transporters participate in phosphoglycan synthesis in *Leishmania major*. *J. Biol. Chem.* **282**, 14006–14017 (2007).

160. Lamotte, S., Späth, G. F., Rachidi, N. & Prina, E. The enemy within: Targeting host-parasite interaction for antileishmanial drug discovery. *PLoS Negl. Trop. Dis.* **11**, e0005480 (2017).
161. Rittig, M. G. & Bogdan, C. Leishmania – Host-cell Interaction : **16**, (2000).
162. Michelle Oppenheimer, Ana L. Valenciano, and P. S. Isolation and characterization of functional Leishmania major virulence factor UDP-galactopyranose mutase. *Biochem. Biophys. Res. Commun* **4**, 552–563 (2011).
163. Rose, N. L. *et al.* Expression , Purification , and Characterization of a Galactofuranosyltransferase Involved in Mycobacterium tuberculosis Arabinogalactan Biosynthesis. *JACS* **128**, 6721–6729 (2006).
164. Zhang, K., Barron, T., Turco, S. J. & Beverley, S. M. The LPG1 gene family of Leishmania major. *Mol. Biochem. Parasitol.* **136**, 11–23 (2004).
165. Spath, G. F. *et al.* Lipophosphoglycan is a virulence factor distinct from related glycoconjugates in the protozoan parasite Leishmania major. *Proc. Natl. Acad. Sci.* **97**, 9258–9263 (2000).
166. Sahdev, S., Khattar, S. K. & Saini, K. S. Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. *Mol. Cell. Biochem.* **307**, 249–264 (2008).
167. Jones, J. D. Leishmania tarentolae: An alternative approach to the production of monoclonal antibodies to treat emerging viral infections. *Infect. Dis. Poverty* **4**, 1–5 (2015).
168. Khan, A. H., Bayat, H., Rajabibazl, M., Sabri, S. & Rahimpour, A. Humanizing glycosylation pathways in eukaryotic expression systems. *World J. Microbiol. Biotechnol.* **33**, (2017).

***Chapter III: LPG1x production in  
prokaryotic host.***

## Abstract

Four genes were identified in *Leishmania major* genome, *lpg1*, *lpg1L*, *lpg1G* and *lpg1R*. They were predicted to encode for four putative galactofuranosyltransferases. Moreover, they are expected to be type II membrane proteins and to be located in the membrane of Golgi apparatus of the parasite. Thus, these proteins are supposed to be glycosylated, which could be an essential element for their structure and activity. However, in this chapter, prokaryotic hosts, *Escherichia coli*, were used to produce the proteins of interest. Since, bacteria lack the post translational mechanism such as glycosylation, the final proteins are non-glycosylated. We chose this heterologous expression system because uncountable active and well folded prokaryotic and eukaryotic recombinant proteins were produced by *E. coli*. In addition, this strategy will allow us to evaluate the effect of the deglycosylation on the activity and stability of proteins of interest.

To facilitate their purification, the transmembrane domain, which is located in the *N*-terminus, was removed. The *lpg1x* genes were first cloned into two expression plasmids pET-28a(+) and pET-32a(+) in order to obtain recombinant proteins fused with hexa-histidine tag in their *N*-termini. The overexpression and purification trials showed that pET-28a(+) was more efficient than pET-32a(+) to express LPG1x proteins. However, we have encountered protein solubilization problems in addition to their co-purification with chaperonin, which indicates that they are unfolded. Then, another plasmid, pMAL-C2X was used that allowed us to obtain pure and soluble LPG1X with high yield.

The production of the four putative galactofuranosyltransferases of *Leishmania* is described for the first time in this chapter. This represents a key step before their enzymatic and structural characterization.

## Résumé

Quatre gènes ont été identifiés dans le génome de *Leishmania major*, *lpg1*, *lpg1L*, *lpg1G* et *lpg1R*. Ils ont été prédits pour coder des galactofuranosyltransférases. Elles doivent être des protéines membranaires de type II qui sont situées dans la membrane de l'appareil de Golgi du parasite. Ainsi, ces protéines sont supposées être glycosylées, ce qui pourrait être un élément essentiel pour leur structure et leur activité. Dans ce chapitre, des hôtes procaryotes, *Escherichia coli*, ont été utilisés pour produire les protéines d'intérêt. Comme les bactéries ne possèdent pas le mécanisme post-traductionnel tel que la glycosylation, les protéines qu'elles expriment ne seront pas glycosylées. Ce système d'expression hétérologue a été choisi car il a permis la production de nombreuses protéines recombinantes procaryotes et eucaryotes fonctionnelles et bien repliées. De plus, cette stratégie nous permettra d'évaluer l'effet de la dé-glycosylation sur l'activité et la stabilité des protéines d'intérêt.

Afin de faciliter leur purification, le domaine transmembranaire, situé dans l'extrémité *N*-terminale, a été clivé. Les gènes *lpg1x* ont d'abord été clonés dans deux plasmides d'expression pET-28a(+) et pET-32a(+) afin d'obtenir des protéines recombinantes fusionnées avec un marqueur hexa-histidine dans leurs extrémités *N*-terminales. Les essais de surexpression et de purification ont montré que le pET-28a(+) était plus efficace que le pET-32a(+) pour exprimer les protéines LPG1x. Cependant, nous avons rencontré des problèmes de solubilisation des protéines en plus de leur co-purification avec une protéine chaperone, ce qui indique qu'elles ne sont pas bien repliées. Ensuite, un autre plasmide, pMAL a été utilisé ce qui nous a permis enfin d'obtenir quatre LPG1x pures et solubles avec un rendement élevé.

La production des quatre potentiels galactofuranosyltransférases de *Leishmania* est décrite pour la première fois dans ce chapitre. Ceci représente une étape clé pour leur caractérisation enzymatique et structurelle

I.	Recombinant protein production.....	87
A.	<i>E. coli</i> as a bacterial expression system: advances and challenges.....	87
1.	The well-suitable expression vectors.....	88
2.	How to “extract” proteins from <i>E. coli</i> ? .....	90
3.	Obtaining pure proteins .....	91
4.	Inclusion bodies and solubility of proteins: Common problems related to recombinant protein production in <i>E. coli</i> .....	91
II.	LPG1x production: A challenging goal .....	91
A.	Genomic and structural study of four putative GalTs.....	91
B.	LPG1x cloning, overexpression and purification .....	93
1.	pET vectors as heterologous expression system .....	93
2.	Cloning.....	95
3.	Overexpression of LPG1x.....	96
a)	Influence of the temperature and the induction time.....	96
b)	Influence of the plasmid .....	98
c)	First strategy of purification .....	99
d)	Detergent effect on protein solubility .....	101
e)	Alternative lysis protocol .....	102
4.	Identification of the contaminant.....	104
5.	Removal of chaperonin GroEL.....	106
a)	ATPase activity.....	106
b)	Other <i>E. coli</i> strains.....	107
c)	C- and N- histidine tags.....	108
6.	pMal vectors as heterologous expression system .....	110
a)	Overexpression .....	111
b)	Purification of LPG1X-MBP .....	112
c)	Protein characterization.....	113
d)	MBP-Tag removal .....	115
III.	Conclusion .....	116
	<b>References</b> .....	118

## I. Recombinant protein production

Protein purification is considered as vital procedure for research or industrial field. There are many reasons for expressing and purifying recombinant proteins, such as obtaining full enzymatic and structural characterization of new therapeutic targets, or producing in large scale, enzymes that are used as biocatalysts which could be commercialized.<sup>1,2,3</sup>

As mentioned before, LPG1x proteins are produced by *Leishmania*, which are eukaryotic species. Moreover, they are predicted to be type II membrane proteins that are characterized by their cytoplasmic *N*-terminal and are located in Golgi apparatus.<sup>4</sup> Thus, the proteins of interest might be glycosylated. It is known that many GTs are glycosylated by other GTs, or in some cases by autocatalytic transferase activities.<sup>5</sup> Here, glycosylation could play an important role for their good folding and stability during synthesis and for their activity.<sup>6</sup> Thus, it is important to choose the well-suitable organism to produce the proteins.<sup>7</sup> In order to get glycosylated GalTs, eukaryotic expression system was used, and is discussed in the third chapter. However, in this chapter, we will present the production of LPG1x using prokaryotic host, *E. coli*. We decided to use bacteria because they have been exploited for the production of a variety eukaryotic GTs.<sup>8,9,10,11</sup>

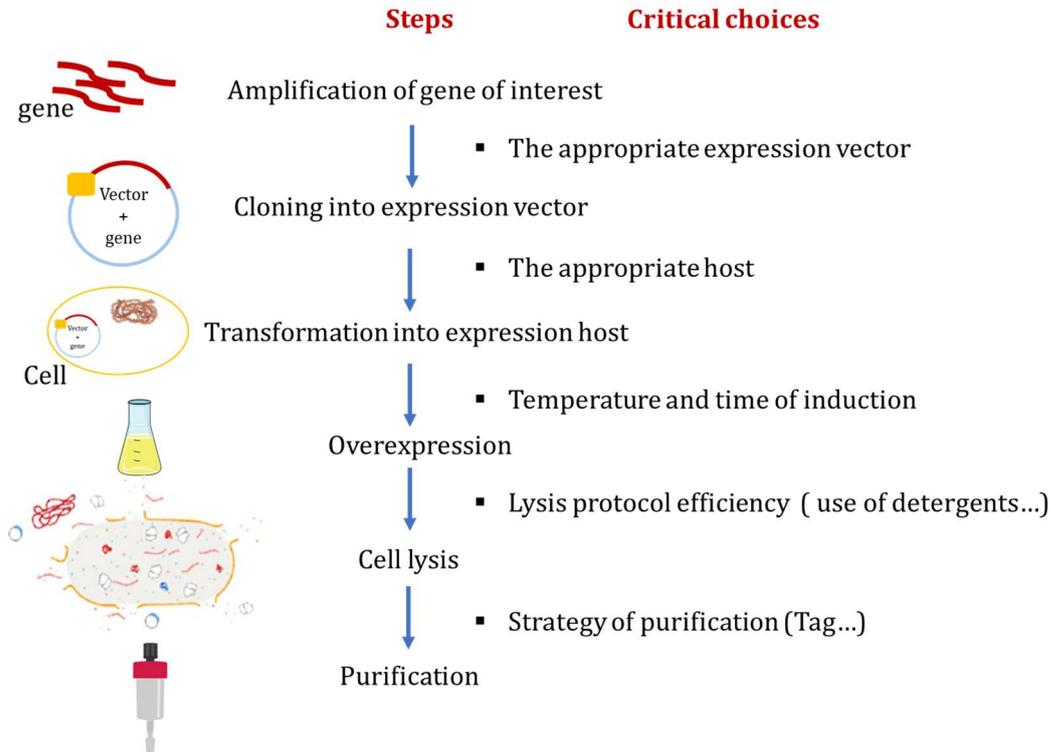
With this strategy, LPG1x will not be glycosylated, because bacteria lack the post-translational mechanism like glycosylation.<sup>12</sup> As a result, it will allow us to see if glycosylation is essential for the activity of these enzymes and compare with results obtained with eukaryotic expression system described in chapter 3.

### A. *E. coli* as a bacterial expression system: advances and challenges

Nowadays, the scientific community has extensive knowledge about the genome and the molecular biology of *E. coli*. Since the production of recombinant insulin and growth hormone earlier in 1980's, *E. coli* has been widely used to produce proteins with complex structure such as antibodies, blood coagulation factors and enzymes.<sup>13,14</sup> Bacteria are the preferred host because they are easily handled, they grow fast and require inexpensive medium.<sup>15</sup> However, these organisms have some limits. Indeed, not every gene that has been inserted into *E. coli* was efficiently expressed.<sup>16</sup> This is related to various factors, that make the production of pure and active proteins very challenging.<sup>14</sup> Ferguson et al., and Teresa San-Miguel et al., had difficulties on getting soluble GTs.<sup>17,18</sup>

Basically, few steps are required to get pure and soluble recombinant proteins. First, genes of interest are isolated and amplified by PCR (Polymerase Chain Reaction). Second, they are inserted into specific cloning or expression vectors. Then the whole construct is transferred into expression-host, that will produce and overexpress desired proteins. Finally, proteins of interest are purified and ready to be characterized. This seems to be easy, but, it is not, because each step

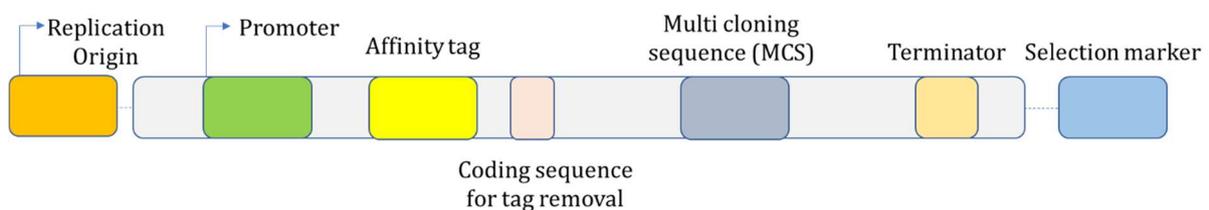
could have its own difficulties and this could be very fastidious. We can face problem growth of the host, difficulties to solubilize the target proteins, and unsuccessful purification step. This is because each protein is unique and needs specific conditions depending on its properties. To overcome these difficulties, there are some essential steps that need to be studied (Figure III-1).



**Figure III- 1:** General steps required to produce recombinant proteins.

1. The well-suitable expression vectors

Various expression vectors are commercially available and suitable for *E. coli* expression system. Unfortunately, the “perfect” plasmid does not exist. A choice has to be taken according some essential elements that will initiate the outline of the whole process (Figure III-2):



**Figure III- 2:** Scheme representing the essential element present into expression vector

- **Replicon**, known as the origin of replication. It is important because it ensures the obtention of moderate to high-level copy number of plasmids. We could think that the more copies we got, the more proteins we will have. But, this could affect the growth of bacteria and the stability of the expression, resulting a low yield of proteins.<sup>19,20</sup>
- **Suitable promoter** that ensures the overexpression of interest. One of the most used one is T7 bacteriophage promoter found in pET (Novagen) vectors.<sup>21</sup> This promoter is recognized by T7 RNA polymerase (T7 RNAP), provided by the bacterial expression strain. The system is controlled by the lactose/operon lactose system which is important to induce and overexpress recombinant proteins (Figure III-3).<sup>22</sup> Indeed, the host DNA contains a *Lac I* gene that encodes for protein referred to lac repressor. In the absence of the IPTG, the lac repressor is on its active state and can then recognize and binds the lac operon site. Therefore, the T7 RNA polymerase is unable to recognize the promoter and the genes of interest are not transcribed. The isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) (Figure III-4), once is added in the medium, binds to the repressor causing its conformational changes, that prevent its interaction with the lac operon, allowing the transcription of target genes. The characteristics of the molecule of IPTG, is that it has a sulfur (S) atom that creates a non-hydrolysable chemical bond.<sup>23</sup> Thus, the concentration of IPTG remains the same and the transcription of the genes is constant during the experiments. In addition, this system keeps the transcription of genes of interest silent in the uninduced state. Consequently, the instability of the plasmid that can come from the continuous production of potential toxic proteins to *E. coli* can be avoided.<sup>14</sup>  
There is also the Taq promoter found in pMAL vectors (NEB) that boosts the expression of recombinant proteins and is induced by the presence of IPTG.<sup>24,25,26,1,22</sup> However, the induction step should be optimized to avoid the recombinant proteins to become toxic for host cells and thus formation of inclusion bodies.<sup>26,27</sup>

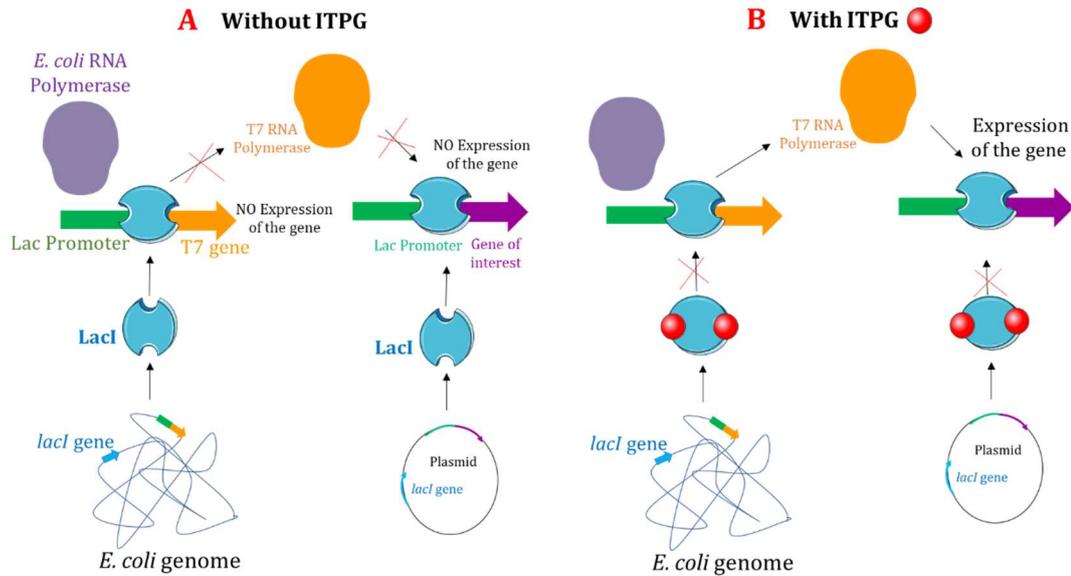


Figure III- 3: Induction mechanism without and with IPTG

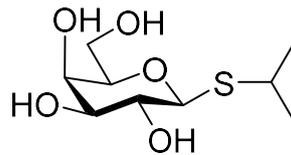


Figure III- 4: IPTG structure

- **Selection marker** allows selection of free plasmids bacteria thanks to a simple addition of appropriate antibiotics.
- **Affinity tag** regions allow the obtention of fused proteins with specific tag (Hexa-histidine, MBP...). It can increase the solubility of proteins, and enable their purification and detection. Using affinity is a good benefit, but it could alter the folding of the proteins.<sup>28</sup>, In some cases, it is possible to remove the tag by enzymatic digestion thanks to the presence of cleavage sites downstream or upstream of the gene coding for the tag.<sup>21</sup>

## 2. How to “extract” proteins from *E. coli*?

Proteins which are produced by *E. coli* are expressed in the cytoplasm. So, breaking the cells is essential and considered as crucial step to liberate target proteins. Different protocols can be performed. We can use mechanical lysis by sonication, freeze-thaw procedure or high-pressure homogenization.<sup>29,27</sup> Also, there is enzymatic lysis method thanks to globular protein, lysozyme, which is capable of bacterial membrane degradation.<sup>30,31,32</sup> However, *E. coli* are Gram negative bacteria which are known to be more resistant against the activity of lysozyme thanks to

lipophospholipids (LPS) that covers their peptidoglycans, which are the target of lysozymes. To optimize the lysis of cells, a combination between different methods is recommended.<sup>28</sup>

### 3. Obtaining pure proteins

Purification is the last step and is related to the chosen expression vector. Indeed, at the beginning of each project, if we decide to produce fused protein with affinity tag, the most common purification method consist on the use of affinity chromatography. It is an efficient method based on the highly interaction between the tag of the protein and specific molecule fixed to a resin stored into column. After different washing steps, protein of interest can be eluted. This method can be followed by ion exchange chromatography based on the separation of ionizable molecules based on their total charge. And the final and essential method, is the size exclusion or gel filtration chromatography.<sup>12</sup> The purpose of this step is to remove aggregates and contaminants, to finally, obtaining pure proteins into a suitable buffer.

### 4. Inclusion bodies and solubility of proteins: Common problems related to recombinant protein production in *E. coli*

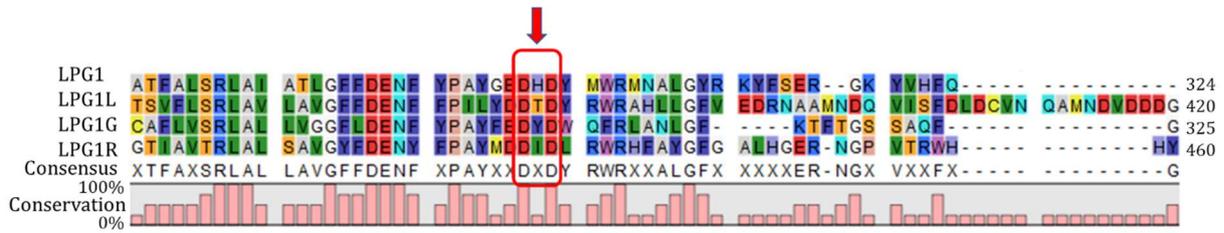
Overexpression of eukaryotic proteins into bacteria often leads to the formation of inclusion bodies. They are characterized by the accumulation of misfolded, insoluble and aggregated proteins. Inclusion bodies remain a significant barrier to gene expression in the cytosol. Some protocols were developed to extract “denaturated” proteins of interest.<sup>33,27</sup> However, to refold correctly the proteins we need to have access to precious information about their structure and activity. To prevent the inclusion bodies, reducing the growth rate of the cells by inducing the expression of proteins for short time and at low temperature is recommended.<sup>16,34,35</sup>

## II. LPG1x production: A challenging goal

### A. Genomic and structural study of four putative GalTs

Up to date, four genes have been identified in the genome of *Leishmania major* that could encode for putative galactofuranosyltransferases. There is one copy of *lpg1* (LmjF.25.0010), *lpg1L* (LmjF.26.0550), *lpg1R* (LmjF.33.0300) and three identical copies of *lpg1G* gene (LmjF.32.3990, LmjF.05.1230, LmjF.19.0650).<sup>4,36</sup> They have respectively 1305, 1617, 1383 and 1800 base pairs (bp). *Leishmania* genes are free of introns, thus for cloning step, chromosomal DNA can be used as template.<sup>37</sup> These genes have never been over-expressed before and are expected to encode for four GalTs with transmembrane domains and DXD conserved catalytic motifs.<sup>4</sup>

The alignment of the four amino acid sequences of LPG1x confirmed the presence of conserved catalytic DXD domain; DHD, DTD, DYD and DID for LPG1, LPG1L, LPG1R and LPG1G respectively (Figure III-5).



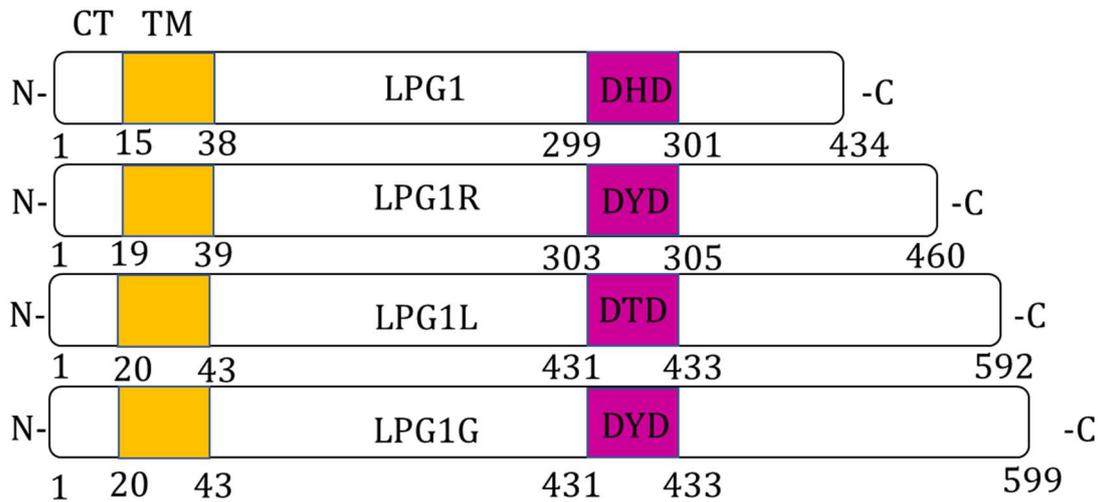
**Figure III- 5:** Partial amino acid sequence alignment of LPG1, LPG1L, LPG1R and LPG1G realized with CLC sequence viewer software. Names of the corresponding coding sequences are shown on the right of the sequences. Conserved DXD-motif is marked in the red box. These DXD-motif positions are located in: 299-301aa for LPG1, 431-433aa for LPG1L, 303-305aa for LPG1R and 431-433aa for LPG1G.

	LPG1	LPG1L	LPG1R	LPG1G	GlT1	GlT2	GfsA	WbbI
LPG1		92% 29%	91% 30%	65% 27%	N.H	N.H	N.H	N.H
LPG1L			74% 28%	48% 29%	N.H	N.H	N.H	N.H
LPG1R				60% 30%	N.H	N.H	N.H	N.H
LPG1G					N.H	N.H	N.H	N.H

**Table III- 1:** Summary table of the homology of LPG1x proteins, GlT, GlT2, GfsA and WbbI. Coverage percentage; Identity percentage; N.H: Non-Homologous (Less than 50% of covered region between two amino acid sequences and less than 50% of homology).

Alignment studies of their correspondent amino acid sequences with protein Blast software (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) were performed (Table III-1), the analysis showed that some of these putative proteins have a good percentage of sequence alignment or coverage which is around 90% (LPG1 and LPG1L; LPG1 and LPG1R), and some other have only 48% of alignment (LPG1L and LPG1G). However, the four proteins share low identity that corresponds to the degree to which two amino acid sequences have the same residues at the same positions in an alignment. This could suggest that they have different activities or use different substrates. Moreover, when we compared these four amino acid sequences with other galactofuranosyltransferases such as GlT1, GlT2 from *Mycobacterium tuberculosis*,<sup>38</sup> GfsA from *Aspergillus fumigatus* and WbbI from *E. coli* -K12,<sup>39,40</sup> no significant homology of sequences could be detected. Indeed, their coverage percentage is below 50%, so it is not significant. This suggests that the available structural data of GlT2 could not be used to help building an homology model for LPG1x proteins.<sup>41,42</sup>

A prediction of transmembrane helices with TMHMM Server v. 2.0 allowed to detect a single transmembrane domain present in the *N*-terminal part of putative galactofuranosyltransferases (Figure III-6).



**Figure III- 6:** Transmembrane domain prediction of LPG1, LPG1L, LPG1R and LPG1G. TM: Transmembrane region, predicted from THMM software, are localized in the *N*-terminus of proteins.

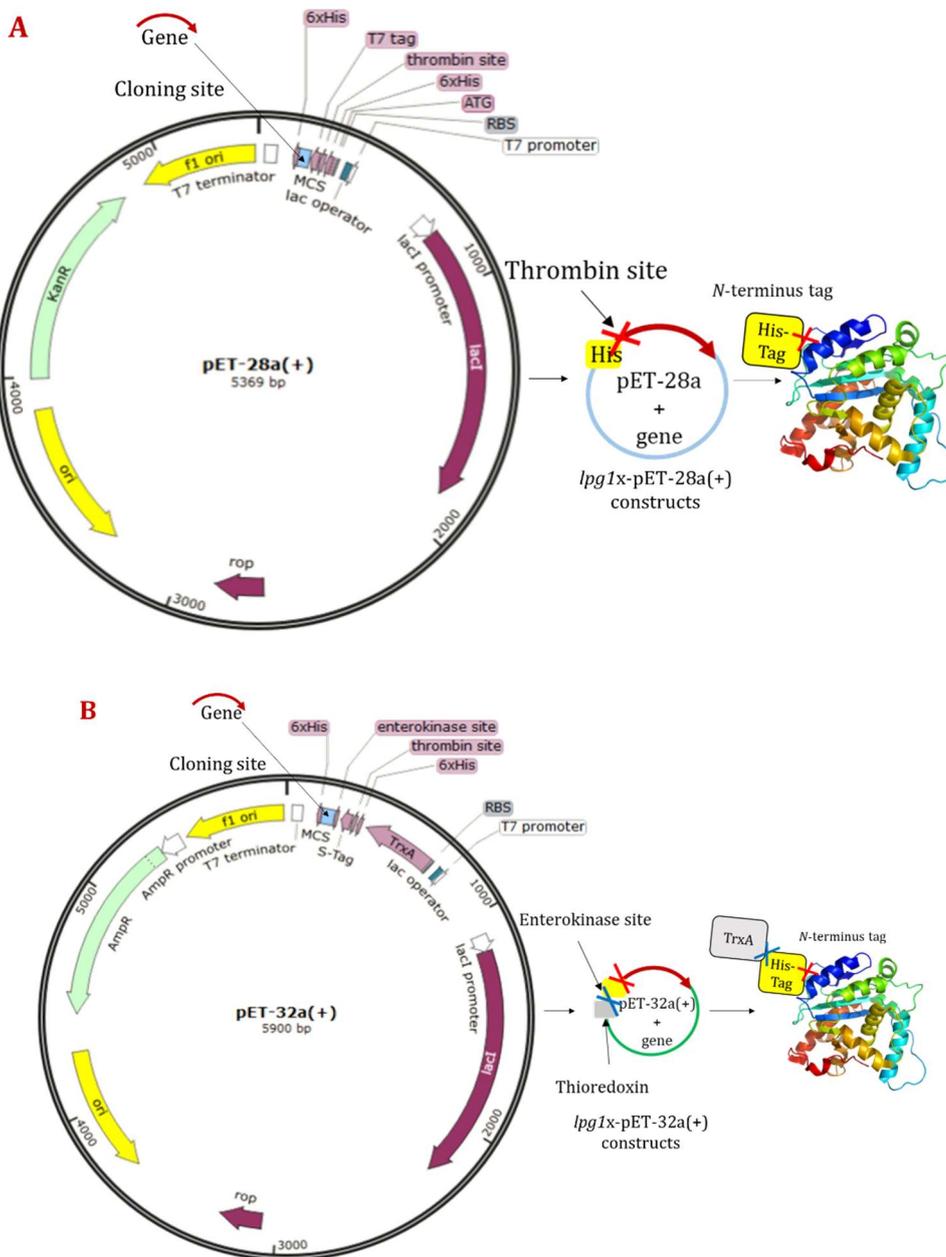
These observations were previously reported by Zhang et al.<sup>4</sup> It is important to highlight the fact that transmembrane domains are localized only in *N*-terminus of proteins. Moreover, they are distant from their respective catalytic site. This is considered as a major piece of information for the establishment of a good purification strategy. As it was mentioned before, transmembrane proteins are hard to produce.<sup>43</sup> In this case, we can adopt a strategy based on the removal of transmembrane domain and hope that it will not affect the folding and activity of our proteins, but facilitate the extraction of LPG1x proteins.

## B. LPG1x cloning, overexpression and purification

### 1. pET vectors as heterologous expression system

For this project, we decided to test first two plasmids, pET-28a(+) and pET-32a(+) that will allow the obtention of recombinant proteins fused to histidine tag in their *N*-terminus. This tag consists of six consecutive histidine residues.<sup>28,44</sup> Here, the two vectors were used in order to find the appropriate plasmid to produce soluble and pure proteins with a good yield. We choose the pET-28a(+) plasmid because it is used in routine in our lab and provided hundreds of milligrams of pure and active proteins in other projects. And the particularity of pET-32a(+) is the presence of

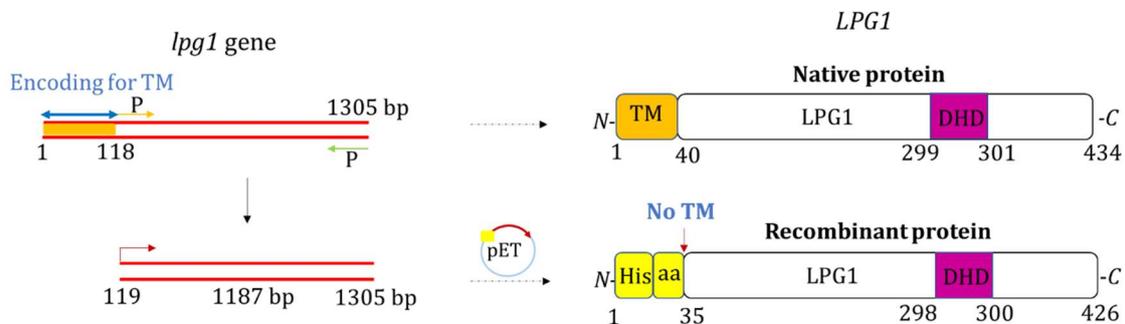
109aa Trx-tag thioredoxin protein and cleavable His-Tag. Trx•Tag™ is a peptide sequence grafted onto the recombinant protein which could be removed by enterokinase.<sup>45</sup> It is known as a solubilization tag, which could assist the proper folding in proteins and avoid their precipitating.<sup>46</sup> In both plasmids, there is thrombin site for the removal of poly-histidine tag.<sup>47</sup> The map of pET-28a(+) and pET-32a(+) and their respective expected recombinant proteins are described in Figure III-7.



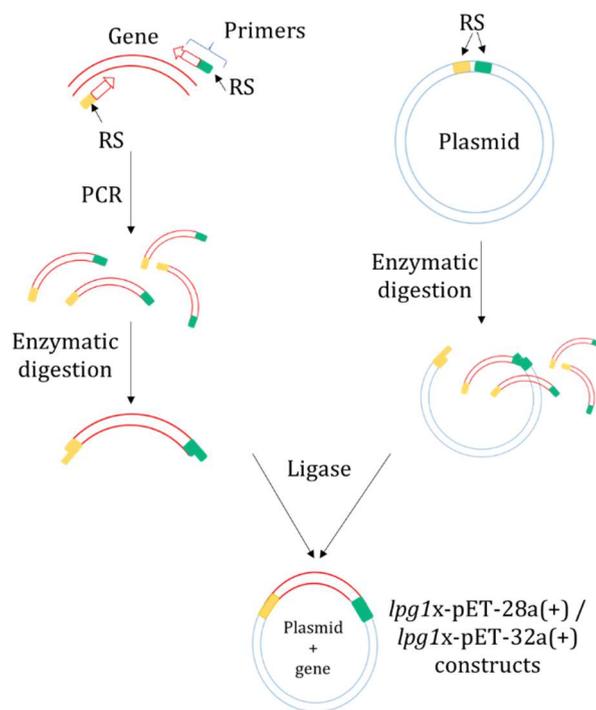
**Figure III- 7:** pET Scheme representing cloning product obtained with pET-28a(+) and pET-32a(+) and their respective recombinant proteins fused with histidine tag. A: pET-28a(+), B: pET-32a(+).

2. Cloning

In order to facilitate purifications of our proteins, nucleotides that encode for transmembrane domains were removed (Figure III-8). Basically, the four genes *lpg1*, *lpg1L*, *lpg1R* and *lpg1G* were inserted in the same manner to both pET-28a(+) and pET-32a(+) vectors thanks to a classical cloning method that consists on amplification of gene of interest by polymerase chain reaction (PCR) from the *L. major* genome using specific primers produced by Eurofins Genomics (<http://www.eurofinsgenomics.eu/>). Then, PCR products were digested by appropriate enzymes. Finally, the digested amplicons are inserted into the expression vector by ligase (Figure III-9). Final constructs *lpg1x*-pET-28a(+) and *lpg1x*-pET-32a(+), were analyzed by sequencing by Eurofins Genomics and then transformed into *E. coli*-Rosetta (DE3). The size of modified genes and their respective recombinant proteins are presented in table II-2.



**Figure III- 8:** Scheme representing the strategy used to remove the transmembrane domain from the recombinant proteins, here LPG1 as example. P: Primers.



**Figure III- 9:** Classical cloning principle RS: Restriction site; PCR: Polymerase Chain Reaction

	 Native gene (bp)	 Removed TM (bp)	 Native protein (AA)	 R.P (N-His) (AA)	 R.P (N-His) (kDa) pET-28a(+)	 R.P (N-His) (kDa) pET-32a(+)
<i>lpg1</i>	1305	118	433	426	50.072	64.230
<i>lpg1L</i>	1617	121	592	580	64.324	78.530
<i>lpg1R</i>	1383	121	460	454	50.918	65.078
<i>lpg1G</i>	1800	139	599	589	66.207	80.458

**Table III- 2:** Size of native genes, the removed transmembrane regions and the molecular weight of native and recombinant proteins without TM (AA): amino acid, (RP): recombinant proteins, bp: base pairs

### 3. Overexpression of LPG1x

#### a) Influence of the temperature and the induction time

As mentioned before, inducing the expression at high temperature can lead to the production of not correctly folded proteins that will aggregate and accumulate in the cytoplasm as inclusion bodies.<sup>48</sup> Some protocols have been developed to extract the proteins from the inclusion bodies. However, they are based on denaturation and refolding steps. In our case, leishmanial GalT's are unknown, consequently, we cannot predict the right conditions for their refolding. It is uncertain to retain their appropriate structure and biological activity. Then, it is better, to avoid the formation of inclusion bodies by finding the best conditions to produce soluble recombinant proteins. Moreover, extracting good yield of pure soluble proteins is closely related to the solubility of proteins.

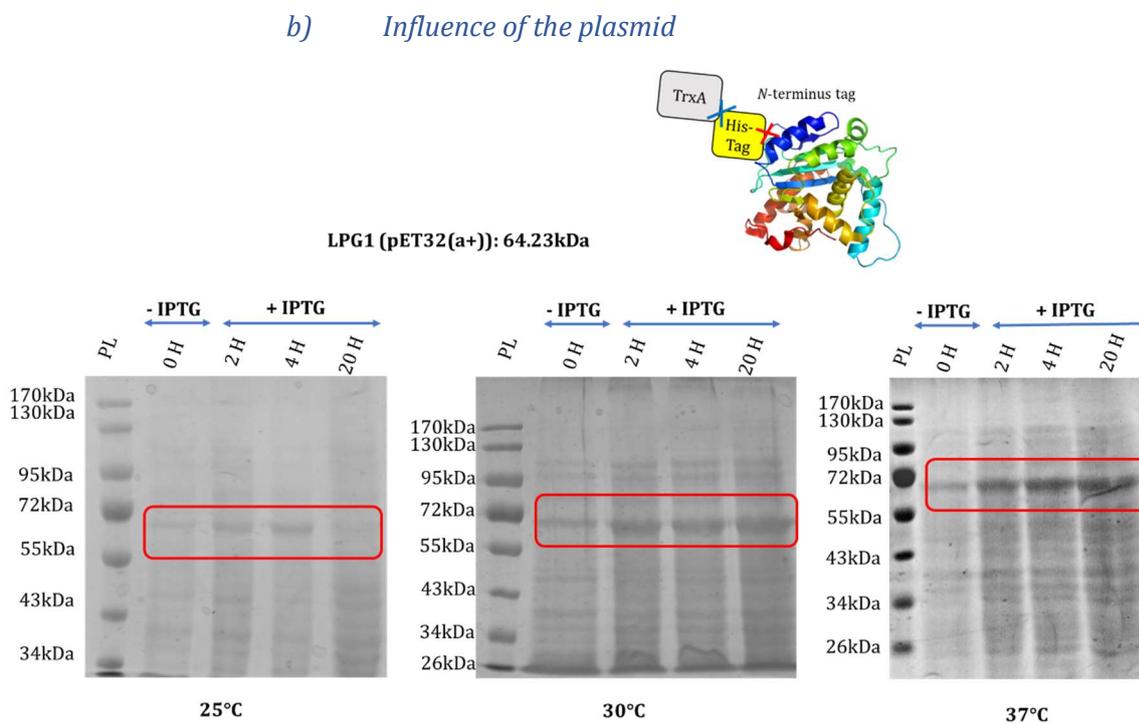
To do that, we decided to evaluate the growth rate of the cells by inducing the expression of recombinant proteins for few hours; 0, 2, 4 and 20 and at different temperature ranging between 25°C and 37°C. Samples were taken from each condition, and were analysed on SDS-PAGE 12%.



and 55kDa. They correspond then to overexpression of LPG1. Indeed, once the IPTG is added, the repressor is inactivated, and the transcription of *lpg1* gene is activated. The maximum of LPG1 overexpression is observed after 20 hours. However, there is no big differences in expression after 2 or 20 hours. Similar bands are observed in the different gels representing the induction at different temperature 25, 30 or 37°C. Same profile was obtained, demonstrating that inducing at 25°C is sufficient to reach the maximum of the overexpression.

Similar results were obtained with the other *lpg1x*-pET-28a(+) constructs (Appendix 1).

Based on this observation, according to Teresa San-Miguel et al.,<sup>50</sup> it is better to induce at lower temperature to avoid the formation of inclusion bodies, we decided to induce the culture at 25°C for 20 hours.



**Figure III-11:** SDS-PAGE (8%) analysis of the overexpression of LPG1 (pET-32(a+)) under different conditions of the induction; 2, 4 and 20 hours of induction after the addition of IPTG in the medium at 25, 30 and 37°C. PL: Protein Ladder.

Similar experiments were performed with *E. coli*-Rosetta (DE3) that were transformed with *lpg1*-pET-32a(+) constructs. Figure III-11 presents the analysis of the overexpression of LPG1 at 25, 30, and 37°C induced for 2, 4 and 20 hours. In the absence of IPTG, and at 25°C, there is a barely visible band below 72kDa, that could correspond to the expected proteins of interest, LPG1 (pET-32a(+)) with the expected molecular weight of 64.32kDa. However, at 30°C and 37°C, there is a ticker band, which is more visible than at 25°C. Thus, the leakiness of the promoter is enhanced at higher temperature. The overexpression at 25°C, in the presence of IPTG, either at 2, 4 or 20 hours, thin bands under 72kDa are visible. This indicates low expression of LPG1 at 25°C even after 20

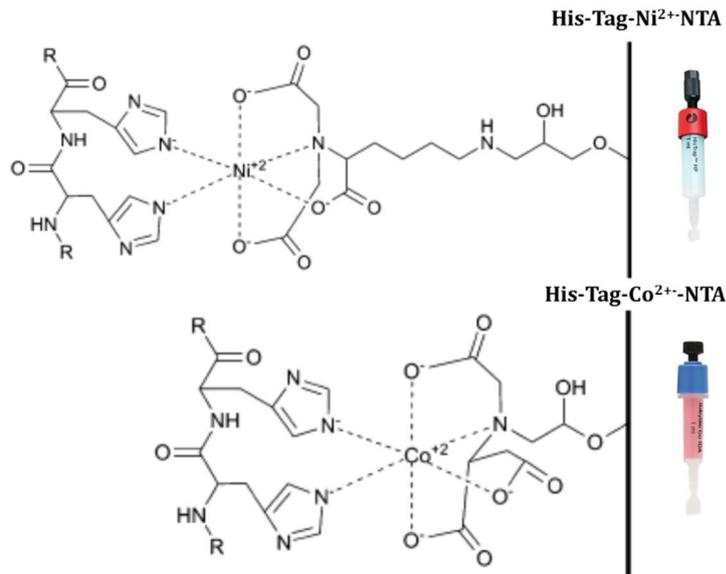
hours of induction with IPTG. The induction of the expression at 30°C is characterized by the presence of thicker bands under 72kDa that corresponds to LPG1 (64.32kDa). However, there is no difference between the different samples that were taken at 2, 4 or 20 hours of induction. This means that the overexpression of LPG1 reached its maximum at 2 hours at 37°C. We observed similar results than the one observed at 30°C. It is thus better to induce the culture at 30°C for 2 hours. Since the temperature was increased to 30°C, compared to 25°C for LPG1 (pET-28a(+)), we decreased the period of the induction to avoid the toxicity effect and the formation of inclusion bodies.

Similar results were obtained with the other *lpg1x*-pET-32a-1H constructs (Appendix 2).

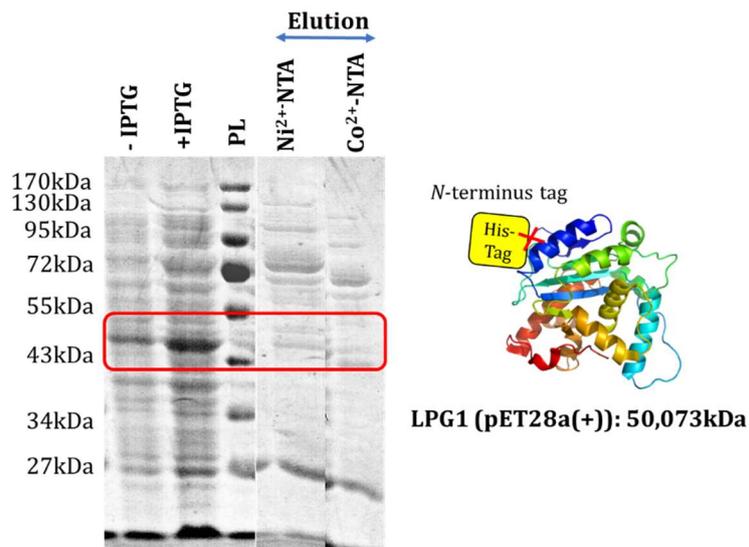
As a conclusion, the overexpression of LPG1x proteins depends on the nature of construct. Indeed, using the *lpg1x*-pET-32a(+) constructs resulted in lower overexpression than *lpg1x*-pET-28a(+). To explain this difference, it is possible that LPG1x proteins fused with Trx-tag thioredoxin protein and poly-histidine tag are toxic to the cells. This could lead to the instability of the plasmid<sup>14</sup> *lpg1x*-pET-32a(+), which can also lead to the death of bacteria. LPG1x proteins extraction and purification

### c) First strategy of purification

After the induction step, bacteria transformed with *lpg1x*-pET-28(a+) and *lpg1x*-pET-32a(+) were harvested and lysed. At the beginning, we used a classical protocol regularly used in the lab. It is based on a combination between different lysis methods in order to enhance the extraction of proteins of interest. Cells were first incubated with lysozyme that degrades bacterial membrane. Since *E. coli* are Gram negative bacteria, they are resistant to the activity of lysozyme. Thus, this step is not sufficient to release the recombinant proteins. Consequently, freeze-thaw procedure was also performed followed by sonication step. Here freeze-thaw helps breaking the membrane cells. But, the sonication, which is a high-energy mechanical disruption method, is more effective.<sup>51</sup> At the end, released proteins were purified either on Ni-NTA or Co-NTA columns (Figure III-12). We decided to test two different columns because according to Joshua A. Bornhorst et al., using cobalt column provides more efficient His-tag-protein purification than Ni-NTA column.<sup>52</sup> Indeed, cobalt exhibited less contaminant protein binding than Ni-NTA due to its weak interactions. Proteins were finally eluted with free imidazole.



**Figure III- 12:** Interaction between the poly-histidine tag fused to the recombinant protein and two different immobilized metal affinity chromatography (IMAC) Ni-NTA and Co-NTA dapted from Julie Credou et al., 2014.<sup>53</sup>



**Figure III- 13:** Analysis of the overexpression of of LPG1(pET- pET28a(+)) and its purification of LPG1(pET- pET28a(+)) on Ni-NTA and Co-NTA column on SDS-PAGE 12%. PL: Protein Ladder.

Figure III-13 presents the analysis of the samples that were taken before and after the induction of LPG1 (pET-28a(+)). The induction was monitored at 25°C for 20 hours. Without IPTG, there is a band around 50kDa that corresponds to LPG1. This band represents the basal expression level. After the addition of IPTG, a thick band is visible, it corresponds to the overexpression of LPG1. The elution sample of purified LPG1 on Ni-NTA column contain many bands that correspond to non-specific proteins. A small thin band is visible, showed int the red box, could correspond to LPG1. Similar results are observed on the elution sample of cobalt purification. At this level,

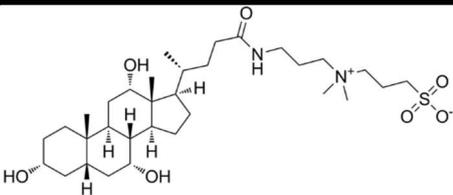
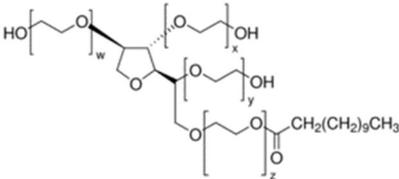
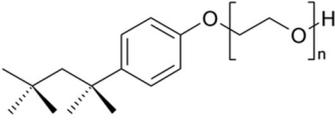
neither Ni-NTA or Co<sup>2+</sup> column allowed the purification of LPG. Moreover, there is no significant difference between the two LPG1-1H elution samples.

Two hypotheses can be proposed: (i) the lysis protocol was not efficient to release the recombinant proteins. Or (ii) the desired proteins are not soluble. Since we used a strong protocol containing three successive lysis methods, the second hypothesis might be favoured.

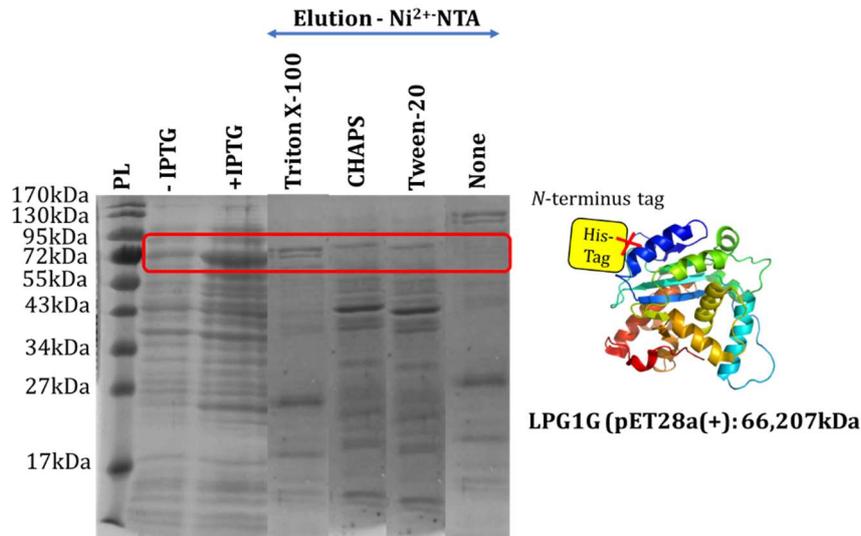
Same experiments were performed with the other putative GalTs in pET-28a(+) or pET-32a(+) vectors and similar results were obtained (Appendix 3). None of LPG1x proteins was correctly purified. However, some cases displayed less bands in the elution sample of cobalt purification step. This confirmed that a cobalt column retains less unwanted proteins. Since there is not much recombinant protein in the elution samples, we decided to use Ni-NTA column for the rest of the experiments.

#### d) Detergent effect on protein solubility

Since the four proteins displayed similar results, we decided to focus on only one protein, LPG1G (PET-28a(+)). In order to decrease the risk of toxicity and aggregation of protein, cultures were induced at 25°C for only 2 hours. And, to help solubilizing the recombinant protein, different detergents were tested. It is important to choose the suitable detergent that will not significantly inactivate the function of the protein of interest.<sup>54,55</sup> Among these detergents, there is Triton X-100 and CHAPS. The Triton X-100, is a non-ionic detergent and CHAPS is a zwitterionic one. We also used another non-ionic detergent, tween 20. (Table III-3). Same previous protocol was used to lyse and purify LPG1G (pET-28(a+)). The only modification consists on the use of the detergent.

	Structures	Type
CHAPS		Zwitterionic
Tween-20		Non-ionic
Triton X-100		Non-ionic

**Table III- 3:** Chemical structures of different used detergents



**Figure III- 14:** SDS-PAGE (12%) analysis of overexpression of LPG1G (pET-28a(+)) proteins after the induction at 25°C for 2 hours, and its purification on Ni-NTA column. Three different detergents were tested solubilize the recombinant proteins. PL: Protein Ladder.

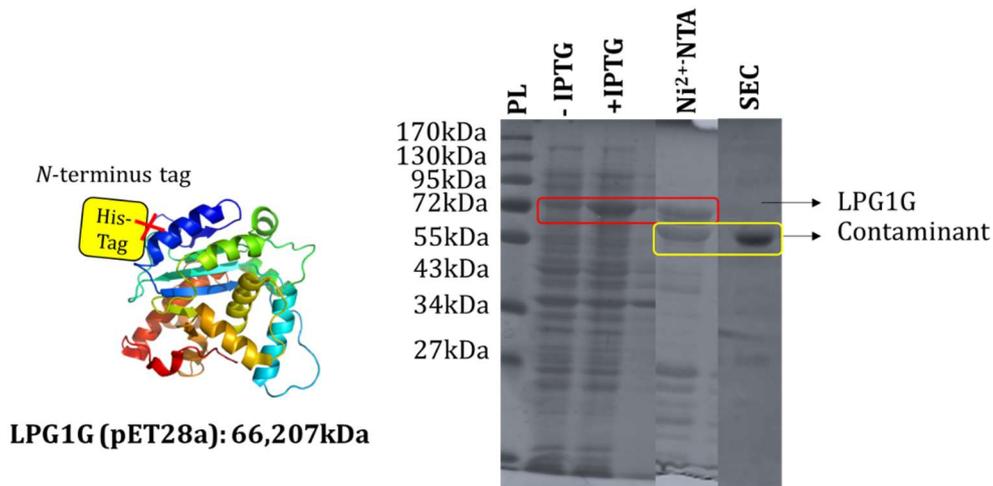
The four experiments were carried out, and elution samples of purification steps were analyzed on SDS-PAGE 12%, as showed in figure III-14. Samples before and after the induction with IPTG were also added on the same gel. In the absence of IPTG, a thin band was visible that could correspond to the expected LPG1G with 66.207kDa. This indicated the basal expression level. However, after 2 hours of incubation at 25°C following the addition of IPTG, one tick band around 70kDa was visible. This confirmed the overexpression of LPG1G.

The analysis of first experiment elution sample shows a thin band, which migrates in the same level than the one observed after the addition of IPTG. Consequently, this could correspond to LPG1G. The elution samples of the 2<sup>nd</sup> and 3<sup>rd</sup> experiments are also characterized by the presence of very thin similar bands around 70kDa that could also correspond to LPG1G.

The use of theses different detergent did not enhance the solubility and purity of the protein of interest.

#### e) *Alternative lysis protocol*

A new protocol that was cited in the Molecular Cloning Laboratory Manual edited in 1989 was tested. The protocol consists of incubation of the cells for 4 hours with 1mg/ml final concentration of lysozyme instead of 0.1µg/ml used in the previous protocol. Here, the buffer contains 5% glycerol and 1mM DTT. DTT is used primarily to reduce protein disulphide bridges, and more generally to prevent the formation of intramolecular and intermolecular disulphide bridges. However, DTT cannot reduce buried disulphide bridges (inaccessible to solvent) unless denaturing conditions are used. The stability of proteins in aqueous solution is routinely enhanced by cosolvents such as glycerol.<sup>28</sup> Samples were loaded on SDS-PAGE (12%).



**Figure III- 15:** SDS-PAGE (12%) analysis of overexpression of LPG1G (pET-28(a+)) proteins after the induction at 25°C for 2 hours, and its purification on Ni-NTA column followed by size exclusion chromatography (SEC). PL: Protein Ladder.

Two liters of *E. coli*-Rosetta (DE3) that were transformed with *lpg1G*-pET-28a(+) were induced at 25°C during 2 hours. Then, cells were harvested and lysed with the second strategy described above. Samples before and after the induction with IPTG were deposited on SDS-PAGE 12%. Figure III-15 shows the presence of thin band around 70kDa before the incubation with IPTG, and the presence of thick band after two hours of induction. This confirmed the overexpression of LPG1G. The elution sample resulting of the purification with Ni-NITA column. Two thick bands are visible between 55 and 72kDa. The highest band, in the red box, migrated at the same level the band corresponding to the overexpression of LPG1G. This band could correspond than to the protein of interest, LPG1G. However, the second band, in the yellow box, which seems also highly expressed still unknown. Size exclusion chromatography (SEC) assay was performed in order to separate proteins according to their size. Only the contaminant was obtained but, LPG1G was not detected.

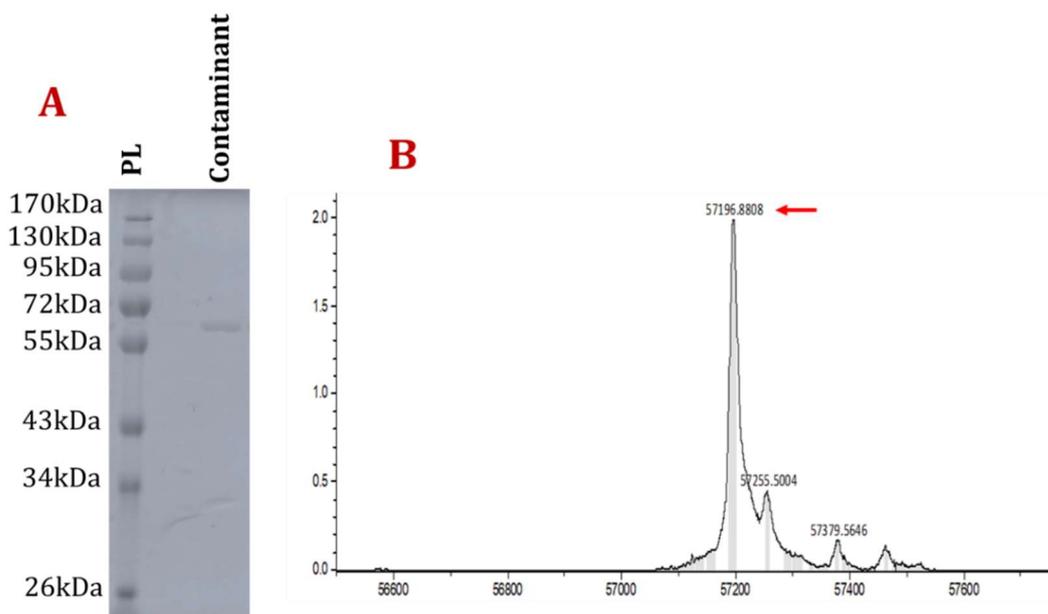
The alternative strategy allowed the obtention of LPG1G. However, the desired protein was not pure, it was contaminated with unwanted protein that migrates between 55 to 72kDa. Size exclusion chromatography was used to get rid of the contaminant, but, it was not successful. In addition, LPG1G was not visible on SDS-PAGE, it was probably too diluted to be detected. Despite the loss of the protein of interest, this strategy enhanced the protein purification and the extraction steps. This could be explained by the increase of lysozyme amount in the buffer that intensified bacteria membrane degradation. Moreover, the addition of glycerol in lysis buffer enhanced the solubility of the protein of interest. In addition, the presence of DTT helped stopping the oxidation of the proteins, and avoided the formation of unwanted disulfide bonds.<sup>12</sup> However, the presence of DTT around LPG1G could alter the binding of recombinant protein on Ni-NTA column. But, the recommended concentration of DTT is 1mM which is the case in this protocol.

Other glycosyltransferases were also successfully extracted using 1mM DTT containing buffers.<sup>56,57</sup>

Similar results were obtained with LPG1 (pET -28a(+)), LPG1L (pET -28a(+)) and LPG1R (pET -28a(+)). However, trials with some *LPG1x* (pET-32(a+)) were not successful. (Appendix 4).

#### 4. Identification of the contaminant

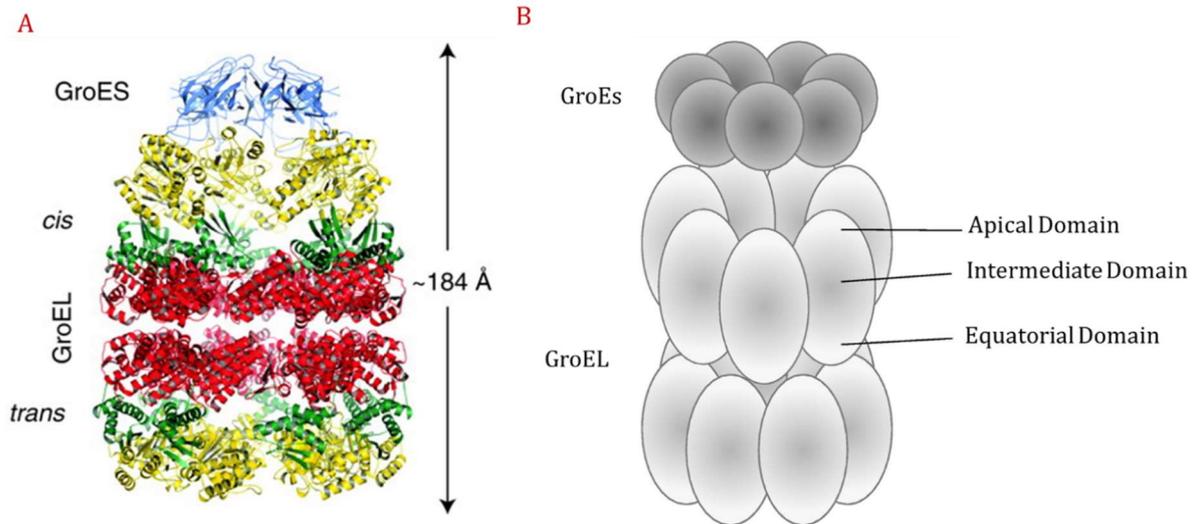
Pure sample of the contaminant were first analysed by mass spectrometry in order to obtain its exact molecular weight. Then, in gel-tryptic digestion of contaminant isolated by gel electrophoresis was performed. The protein was enzymatically cleaved into a limited number of shorter fragments by trypsin. This enzyme is an endopeptidase that hydrolysis the protein at the carboxyl site of specific amino acids; *i.e.* lysine and arginine. Each peptide fragment is analysed by mass spectrometry, this method is called peptide mass fingerprint.<sup>58</sup> The mass of peptides was compared with proteins data base using MASCOT server (<http://www.matrixscience.com>). At the end, the information is stitched together to divulge the identity of the protein.



**Figure III-16:** Electrophoresis of contaminant on SDS-PAGE (12%). B-mass spectrometry analysis of pure contaminant. PL: Protein Ladder

Results showed that the molecular weight of pure contaminant is 57kDa (Figure III-16). Peptides analysis revealed 60% of coverage sequences with bacterial chaperonin GroEL. This protein is composed of two 57kDa subunits which correspond with the estimated molecular weight. Contamination with chaperonin is a common phenomenon<sup>29</sup>. GroEL is one the best studied bacterial chaperonin. <sup>59</sup> It belongs to the class of protein, Hsp60 (Heat shock proteins). This chaperonin is known to be involved as accessory protein in order to ensure the correct folding of “denaturated” protein.<sup>60</sup> Its crystal structure was identified. GroEL, is composed of 14 identical

subunits, which are arranged as two heptameric rings stacked back to back. A large central cavity binds and folds the substrate protein. Each subunit contains three domains, apical, intermediate and equatorial.<sup>59</sup> GroEL cooperate with GroES (~10kDa) for the folding of the proteins. The co-chaperonin GroEs is composed of a single heptameric ring. It interacts with GroEL ring and acts as a cap on the central cavity (Figure III-17).<sup>61</sup>



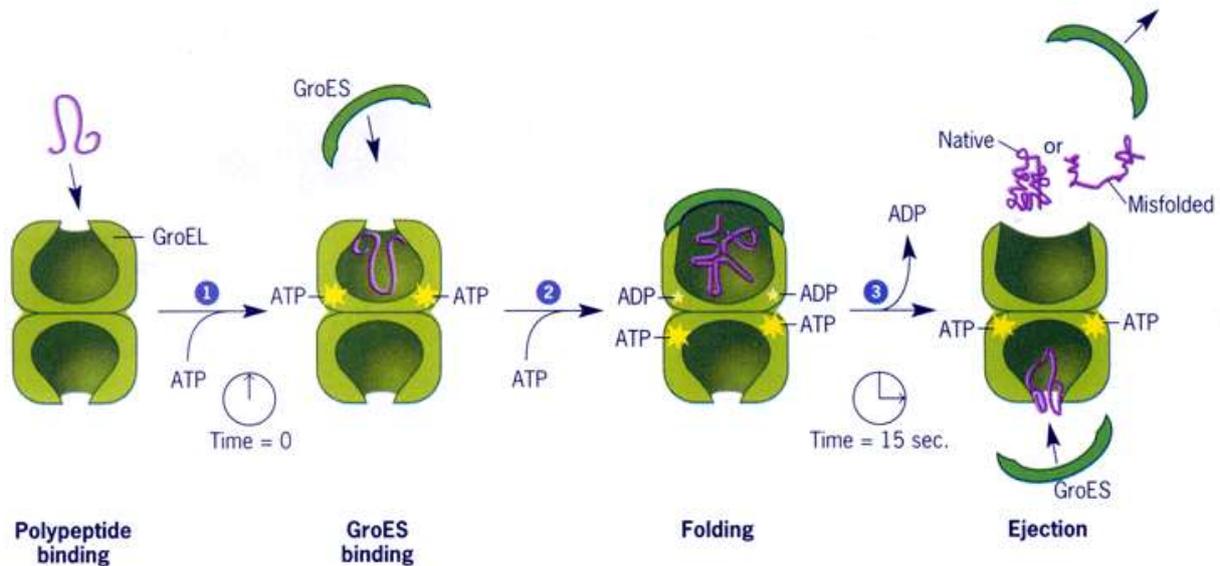
**Figure III-17:** GroEl-GroEs complex. (A): The Crystal structure of GroEl-GroEs complex (PDB 1AON) adapted from R. Martin Vabulas et al.,2018.<sup>62</sup> Cis, the GroES-bound chamber of GroEL. Trans, the opposite GroEL ring. (B): Schematic representation of the complex, showing the two rings of GroEL composed of the seven subunits with their three domains, and the co-chaperonin GroES that caps one end of the end of GroEL. Adapted from Gabriel M. Altschuler et al., 2008.<sup>63</sup>

The presence of chaperonin GroEL indicates then that the protein of interest, LPG1G is not correctly folded, and need then the complex GroEL-GroES. In this case, we suppose that the induction and the overexpression of the proteins are stressful conditions for the bacteria. Such observation were also reported in many articles, highlighting the effect of induction conditions on the co-purification of recombinant proteins with chaperonin.<sup>16,34,35</sup>

It will be better to remove these chaperonins, and obtain pure LPG1x proteins their enzymatic and structural characterization.

## 5. Removal of chaperonin GroEL

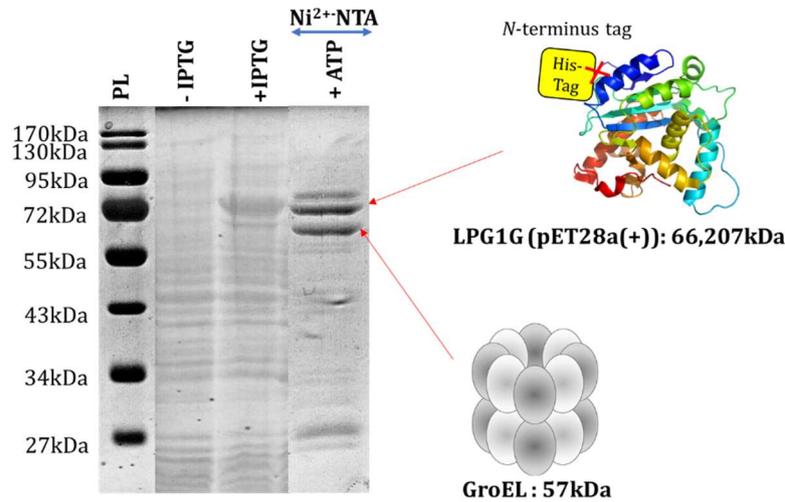
## a) ATPase activity



**Figure III-18:** Protein folding cycle of *Escherichia coli* complex GroEL-GroES.

The complex GroEL-GroES uses ATP hydrolysis to promote the folding of newly-synthesized proteins.<sup>64</sup> Only single substrate protein up to approximately 60 kDa can be encapsulated in the compartment of the complex.<sup>62</sup> First GroEL captures the non-folded substrate protein thanks to its hydrophobic amino acids of the apical domain. Then ATP molecules attach the equatorial domain of the ring that loaded the protein. This results in conformational modification in GroEL. Consequently, GroES is able to bind and cap the apical domain of GroEL. This leads to the release of the non-folded protein into the cavity of GroEL. The complex GroEL-GroES is in its “cis” active conformation. During this state, the ATP is hydrolyzed within 10 to 15 seconds. During this short period, the folding of the protein substrate occurs. The hydrolysis reaction leads to the binding of a second ATP molecule to the non-occupied ring of the GroEL.<sup>65</sup> This induces the release of GroES and the substrate protein into the cell. The non-yet folded protein substrate can rebind to GroEL for further folding trials.<sup>66</sup>

Based on this, Raji E. Joseph and Amy H. Andreotti used an ATP-containing buffer to purify their protein.<sup>67</sup> The presence of ATP is supposed to induce the release of the target protein. We decided then to use the same procedure.



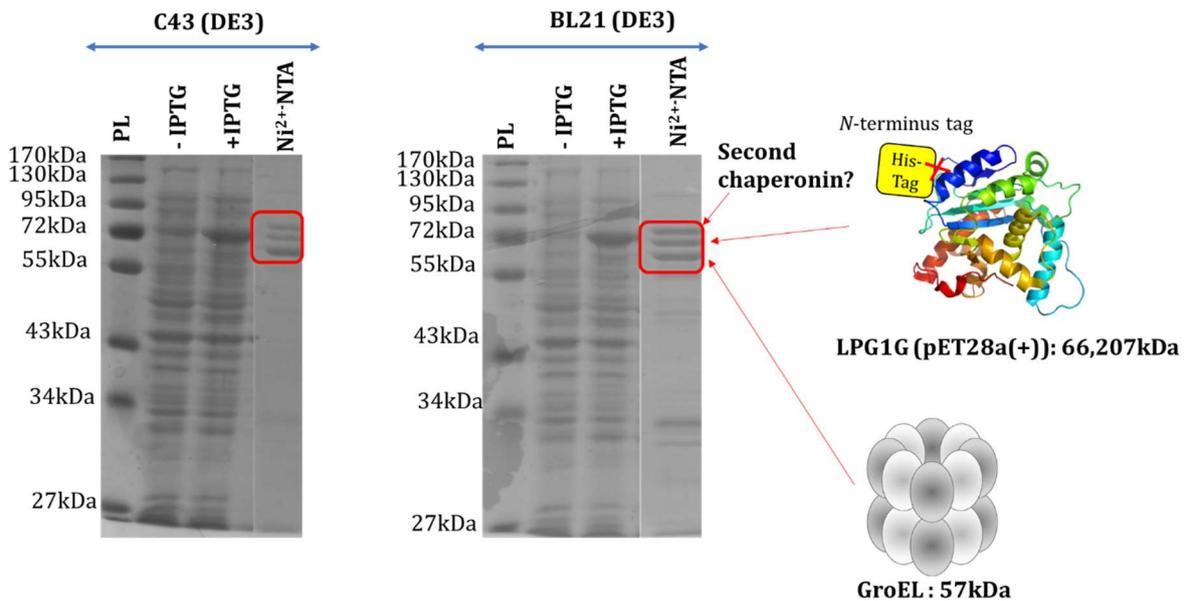
**Figure III-14:** SDS-PAGE (12%) analysis of overexpression of LPG1G (pET-28a(+)) proteins after the induction at 25°C for 2 hours, and its purification after lysis strategy 2 on Ni-NTA column with ATP. PL: Protein Ladder.

The analysis of the samples before and after the addition of IPTG on SDS-PAGE (12%) indicates the overexpression of LPG1G (pET-28a(+)) in Figure III-19. Elution sample that was taken after the purification of LPG1G-1H with ATP-containing buffer was also loaded on SDS-PAGE. Similar bands to those described in the previous purification experiments were observed. The highest one corresponds to LPG1G, and the one below represents the GroEL chaperonin. Here again, the wanted protein, LPG1G-1H, is not pure. The addition of ATP seems not to be efficient enough to remove the complex GroEL-GroES. The conditions are probably not optimum to induce the interaction between ATP molecules and the chaperonin.

#### b) Other *E. coli* strains

It was reported by Dumon-Seignovert *et al*, that using different strains of *E. coli* such as C43 (DE3) can overcome toxicity related to the production of recombinant proteins. C43 (DE3) is a mutant strain from BL21(DE3). C43 strains were discovered by Miroux and Walker in 1996 and are able to better the overexpression of proteins thanks to their lacUV5 promoter. It displays two mutations on its -10 region compared to the lac promoter. This leads to less level of synthesis of T7 RNA polymerase. Thus the overproduction of recombinant proteins is more tolerated by this strain.<sup>14,68</sup>

We decided to transform *lpg1G*-pET-28a(+) into two new *E. coli* strains; i.e. C43(DE3) and BL21(DE3).



**Figure III-20:** SDS-PAGE (12%) analysis of overexpression of LPG1G (pET-28a(+)) proteins after the induction at 25°C for 2 hours using C43 (DE3) and BL21 (DE3) *E. coli* strains, and its purification after lysis strategy 2 on Ni-NTA column. PL: Protein Ladder.

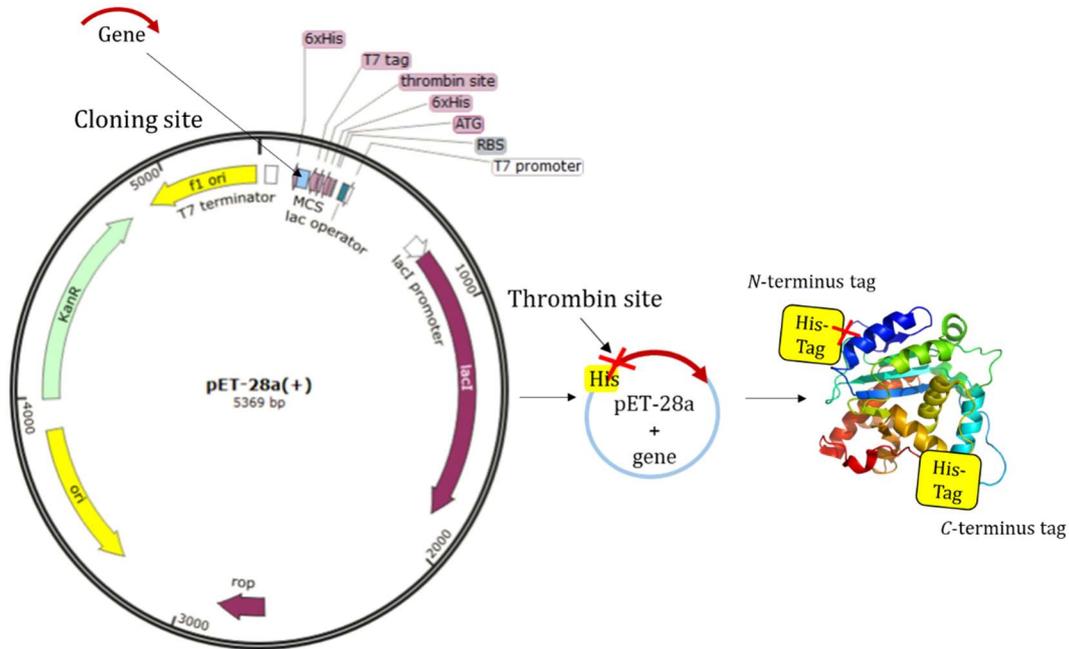
The Figure III-20 presents the results of the overexpression and purification of LPG1G (pET-28a(+)) proteins that were produced respectively by C43 (DE3) and BL21 (DE3) *E. coli* strains. The analysis of samples of the induction and the purification obtained in both strains present similar profile. Indeed, in both gels, the presence of thick band that corresponds to the proteins of interest (66.207 kDa) after the addition of IPTG confirms the success of protein overexpression. The samples obtained from the elution step of LPG1G-1H on Ni-NTA column are characterized by the presence of intensive bands that correspond to LPG1G, and to the chaperonin GroEL. Surprisingly, third intensive band appeared above LPG1G and GroEL. This could correspond to another chaperonin such as HSP70 (Heat shock protein 70), 70 referring to its molecular weight.<sup>69</sup> Neither C43 (DE3) or BL21 (DE3) prevented the interaction of LPG1G with GroEL-GroEs complex.

### c) C- and N- histidine tags

Since all our previous trials to get rid of the chaperonin failed, we decided to produce recombinant proteins fused with histidine tag in N- and C-termini (Figure III-21) using pET-28a(+) plasmid. The purpose of this experiment was to see if this new form of the protein could prevent the interaction with chaperonin.

Constructs were produced following the same strategy of cloning described before. Specific primers were designed and produced by Eurofins Genomics (<http://www.eurofinsgenomics.eu/>).

Final constructs *lpg1x*-2H-pET-28a(+) showed in Figure III-21 were confirmed by sequencing by Eurofins Genomics.

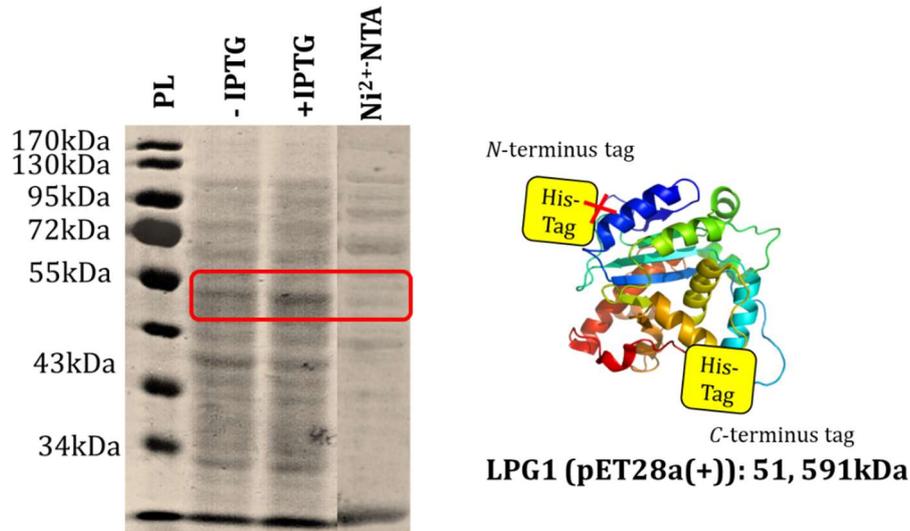


**Figure III-21:** Representation of the strategy of cloning of *lpg1x* genes into pET-28(a+) expression vector. The expressed recombinant proteins should be fused on its both *N*- and *C*- termini with histidine tag.

	N-     -C
	<b>R.P (N- and C-His) (kDa)</b>
	<b>pET-28a</b>
<b>LPG1</b>	51.591
<b>LPG1L</b>	65.602
<b>LPG1R</b>	52.438
<b>LPG1G</b>	67.726

**Table III- 4:** The expected molecular weight of the recombinant proteins (RP) LPG1x fused with His-Tag in both *N*- and *C*- termini.

Two liters of *E. coli*-Rosetta (DE3) that were transformed with the construct *lpg1*-pET-28a-2H, and were induced and at 25°C for 2 hours. Lysis strategy 2 was performed followed by affinity purification using Ni-NTA column. Samples before and after the induction and after the elution steps were taken and loaded on SDS-PAGE 12% (Figure III-22).



**Figure III-22:** SDS-PAGE (12%) analysis of overexpression of LPG1G-2H (pET-28a(+)) proteins after the induction at 25°C for 2 hours, and its purification after lysis alternative strategy on Ni-NTA column. PL: Protein Ladder.

The Figure III-22 shows the results of the production and purification of LPG1-2H (pET-28a(+)). We attempted to overproduce the protein but according to the gel, there is very thin band after the induction with IPTG (red box) that could correspond to the expected protein with ~52.951 kDa. Thus, the overexpression was not efficient. Moreover, the analysis of the elution sample presented many bands corresponding to contaminant proteins and the absence of LPG1G-2H.

Using two histidine tags in both termini of the protein seems not to be appropriate to produce and purify LPG1x proteins.

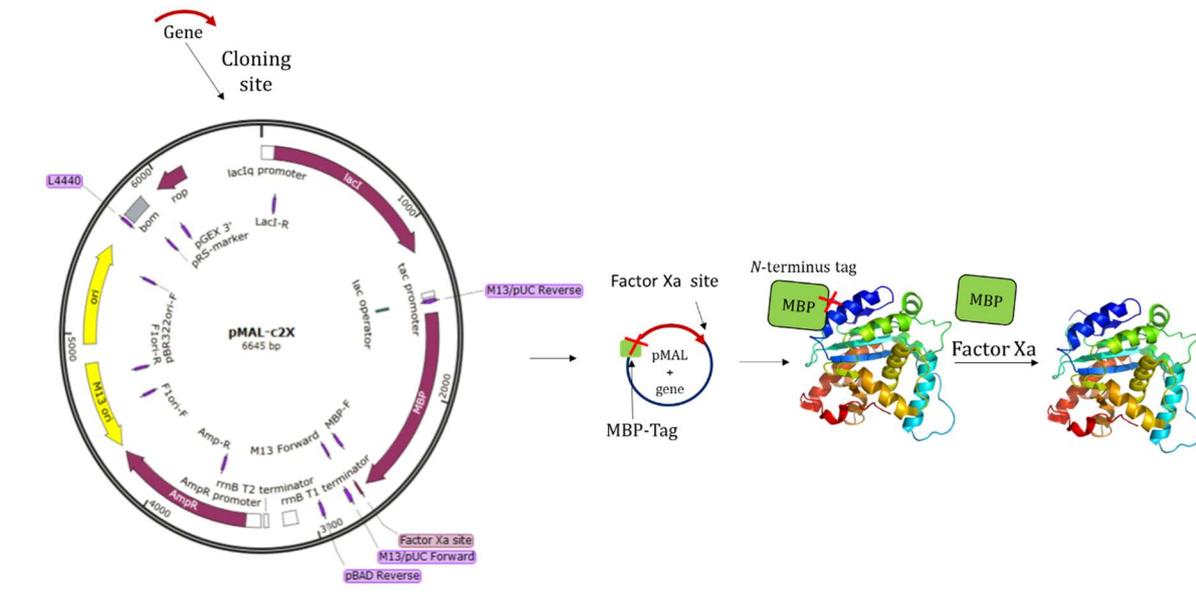
#### 6. pMal vectors as heterologous expression system

All trials to obtain pure LPG1x with pET vectors were not successful. We decided then, to start over again with another expression plasmid: pMal.

The four genes were cloned into new vector pMAL-c2X thanks to classical cloning methods. Constructs were confirmed by sequencing and transformed into *E. coli* – Rosetta (DE3). Using this plasmid will allow the obtention of proteins fused with a very large maltose binding proteins tag (MBP) in their *N*-terminus. Proteins of interest can be cleaved from MBP tag with Factor Xa protease (Figure III-23). The size of the MBP tag is 45kDa, which could affect the folding and the activity of the target protein.<sup>44</sup> Many proteins were expressed using pMAL vector, as example we can cite the bacterial GalT WbbI<sup>39</sup>.

Specific primers were designed and produced by Eurofins Genomics (<http://www.eurofinsgenomics.eu/>). Final constructs *lpg1x*-MBP were confirmed by sequencing

by Eurofins Genomics. The expected molecular weight of LPG1X-MBP recombinant proteins are detailed in table III-5.



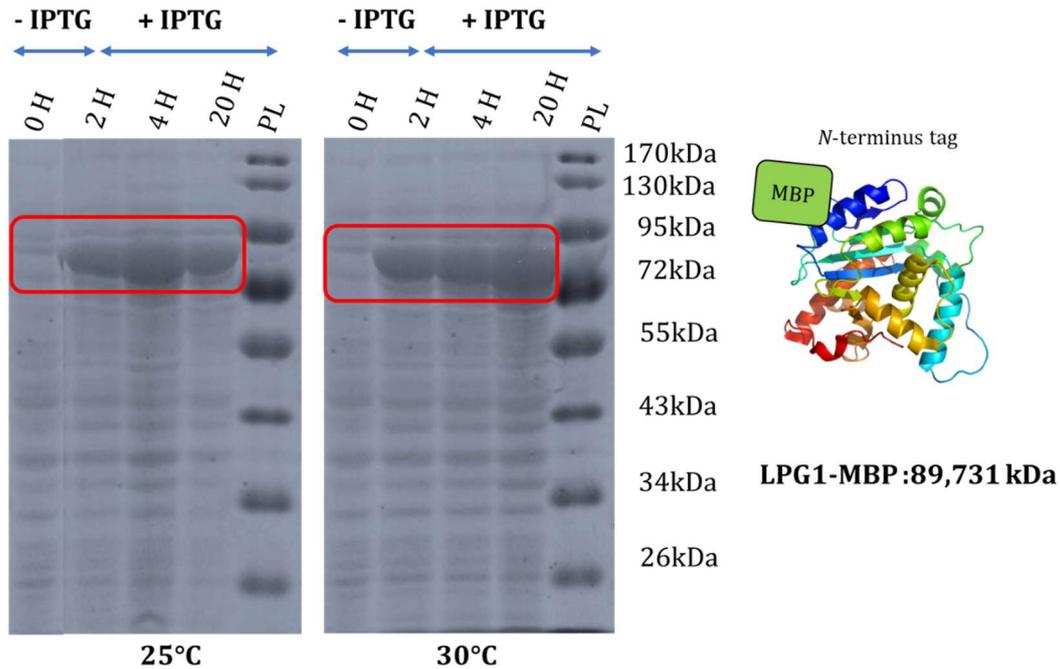
**Figure III-23:** pMAL-C2X Scheme representing Strategy of cloning and the expected recombinant proteins fused with MBP-Tag in *N*-terminus.

	N-   -C
	<b>R.P (N-MBP-Tag) (kDa)</b>
<b>LPG1</b>	89.731
<b>LPG1L</b>	103.983
<b>LPG1R</b>	90.459
<b>LPG1G</b>	105.445

**Table III- 5:** Expected molecular weight of the recombinant proteins (RP) LPG1x fused with MBP-Tag *N*-termini.

*a) Overexpression*

Same procedure described before was followed to evaluate the overexpression of LPG1x-MBP at different temperature (25 and 30°C) and induced for different period (2, 4 and 20 hours). Samples of each condition were loaded on SDS-PAGE 8%.



**Figure III-24:** SDS-PAGE (8%) analysis of the overexpression of LPG1-MBP under different conditions of the induction; 2,4 and 20 hours of induction after the addition of IPTG in the medium at 25 and 30°C. (PL: protein ladder).

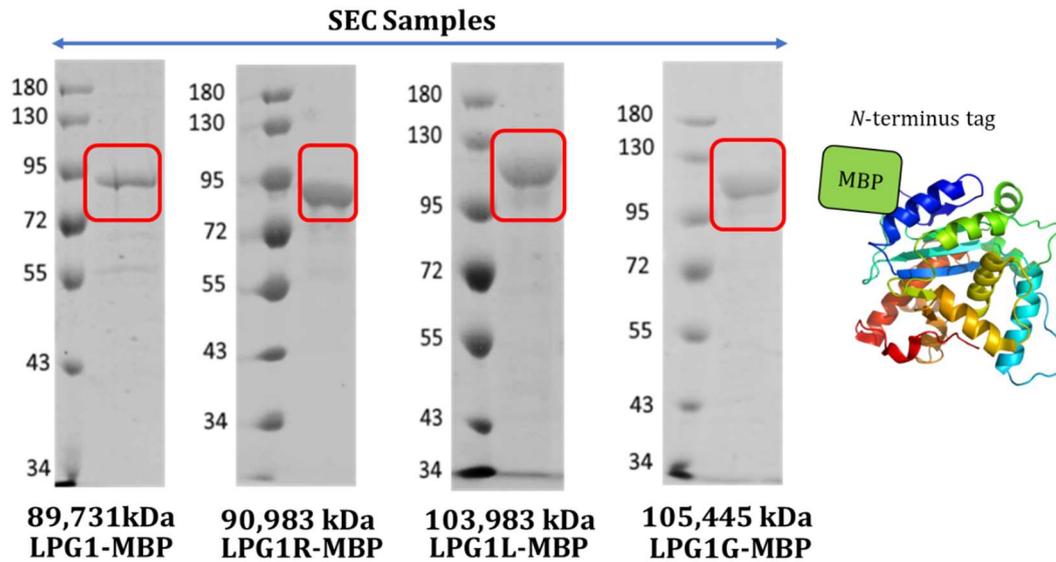
Figure III-24 presents the results of the overexpression of LPG1-MBP protein at 25°C and 30°C after induction with IPTG for two hours. Both gels are characterized by the presence of an intensive band between 92 and 75kDa after the addition of IPTG, while it is absent before the induction. This band corresponds to the overexpression of the expected protein of interest, LPG1-MBP with size of 89.732 kDa.

Similar results were obtained with LPG1L-MBP, LPG1R-MBP and LPG1G-MBP (Appendix 5).

The absence of the bands that represent LPG1-MBP in the sample without IPTG indicates that the absence of leaky promoter phenomena observed with pET vectors. Moreover, the overexpression is high at 25°C and reaches saturation only after 2 hours of induction. Thus, for the future experiments, these conditions will be used to produce LPG1X-MBP proteins.

*b) Purification of LPG1X-MBP*

The initial purification strategy (see pages 103-104) that is based on three successive different lysis methods was followed to extract LPG1X-MBP proteins. The proteins were purified with MBP-column, that contain amylose resin which fixe the MBP-Tag, and targeted proteins are eluted with maltose. Then, size exclusion chromatography was performed.



**Figure III-25:** SDS-PAGE (8%) analysis of the purification of LPG1x-MBP by affinity column followed by SEC. (PL: protein ladder).

Analysis of purified proteins samples on SDS-PAGE 8% (Figure III-25) showed the presence of only one and intensive band in each gel (in the red box). These bands correspond to the expected molecular weight of the four proteins of interest fused with MBP-Tag.

Finally, and after numerous trials and optimization procedures of our production's protein protocols, we obtained pure LPG1, LPG1L, LPG1R and LPG1G fused with MBP-Tag. Thus, the pMAL vector seems to be the most appropriate one to produce LPG1x proteins of *Leishmania*. Proteins of interest are soluble and do not interact with chaperonin. This could suggest that they are well folded and they did not aggregate.

Proteins concentrations were determined by Bradford assays. About 5mg of LPG1 and LPG1R were obtained from 1L of culture, and 10mg of LPG1L and LPG1G per litre, which can be considered as very good yield for this type of enzymes. Obtaining such good yield allowed us to consider crystallogenesis trials, an important step to determine structural configuration of LPG1x proteins, which will be discussed in chapter 4.

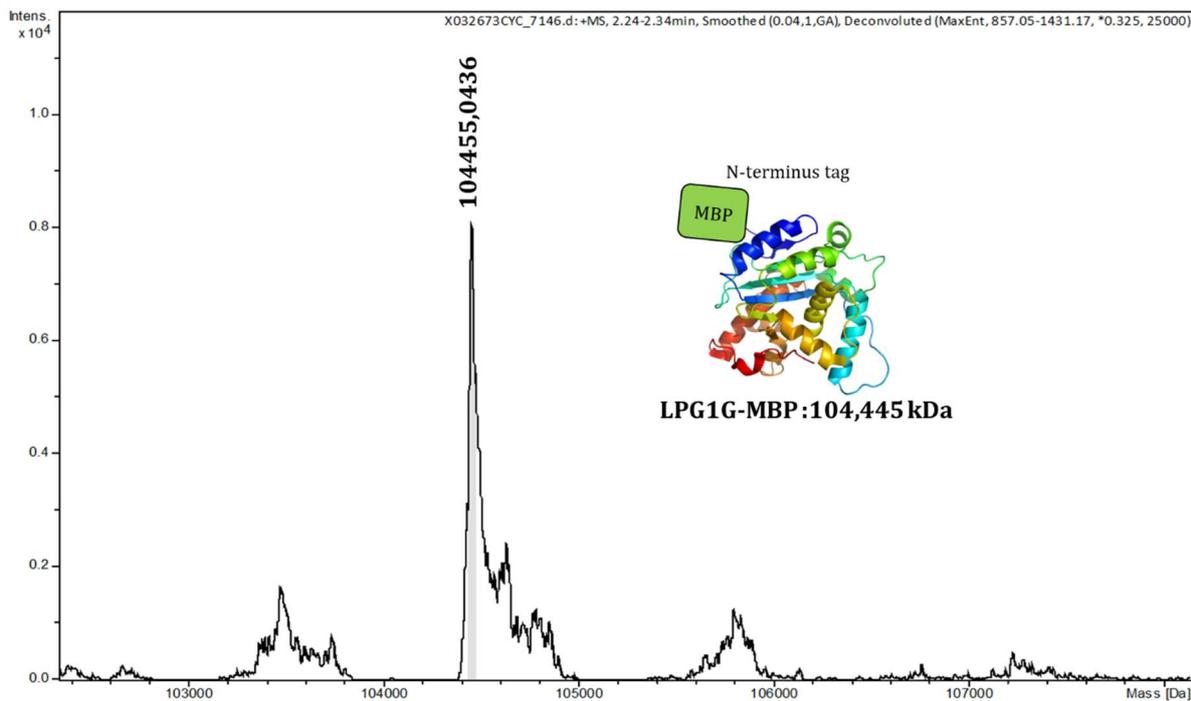
### c) Protein characterization

In-gel digestion of LPG1G-MBP and LPG1R-MBP isolated by gel electrophoresis was performed. Then, they were digested by trypsin and the obtained peptides fragments were analysed by mass spectrometry. The mass of protein fragments were compared with the theoretical peptide mass of digested LPG1G-MBP and LPG1R-MBP.

	LPG1G-MBP	LPG1R-MBP
Sequences coverage %	59	52.3

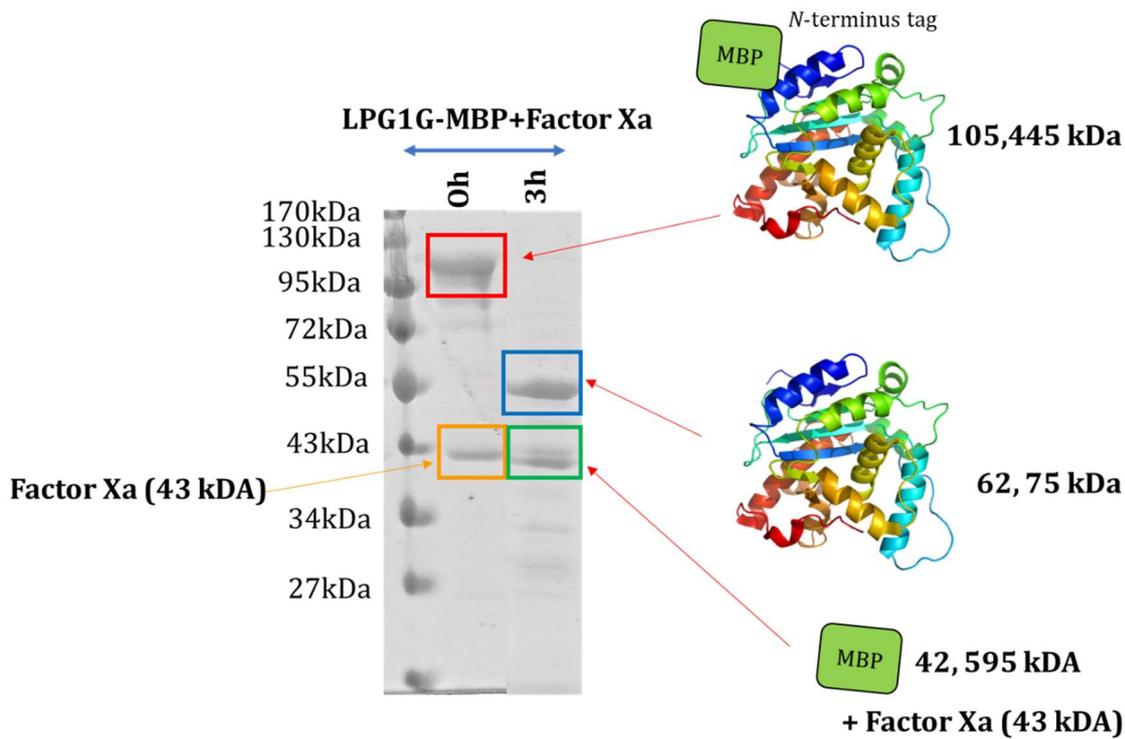
**Table III- 6:** Coverage percent of digested peptide with LPG1G-MBP and LPG1R-MBP amino acid sequence.

Peptides analysis revealed the released peptides from digested LPG1G-MBP and LPG1R-MBP proteins covered 59% and 52.3% of their respective total sequences. This confirms that purified proteins are my proteins of interest. Moreover, small amount of pure LPG1G-BMP was analysed by mass spectrometry at 104.455 kDa. (Figure III-26).



**Figure III- 26:** Mass spectrometry analysis of LPG1G-MBP

d) MBP-Tag removal



**Figure III-27:** Analysis of the removal of MBP-Tag with factor Xa from LPG1G-MBP proteins.

The MBP-Tag is large (42.595 kDa), so it is possible that it could affect the structure of the proteins or its activity. Trials to remove the MBP-Tag from LPG1G-MBP protein were performed. The purpose of this experiment is to have the possibility to study only the proteins of interest. Factor X was used to cleave the tag thanks to the specific site provided by pMAL-C2X vectors.

As showed in the Figure III-27, samples before and after three hours of digestion of purified LPG1G-MBP with factor Xa were analyzed on SDS-PAGE 12%. At t = 0 hour, two bands are visible, the highest one corresponds to the fused protein LPG1G-MBP (105.445 kDa) and the one below corresponds to the factor Xa with 43 kDa. After three hours of digestion, the band that corresponds to the fused protein disappeared and new one appeared. This corresponds to the cleaved protein LPG1G (62.75kDa) without its tag MBP. Similar band that the one observed at 0 hours is visible around 43 kDa, it corresponds to the overlapping of the factor (42 kDa) and the released Tag (42.459 kDa). This confirms the possibility to remove the MBP if we want to continue our experiments with only the protein of interest. However, further purification experiments should be carried out such ion exchange chromatography or SEC to purify the target protein, and remove the cleaved MBP-Tag and the factor Xa.<sup>70</sup> This additional steps will decrease the final protein concentration.

As was mentioned before, many proteins were characterized without removing their MBP-Tag, we decided then to continue my experiments and start the characterization of the protein of interest with their fused form.

### III. Conclusion

*lpg1*, *lpg1L*, *lpg1R* and *lpg1G* genes were amplified from *L. major* genomic DNA. These genes were predicted to encode for four putative galactofuranosyltransferases of *Leishmania*. Structural analysis predicted that the four putative genes contain transmembrane domain at their *N*-terminal and conserved catalytic DXD domain. Analyses showed that these putative GalT's do not share any significant homology of amino acid sequence with the other galactofuranosyltransferases described in literature, *i.e.* GlfT1, GlfT2 from *Mycobacterium tuberculosis*, GfsA from *Aspergillus nidulans* and WbbO from *Escherichia coli*-K12. Being different from the other GalT's, LPG1x studies could lead to the identification of new type of galactofuranosyltransferases.

In order to characterize the four leishmanial GalT's, the four genes were cloned into different vectors and transformed into bacterial host, *E. coli*. In order to facilitate the extraction of their respective recombinant proteins, regions which encode for transmembrane domain were removed. These regions are located at *N*-terminus of protein and distant from the catalytic site, making their removal possible without affecting their activity. First trials were performed with pET construction using pET-28a(+) and pET-32a(+) plasmids. This allowed us the expression of proteins fused with Histidine-tag in their *N*-termini. During the lysis and purification process, we encountered difficulties mostly related to the insolubility of the proteins of interest. We tried to overcome this problem by optimizing every step of our protocols in order to enhance the extraction and the solubility of the proteins. Finally, if we got better solubilization of LPG1x, it appeared that they were co-purified with chaperonin, a protein that help folding of non-folded proteins. Several trials were tested in order to get rid of this chaperonin but they were not successful. It is important to get pure proteins if we need to ensure their enzymatic characterization and determine their three-dimension structures. Moreover, they must be well folded, which is not the case here. Even if we succeeded in removing the chaperonin, it is not sure to regain native folding of proteins of interest. Thus, we decided to start the process with a new plasmid, pMal. Finally, we succeeded on getting pure proteins with good yield ranging from 5mg/L to 10mg/L of culture (figure III-28).

Producing recombinant LPG1X was very challenging, but we managed to obtain soluble LPG1, LPG1L, LPG1R and LPG1G. Consequently, it is now possible to start the enzymatic characterization in order to identify their substrate and kinetic parameters.

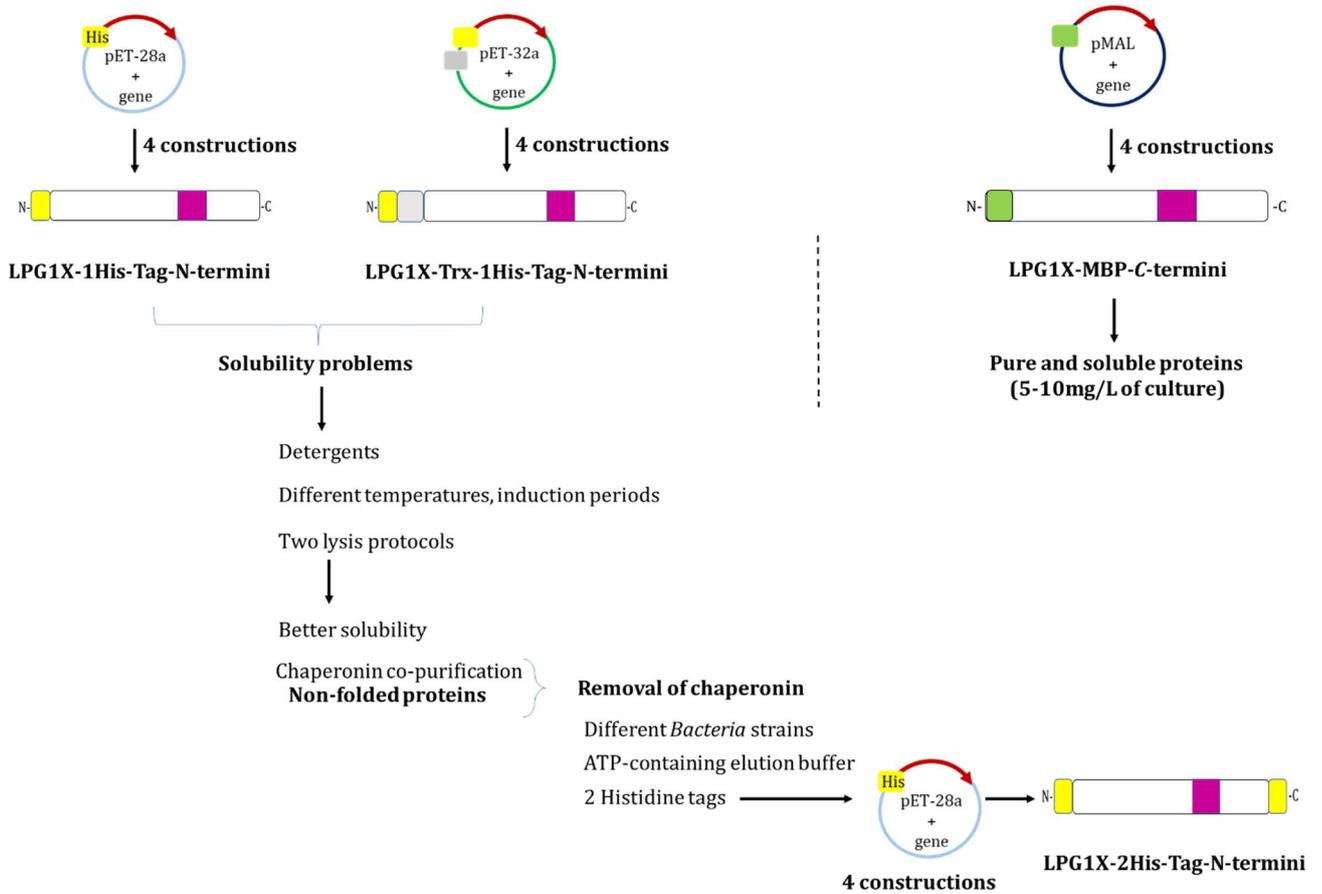


Figure III-28: Summary of the strategies used to produce LPG1x proteins.

## References

1. Makrides, S. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **60**, 512–538 (1996).
2. Saraswat, M. *et al.* Preparative purification of recombinant proteins: Current status and future trends. *Biomed Res. Int.* **11**, 1–18 (2013).
3. Terstappen, G. C. & Reggiani, A. In silico research in drug discovery. *Trends Pharmacol. Sci.* **22**, 23–26 (2001).
4. Zhang, K., Barron, T., Turco, S. J. & Beverley, S. M. The LPG1 gene family of *Leishmania major*. *Mol. Biochem. Parasitol.* **136**, 11–23 (2004).
5. Alonso, M. D., Lomako, J., Lomako, W. M. & Whelan, W. J. Catalytic activities of glycogenin additional to autocatalytic self-glycosylation. *Journal of Biological Chemistry* **270**, 15315–15319 (1995).
6. Parekh, R. B. Effects of glycosylation on protein function. *Curr. Opin. Struct. Biol.* **1**, 750–754 (1991).
7. Solá, R. J. & Griebenow, K. A. I. Effects of Glycosylation on the Stability of Protein Pharmaceuticals. *Biochemistry* **98**, 1223–1245 (2010).
8. Talhaoui, I. *et al.* Identification of key functional residues in the active site of human  $\beta$ 1,4-galactosyltransferase 7: A major enzyme in the glycosaminoglycan synthesis pathway. *J. Biol. Chem.* **285**, 37342–37358 (2010).
9. Fujiyama, K. *et al.* Human N-Acetylglucosaminyltransferase I. Expression in *Escherichia coli* as a soluble enzyme, and application as an immobilized enzyme for the chemoenzymatic synthesis of N-linked oligosaccharides. *J. Biosci. Bioeng.* **92**, 569–574 (2001).
10. Fengbin Wang, Maoquan Zhou, Shanteri Singh, Ragothaman M. Yennamalli1, Craig A. Bingman, Jon S. Thorson, and G. N. P. J. Crystal structure of SsfS6, the putative C-glycosyltransferase involved in SF2575 biosynthesis. *Proteins* **81**, 1277–1282 (2013).
11. Warnecke, D. *et al.* Cloning and functional expression of UGT genes encoding sterol glucosyltransferases from *Saccharomyces cerevisiae*, *Candida albicans*, *Pichia pastoris*, and *Dictyostelium discoideum*. *J. Biol. Chem.* **274**, 13048–13059 (1999).
12. Sahdev, S., Khattar, S. K. & Saini, K. S. Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. *Mol. Cell. Biochem.* **307**, 249–264 (2008).
13. Choi, S. J., Jang, K. J., Lim, J. A. & Kim, H. S. Human coagulation factor VIII domain-specific recombinant polypeptide expression. *Blood Res.* **50**, 103–108 (2015).
14. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* **5**, 1–17 (2014).
15. Jia, B. & Jeon, C. O. High-throughput recombinant protein expression in *Escherichia coli* : current status and future perspectives. *Open Biol.* **6**, 1–17 (2016).
16. Kwon, S., Kim, S. & Kim, E. Effects of glycerol of beta-lactamase production during high cell density cultivation of recombinant *Escherichia coli*. *Biotechnol. Prog.* **12**, 205–8 (1996).
17. Lodge, R. & Descoteaux, A. Modulation of phagolysosome biogenesis by the lipophosphoglycan of *Leishmania*. *Clin. Immunol.* **114**, 256–265 (2005).

18. Menon, A. K., Mayor, S. & Schwarz, R. T. Biosynthesis of glycosyl-phosphatidylinositol lipids in *Trypanosoma brucei*: involvement of mannosyl-phosphoryldolichol as the mannose donor. *EMBO J.* **9**, 4249–58 (1990).
19. Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H. & Kompala, D. S. Plasmid-encoded protein: The principal factor in the 'metabolic burden' associated with recombinant bacteria. *Biotechnol. Bioeng.* **35**, 668–681 (1990).
20. Wang, R. F. & Kushner, S. R. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene* **100**, 195–199 (1991).
21. Sørensen, H. P. & Mortensen, K. K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* **115**, 113–128 (2005).
22. Jana, S. & Deb, J. K. Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **67**, 289–298 (2005).
23. Ati, J., Lafite, P. & Daniellou, R. Enzymatic synthesis of glycosides: from natural *O*- and *N*-glycosides to rare *C*- and *S*-glycosides. *Beilstein J. Org. Chem.* **13**, 1857–1865 (2017).
24. Moffatt, B. A. & Studier, F. W. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49**, 221–227 (1987).
25. Doherty, A. J., Connolly, B. A. & Worrall, A. F. Overproduction of the toxic protein, bovine pancreatic DNaseI, in *Escherichia coli* using a tightly controlled T7-promoter-based vector. *Gene* **136**, 337–340 (1993).
26. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89 (1990).
27. Villaverde, A. & Carrió, M. M. Protein aggregation in recombinant bacteria: Biological role of inclusion bodies. *Biotechnol. Lett.* **25**, 1385–1395 (2003).
28. Gräslund, S. *et al.* Protein production and purification. *Nat. Methods* **5**, 135–146 (2008).
29. Wingfield, P. T. *Overview of the Purification of Recombinant Proteins. Curr Protoc Protein Sci* **80**, (2016).
30. GE Healthcare. Purifying Challenging Proteins. *Gen. Electr. Co.* **28-9095-31**, 1–107 (2007).
31. REPASKE, R. Lysis of gram-negative bacteria by lysozyme. *Biochim. Biophys. Acta* **22**, 189–191 (1956).
32. FEINER, R. R., MEYER, K. & STEINBERG, A. Bacterial lysis by lysozyme. *J. Bacteriol.* **52**, 375–384 (1946).
33. Burgess, R. R. Solubilized Inclusion Body Proteins. *Methods Enzymol.* **463**, 259–282 (2009).
34. Dumon-Seignovert, L., Cariot, G. & Vuillard, L. The toxicity of recombinant proteins in *Escherichia coli*: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr. Purif.* **37**, 203–206 (2004).
35. Schlicke, M. & Brakmann, S. Expression and purification of histidine-tagged bacteriophage T7 DNA polymerase. *Protein Expr. Purif.* **39**, 247–253 (2005).
36. Ryan, K. A., Garraway, L. A., Descoteaux, A., Turcot, S. J. & Beverley, S. M. Isolation of virulence genes directing surface glycosyl- phosphatidylinositol synthesis by functional complementation of *Leishmania*. *Proc. Natl. Acad. Sci. USA.* **90**, 8609–8613 (1993).
37. Kazemi, B. Genomic Organization of *Leishmania* Species. *Iran. J Parasitol* **6**, 1–18 (2011).

38. Rose, N. L. *et al.* Expression , Purification , and Characterization of a Galactofuranosyltransferase Involved in Mycobacterium tuberculosis Arabinogalactan Biosynthesis. *J. Am. Chem. Soc.* **128**, 6721–6729 (2006).
39. Wing, C., Errey, J. C., Mukhopadhyay, B., Blanchard, S. & Field, R. A. Expression and initial characterization of WbbI, a putative D -Gal f : a - D -Glc b -1 , 6-galactofuranosyltransferase from Escherichia coli K-12. *Org. Biomol. Chem.* **21**, 3945–3950 (2006).
40. Komachi, Y. *et al.* gfsA encodes a novel galactofuranosyltransferase involved in biosynthesis of galactofuranose antigen of O-glycan in Aspergillus nidulans and Aspergillus fumigatus. *Mol. Microbiol.* **90**, 1054–1073 (2013).
41. Szczepina, M. G., Zheng, B., Completo, G. C., Lowary, T. L. & Pinto, B. M. STD-NMR Studies Suggest that Two Acceptor Substrates for GlfT2 , a Bifunctional Galactofuranosyltransferase Required for the Biosynthesis of Mycobacterium tuberculosis Arabinogalactan , Compete for the Same Binding Site. *ChemBioChem* **10**, 2052–2059 (2009).
42. Wheatley, R. W., Zheng, R. B., Richards, M. R., Lowary, T. L. & Ng, K. K. S. Tetrameric Structure of the GlfT2 Galactofuranosyltransferase Reveals a Scaffold for the Assembly of Mycobacterial. *J. Biol. Chem.* **287**, 28132–28143 (2012).
43. Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D. S. Membrane protein structure determination - The next generation. *Biochim. Biophys. Acta - Biomembr.* **1838**, 78–87 (2014).
44. Kimple, M. E., Brill, A. L. & Pasker, R. L. Overview of affinity tags for protein purification. *Curr. Protoc. Protein Sci.* **73**, 608–616 (2013).
45. Yuan, L.-D. & Hua, Z.-C. Expression, purification, and characterization of a biologically active bovine enterokinase catalytic subunit in Escherichia coli. *Protein Expr. Purif.* **25**, 300–304 (2002).
46. Liu, Z. Q. & Yang, P. C. Construction of pET - 32  $\alpha$  ( + ) Vector for Protein Expression and Purification. *N. Am. J. Med. Sci.* **4**, 651–656 (2012).
47. Jenny, R. J., Mann, K. G. & Lundblad, R. L. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr. Purif.* **31**, 1–11 (2003).
48. Groot, D. & Ventura, S. Effect of temperature on protein quality in bacterial inclusion bodies. *FEBS Lett.* **580**, 6471–6476 (2006).
49. Huang, L., Yuan, Z., Liu, P. & Zhou, T. Effects of promoter leakage on dynamics of gene expression. *BMC Syst. Biol.* **9**, 1–12 (2015).
50. San-miguel, T., Pérez-bermúdez, P. & Gavidia, I. Production of soluble eukaryotic recombinant proteins in E . coli is favoured in early log-phase cultures induced at low temperature. *SpringerOpen J.* **2**, 2–5 (2013).
51. KOHN, A. Lysis of frozen and thawed cells of Escherichia coli by lysozyme and their conversion into spheroplasts. *J. Bacteriol.* **79**, 697–706 (1960).
52. Joshua A. Bornhorst and Joseph J. Falke. Purification of Proteins Using Polyhistidine Affinity Tags. *Methods Enzym.* **6**, 247–253 (2009).
53. Credou, J. & Berthelot, T. Cellulose: From biocompatible to bioactive material. *J. Mater. Chem. B* **2**, 4767–4788 (2014).
54. Rawlings, A. E. Membrane proteins: always an insoluble problem? *Biochem. Soc. Trans.* **44**, 790–795 (2016).

55. Lin, S. H. & Guidotti, G. Purification of Membrane Proteins. *Methods Enzymol.* **463**, 619–629 (2009).
56. Wu, R., Zhou, M. & Wu, H. Purification and Characterization of an Active N-Acetylglucosaminyltransferase Enzyme Complex from Streptococci. *Appl. Environ. Microbiol.* **76**, 7966–7971 (2010).
57. Rose, N. L. *et al.* Development of a coupled spectrophotometric assay for GlfT2 , a bifunctional mycobacterial galactofuranosyltransferase. *Carbohydr. Res.* **343**, 2130–2139 (2008).
58. Thiede, B. *et al.* Peptide mass fingerprinting. *Methods* **35**, 237–247 (2005).
59. Braig, K. *et al.* The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**, 578–586 (1994).
60. Wang, J. D. & Weissman, J. S. Thinking outside the box: new insights into the mechanism of GroEL-mediated protein folding. *Nat. Struct. Mol. Biol.* **6**, 597–600 (1999).
61. Mark Mayhew, Ana C.R da Silva, Jorg Martin, Hediye Erdjument-Bromage, P. tempst and U. H. Protein folding in the central cavity of the GroEL-GroES chaperonin complex. *Nature* **379**, 420–426 (1996).
62. Vabulas, R. M., Raychaudhuri, S., Hayer-hartl, M. & Hartl, F. U. Protein Folding in the Cytoplasm and the Heat Shock Response. *CSH Perspect.* **11**, 1–19 (2018).
63. Altschuler, G. M. & Willison, K. R. Development of free-energy-based models for chaperonin containing TCP-1 mediated folding of actin. *J. R. Soc. Interface* **5**, 1391–1408 (2008).
64. Liberek, K., Lewandowska, A. & Ziętkiewicz, S. Chaperones in control of protein disaggregation. *EMBO J.* **27**, 328–335 (2008).
65. Molugu, S. K., Li, J. & Bernal, R. A. Separation of E. coli chaperonin groEL from β-galactosidase without denaturation. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **1007**, 93–99 (2015).
66. Keskin, O., Bahar, I., Flatow, D., Covell, D. G. & Jernigan, R. L. Molecular mechanisms of chaperonin GroEL-GroES function. *Biochemistry* **41**, 491–501 (2002).
67. Joseph, R. E. & Andreotti, A. H. Bacterial expression and purification of Interleukin-2 Tyrosine kinase: Single step separation of the chaperonin impurity. *Protein Expr. Purif.* **60**, 194–197 (2008).
68. Wagner, S. *et al.* Tuning Escherichia coli for membrane protein overexpression. *PNAS* **105**, 14371–14376 (2008).
69. Motojima, F. How do chaperonins fold protein? *Biophysics (Oxf)*. **11**, 93–102 (2015).
70. Krisna C. Duong-Ly and Sandra B. Gabelli. Affinity Purification of a Recombinant Protein Expressed as a Fusion with the Maltose-Binding Protein (MBP) Tag. *Methods Enzym.* **91**, 165–171 (2015).

***Chapter IV: Characterization of  
prokaryotic LPG1x.***

## Abstract

Since galactofuranose residues are absent from mammalian cells and present in several pathogenic species, the galactofuranosyltransferases that are responsible of their addition into the polysaccharide's structures were considered as potent therapeutic targets. However, up to date, we lack information about these GTs. Indeed, only the mycobacterial GalT; GlfT2, has been fully characterized and described. However, the kinetic parameters of GfsA and WbbI are not available. This is related the accessibility to the substrate, UDP-Galactofuranose that is not commercial as its synthesis is considered to be a very challenging task. Moreover, most of the reported methods used to evaluate enzymes activity are not appropriate for transferases in a generic manner.

In this chapter, spectrophotometric assays allowed the evaluation of the activity leishmanian galactofuranosyltransferases that were produced by *E. coli*. Indeed, we demonstrated that LPG1, LPG1R, LPG1L and LPG1G were not only able to transfer UDP-Galf onto acceptor, Me-Man<sub>p</sub>, but also recognize different NDP-sugars as donors. This consist of the first enzymatic characterization of eukaryotic GalTs.

## Résumé

Les galactofuranosyltransferases, qui sont responsables de l'incorporation des résidus galactofuranose, sont absent chez les mammifères mais présent dans plusieurs espèces pathogènes, sont considérées comme des cibles thérapeutiques intéressantes. Cependant, à ce jour, nous manquons d'informations sur ces familles de glycosyltransferases. En effet, seul la GalT de *Mycobacterium tuberculosis*; GlfT2, a été entièrement caractérisée. Cependant aucune donnée sur les paramètres cinétiques des autres GalTs ; GlfT1, GfsA et WbbI ne sont disponible. Ceci est lié à l'indisponibilité du substrat, UDP-Galf. En effet ce donneur de sucre n'est pas commercial et sa voie synthèse est considérée comme une tâche très laborieuse. De plus, la plupart des méthodes utilisées pour évaluer l'activité des enzymes ne conviennent pas aux transférases de manière générale.

Dans ce chapitre, les analyses spectrophotométriques ont permis d'évaluer les activités des galactofuranosyltransférases qui ont été produite précédemment par *E. coli*. En effet, nous avons démontré que LPG1, LPG1R, LPG1L et LPG1G étaient non seulement capables de transférer UDP-Galf sur l'accepteur, Me-Man<sub>p</sub>, mais reconnaissaient également différents sucres NDP en tant que donneurs. Il s'agit de la première caractérisation enzymatique de GalT eucaryotes.

I-	The characterization of GalT's is an ongoing challenge .....	125
A.	UDP-Galactofuranose synthesis .....	125
1.	Chemical approaches .....	125
2.	Enzymatic approaches.....	127
a)	Mutase reaction .....	127
b)	GalPUT reaction.....	127
B.	Enzymatic assays of GT .....	128
1.	Spectrophotometric assays .....	129
a)	The coupled spectrophotometric assays.....	129
b)	Malachite green assay .....	130
2.	High performance liquid chromatography .....	130
II.	Enzymatic characterization of LPG1x proteins.....	131
A.	Coupled Spectrophotometric assay .....	131
1.	Galactofuranosyltransferase activity of LPG1x proteins.....	131
2.	Product characterization.....	135
B.	Substrate specificity .....	137
1.	Donor specificity .....	137
a)	Screening assays.....	137
b)	Product characterization.....	140
2.	Acceptors specificity.....	142
a)	Screening of hexoses as acceptors .....	142
b)	Screening of PNP-sugars as acceptors.....	142
3.	Temperature and pH effect.....	144
III.	Crystallization of LPG1x.....	145
IV.	Conclusion .....	149
	References.....	151

## I- The characterization of GalT's is an ongoing challenge

The cells surface is heavily covered by oligosaccharides, which play numerous biological roles, for examples in cell-cell communication, signaling, development, pathogen adhesion and infection.<sup>1,2</sup> Thus, understanding the implication and specific roles of oligosaccharides in these processes became one of biggest interest of glycoscientists. Indeed, there is a great interest for the access to defined glycosyltransferases to identify their substrates and activities. Among these GTs, galactofuranosyltransferases have attracted lot of attention as a potent therapeutic target for some infectious diseases, such aspergillosis, tuberculosis and leishmaniosis.<sup>3,4</sup> However, as mentioned before, only mycobacterial GalT2 was fully characterized and described.<sup>5,6</sup> It is important to highlight the fact that working on putative GTs and identifying their donors and acceptors substrates remain a hard task. Currently there are 446370 entities for GTs in CAZy data base (<http://www.cazy.org/>; September 2018), however there are only 2002 GTs that have been characterized and only 223 structural data are available, compared to 7336 characterized glycosyl hydrolases with 1286 crystal structure for 526565 entities. This is practically due to the lack of efficient tools to functionally characterize GTs, in addition to problems related to the availability of their substrates. In our case, we are interested in putative GalT's of *Leishmania*, the causing agent of leishmaniosis disease.<sup>7,8</sup> In the previous chapter, we have described the production of four pure and soluble prokaryotic LPG1x, which allowed us to move the next step, their enzymatic characterization. However, their putative substrate, UDP- $\alpha$ -D-Galf, is not commercially available and its synthesis remains an ongoing challenge.<sup>9</sup> Moreover, few assays are suitable to analyze their activities.

### A. UDP-Galactofuranose synthesis

#### 1. Chemical approaches

Tsvetkov and Nikolaev proposed a method based on the coupling between  $\alpha$ -D-galactofuranosyl phosphate and imidazoluridine 5-monophosphate to produce yield of 45-50% of UDP-Galf after 19 hours of reaction.<sup>10</sup> (Figure IV-1) Then, Kiessling and coworkers proposed another strategy by using activated 5'-N-methyl phosphorylimidazolide with Galf 1-phosphate and obtained 35% of UDP-Galf after in short reaction time; 2 hours. (Figure IV-2).<sup>11</sup> An alternative method was developed later by Ferrières and co-workers that involves direct coupling between an unprotected thioimidate and a phosphoric acid<sup>12,13</sup> (Figure IV-3). Nevertheless, these methods are not very effectives.

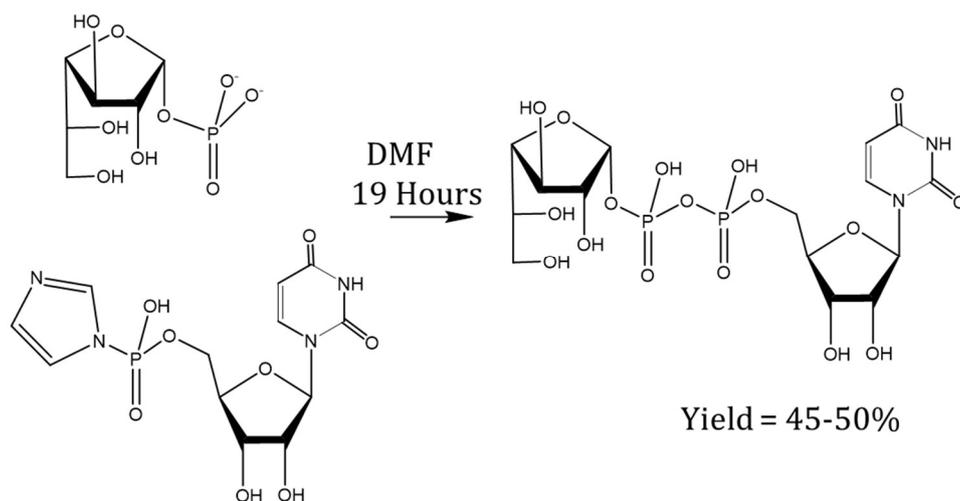


Figure IV-1 : Synthesis of UDP-Galf described by Tsvetkov and Nikolaev.<sup>10</sup>

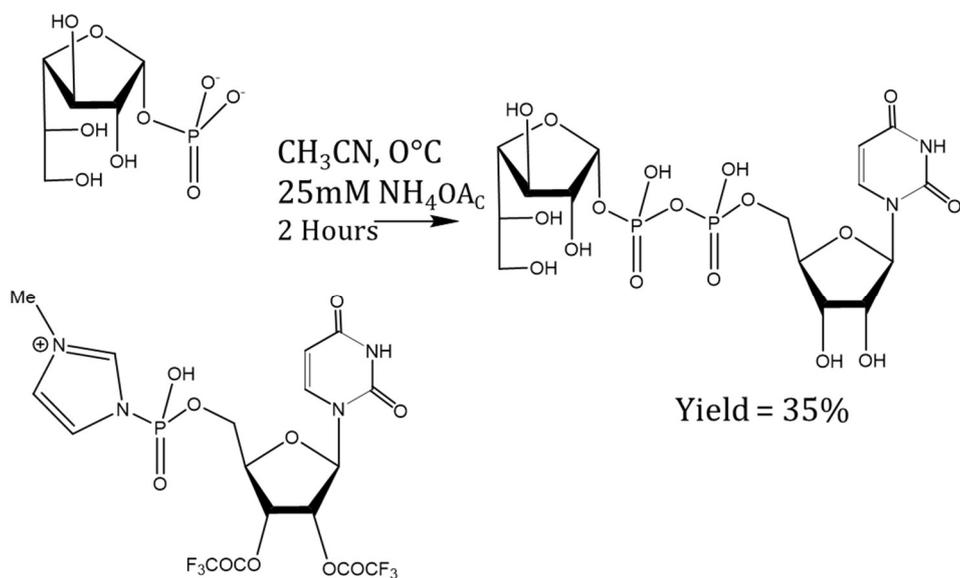
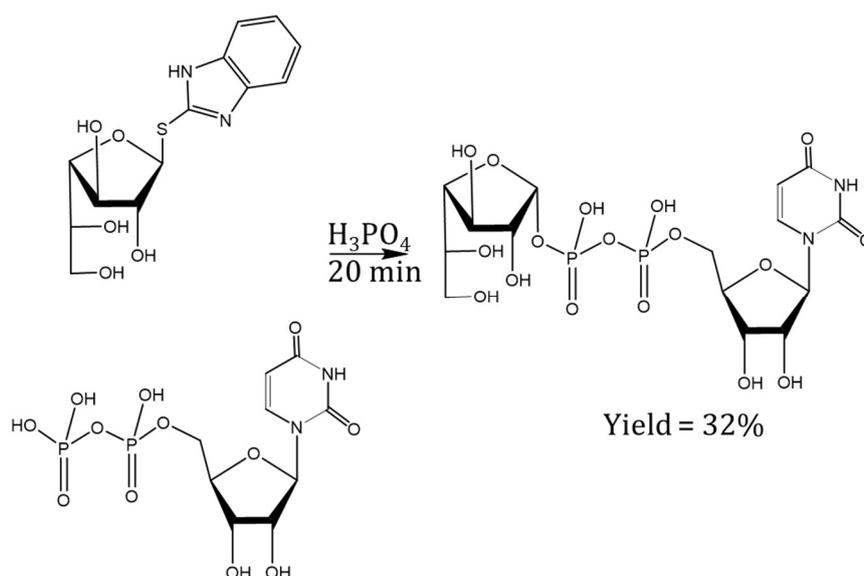


Figure IV-2 : Synthesis of UDP-Galf described Kiessling and co-workers starting from an activated 5'-N-methyl phosphorylimidazole.<sup>11</sup>

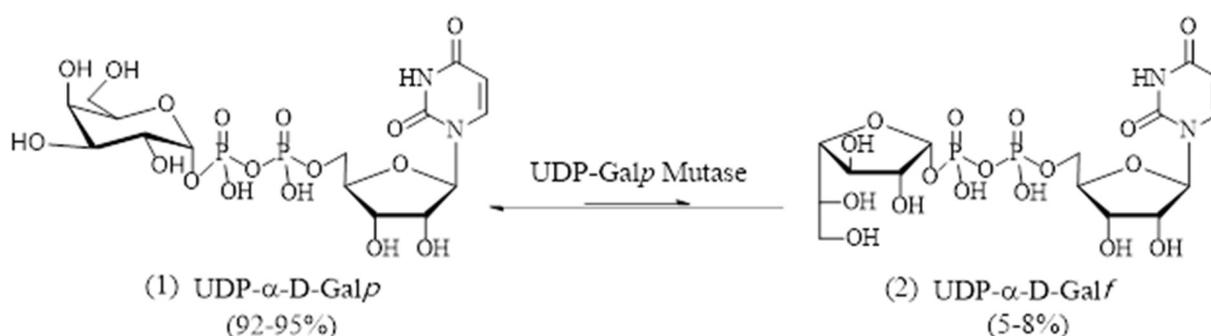


**Figure IV-3:** Synthesis of UDP-Galf described by *Ferrières* and co-workers starting from thiomidates

## 2. Enzymatic approaches

### a) Mutase reaction

It is based on the activity of a specific enzyme, UDP-galactopyranose mutase that catalyzes the conversion of UDP-Galp (1) to UDP-Galf (2).<sup>11</sup> However, this reaction is not efficient enough to give sufficient quantities of UDP-Galf to perform biological and biochemical studies. Indeed, the yield of this step is around 5-8%. The low efficiency is related to equilibrium position of the reaction in favor of UDP-Galp.<sup>14 13</sup> (Figure IV-4)

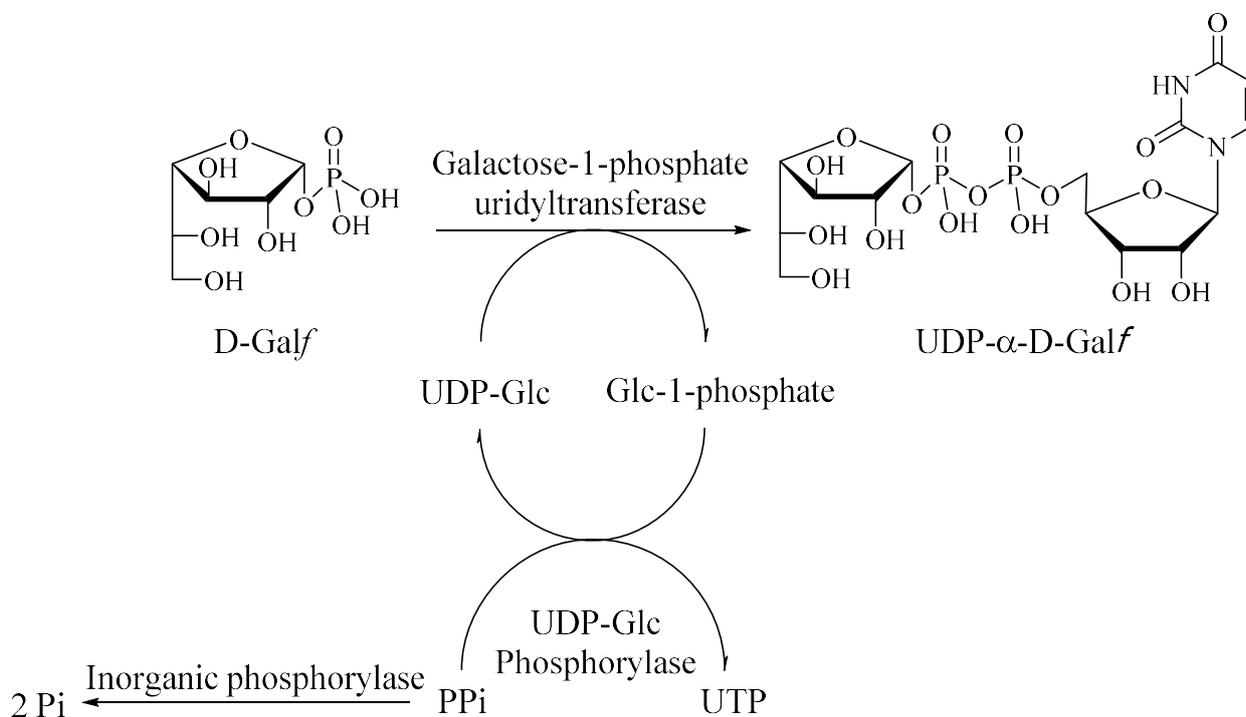


**Figure IV-4:** UDP-Galp Mutase catalyses the conversion of UDP-Galp (1) into UDP-Galf (2). UDP-Galp is favoured. (Adapted from Allison L. Marlow and Laura L. Kiessling, 2001).<sup>11</sup>

### b) GalPUT reaction

This approach was proposed by Field and co-workers, and is based on the previous work of Wang and co-workers.<sup>15</sup> (Figure IV-5) A key enzyme of Leloir pathway of galactose metabolism, Galactose-1-phosphate uridylyl transferase (GalPUT), is used. It catalyses the conversion of UDP-D-glucopyranose (UDP-D-Glcp) and D-galactopyranose-1-phosphate (D-Galp) to D-glucopyranose-1-phosphate (D-Glcp-1-P) and UDP-D-galactopyranose (UDP-D-Galp) respectively. In this case,

GalPUT, thanks to its relaxed substrate specificity was also able to accept D-galactofuranose-1-phosphate as a substrate (D-Galf-1-P), to generate UDP-D-galactofuranose. This method is considered as the best and most efficient one because of its good yield of 79%.<sup>16,17</sup>



**Figure IV-5:** Chemoenzymatic strategy to synthesize UDP- $\alpha$ -D-Galf described by Field and coworkers.<sup>16</sup>

Daniellou and co-workers used the same the chemoenzymatic procedure and obtained natural UDP- $\alpha$ -D-Galf as well as three 1,2-cis synthetic analogues using anomeric mixture of the of the corresponding phosphates.<sup>14</sup>

### B. Enzymatic assays of GT

To evaluate the activity of enzymes, different techniques can be performed. The most known techniques are fluorescence, chromatography, colorimetry, spectroscopy and potentiometry, capillary electrophoresis, nuclear magnetic resonance, mass spectrometry and high liquid performance. Most of the reported methods tend to be very specific and are not appropriate to transferases in a generic manner.<sup>18</sup> It is a demanding task to evaluate the activity of GalT's, and this due to many factors; i) the availability of the putative substrate, UDP-Galf. ii) the absence of a direct reaction that could detected using spectrophotometric techniques, which is also closely related to the lack of colorimetric substrates that could be recognized by these enzymes. iii) The obtention of poor amount of pure product that makes difficult to detect with NMR strategies as example and reveal their structures. Nevertheless, some techniques, described above, were adopted to study the activity of mycobacterial GalT's as much as other GT's and can be considered

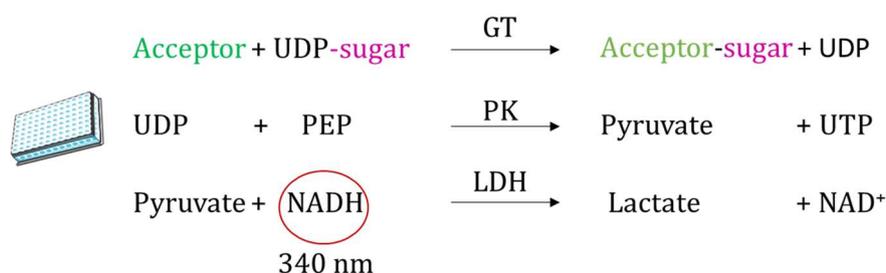
to be used with for our LPG1x enzymes. These methods were chosen because they are easy to perform, reliable, reproducible, and the needed chemicals and materials are not expensive.

### 1. Spectrophotometric assays

Spectrophotometric assays are widely used to study the progress of the reactions, thus the disappearance of substrates and the formation of products. It is simple and very sensitive test, based on the measurement of the amount of light a substance absorbs during the reaction. Then, a shift of the absorbance spectrum indicates that a change in the substrate structure has taken place.

#### a) The coupled spectrophotometric assays

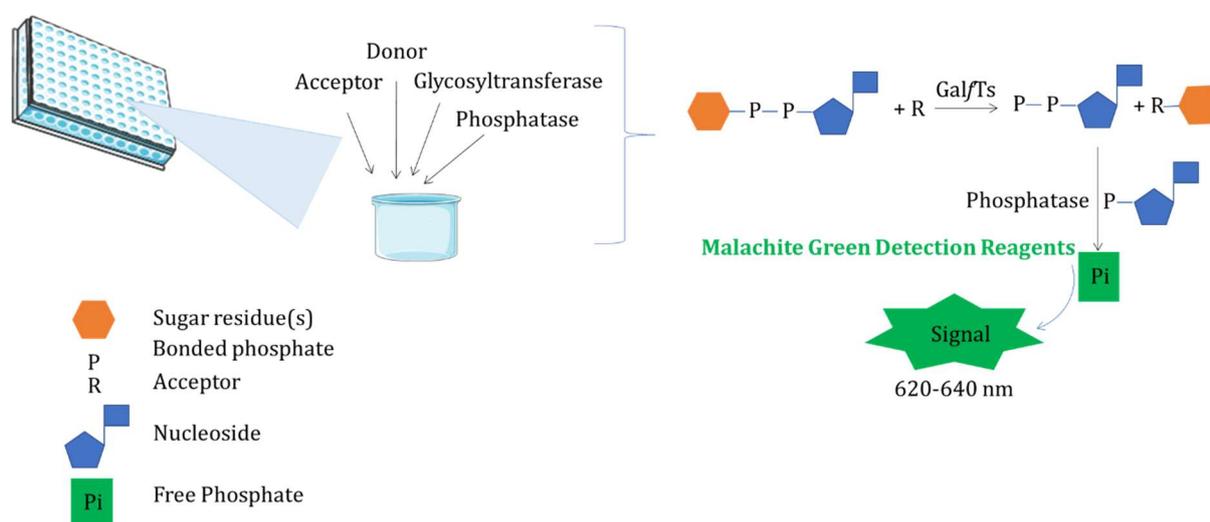
Gosselin et al., described enzymatic assays that consists of two consecutives reactions following the one that is carried out by one of the glycosyltransferases.<sup>19</sup> (Figure IV-6) First, the transfer of a monosaccharide moiety from the sugar donor to an acceptor will generate the formation of disaccharide and liberate uridyl diphosphate (UDP). Then, pyruvate kinase (PK) catalyzes the transfer of phosphate group from phosphoenolpyruvate, that was added in the mixture reaction from the beginning, to the liberated UDP. This results the formation of pyruvate and uridyl triphosphate (UTP).<sup>20</sup> Finally, lactate dehydrogenase (LDH) catalyzes the conversion of pyruvate to lactate, as it converts NADH to NAD<sup>+</sup>.<sup>21</sup> Here, NADH absorbs light at a specific wavelength, 340nm, while NAD<sup>+</sup> does not. So, the shift of the absorbance indicates the oxidation of an amount of NADH, which reflect the transfer of an equal amount of transferred galactofuranose. This assay allowed them to evaluate the activity of galactosyltransferases, fucosyltransferases and *N*-acetylglucosaminyltransferase that respectively utilize UDP-Gal, GDP-Fuc, and UDP-GlcNac. This was accomplished thanks to the fact that the pyruvate kinase is able to recognize UDP, GDP, CDP and ADP.<sup>22</sup> Thus, it is possible to expand this assay for various GTs using different nucleotide donors. This same strategy was adopted by Todd L. Lowary and co-workers to successfully evaluate the activity of mycobacterial galactofuranosyltransferase GlfT2.<sup>5</sup>



**Figure IV-6:** Coupled spectrophotometric assay to measure the activity of GalT by following the consumption of NADH at 340nm. <sup>19,17</sup>

## b) Malachite green assay

Zhengliang LWu et al., proposed a Universal phosphatase-coupled glycosyltransferase assay, known also as Malachite green assay, as a rapid method to assess the activity of GTs.<sup>23,24,25,26</sup> It is based on the quantitative detection of released inorganic phosphate (Pi) by phosphatase from liberated UDP. It is important to highlight the fact that phosphatase do not use UDP-sugars as substrate, so, the detected Pi is directly proportional to the transferred sugar.<sup>23</sup> This assay was used to screen different GT activities.<sup>26,27</sup> The concentration of released Pi is measured with malachite green detection reagents. Obtained level allow the measurement then and the evaluation of the activity of target enzymes. Indeed, Pi molecules form a complex with malachite green molybdate under acidic conditions. This complex can be measured at 620-640nm. (Figure IV-7)



**Figure IV-7:** Malachite green assay principle. Glycosyltransferases transfer sugars residues from an activated nucleotide sugar donor to an acceptor. Phosphatase generates phosphate from liberated UDP. Pi form complex with malachite Green reagent that is detected at 620-640 nm.

## 2. High performance liquid chromatography

High performance liquid chromatography (HPLC) is highly versatile techniques widely used to separate and measure components of a liquid samples. It is based on the interaction of the reaction components, which is in the mobile or liquid phase, with the stationary phase (SP).<sup>28</sup> Here, the column is packed with micro-scale beads. These beads are functionalized with chemical groups that increase the interaction between the beads and the component of the mixture. The components interact with the stationary phase differently and therefore travel the full length of the column into the detector at a different rate. The time required for a component to leave the column or elute is called the retention time.<sup>29</sup> The result is a plot of retention vs. intensity or a chromatogram. Retention time is used to identify the components. The peak size, particularly the

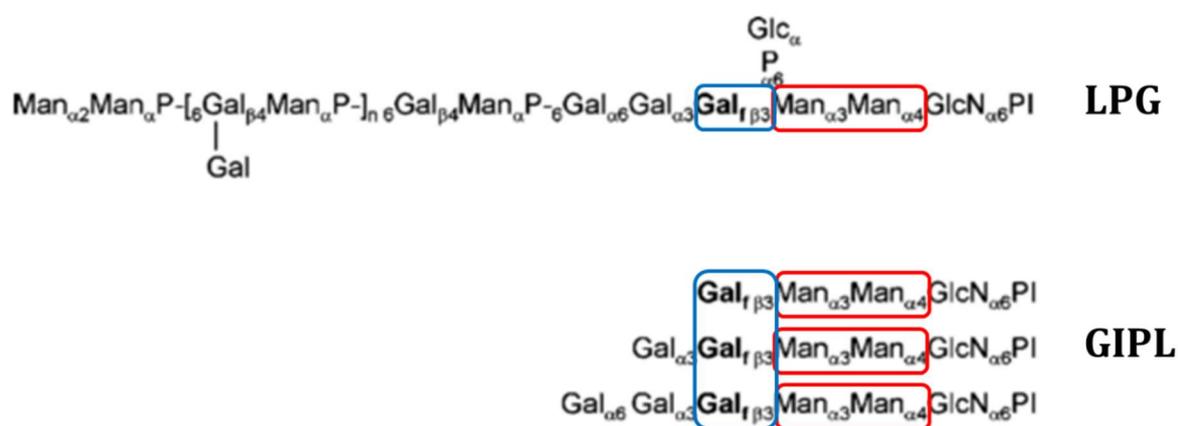
region under the peak, is used to quantify the amount of the compound in the initial solution.<sup>30</sup> The choice of the stationary phase depends on the properties of the components in the sample mixture.<sup>31</sup> The most commonly used stationary phase is silica beads, as they are an inert and non-polar material that forms micro-scale beads and provides a sufficient packing density. The most common type of HPLC is reverse phase chromatography, which uses a hydrophobic stationary phase, usually silica beads with C<sub>18</sub> chains adhered to the surface of the beads. The components are eluted in order of decreasing polarity.<sup>32</sup> Two types of HPLC, analytical and preparative, can be used in the laboratory.<sup>33</sup> First, the analytical one is used to identify components; however, the analyzed samples are discarded as wastes. Then, the preparative HPLC allow them to recover compounds into collected fractions. As an example, preparative HPLC was used to purify the UDP-Galf used as donor for mycobacterial GlfT2 and its products.<sup>17</sup>

## II. Enzymatic characterization of LPG1x proteins

### A. Coupled Spectrophotometric assay

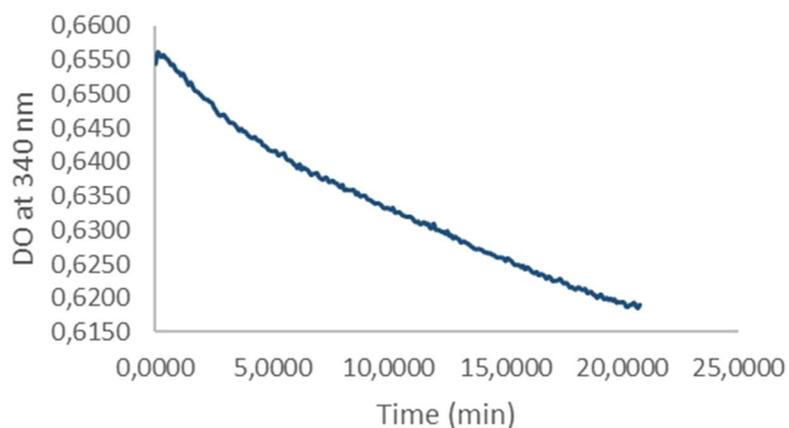
#### 1. Galactofuranosyltransferase activity of LPG1x proteins

Three-enzymatic assay that was developed by Todd L. Lowary and co—workers was used.<sup>5</sup> We decided to start with methyl- $\alpha$ -D-mannopyranoside (Me-Man<sub>p</sub>) and D-mannopyranose (D-man<sub>p</sub>) as acceptors. This choice was made based on the structures of LPG and GIPLs of *Leishmania* that were described by Zhang et al. Indeed, there is galactofuranose residues linked to mannose (Figure IV-8).



**Figure IV-8:** LPGs and GIPLs structures in *Leishmania major* containing galactofuranose residue linked to mannose residue. Galf: Galactofuranose, Man: Mannose, Gal: Galactose, Glc: Glucose.

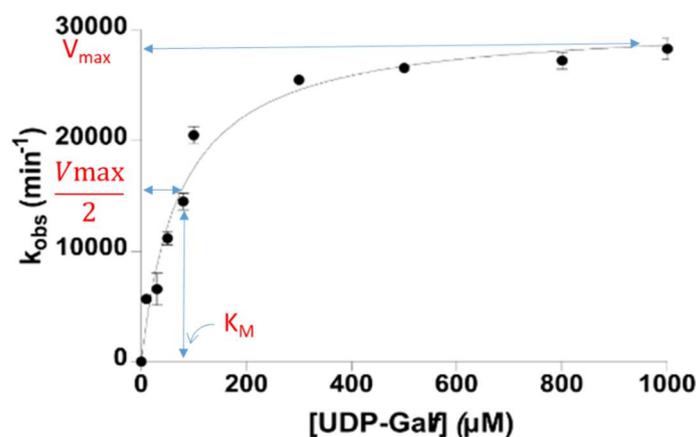
At the beginning, reactions were monitored at 10s intervals for up to 25 min. First trials were tested with 1mM final concentration of UDP- $\alpha$ -D-Galf and 0.2 to 10 $\mu$ g of enzymes. Several negative controls have been carried out for each reaction *i.e.* reaction without donors, without enzymes, without acceptors, without and PK/LDH.



**Figure IV-9:** Oxidation of NADH over time detected at 340nm in 200 $\mu$ l reaction mixture containing 1Mm UDP-Galf, 1Mm Me-Manp and 0.2 $\mu$ g of LPG1.

Linear decrease of the absorbance at 340nm for 20 min was registered during the reaction with LPG1, UDP-Galf as donor and Me-Manp as acceptor (Figure IV-9). The shift of the absorbance is related to the oxidation of an amount of NADH. This is stereochemically identical to the amount of galactofuranose transferred on Me-Manp. Nevertheless, we did not detect any activity of the enzymes for reactions that were run with D-mannopyranose as acceptor. Therefore, my only hit was with Me-Manp as acceptor for LPG1. Similar results were obtained with LPG1L, LPG1R and LPG1G. They also seem to accept only methyl- $\alpha$ -D-mannopyranoside as acceptor, with UDP- $\alpha$ -D-galactofuranose as donor.

Based on these preliminary results, we decided to incubate all our reactions up to 20 min. Reactions contained similar concentration of enzyme and Me- $\alpha$ -D-mannopyranoside, but with different concentrations of UDP-Galf ranging from 10 to 1000 $\mu$ M, in order to evaluate the enzymes activities.



**Figure IV-10:** Michaelis-Menten plot that corresponds to UDP-Galf kinetics as donor substrate of LPG1. Reactions were performed in triplicate with 0.2 $\mu$ g of enzyme and 1mM of Me-Manp.  $V_{max}$

is the maximal velocity of the enzymatic reaction. The Michaelis constant ( $K_M$ ) is the substrate concentration yielding a velocity of  $V_{max}/2$ .

Using 1mM of Me-Manp as acceptor and different concentrations of UDP-Galf as donor, Michaelis-Menten plots were obtained (Figure IV-10), which is one of the best-known models and most used approaches to identify enzyme kinetics and its equation was proposed in in 1913 by German Leonor Michaelis and Canadian Maud Menten (Equation IV-1). Kinetics parameters can be distinguished; the maximal velocity ( $V_{max}$ ), which corresponds to the saturation state of the enzyme at saturating substrate concentrations, and the Michaelis constant ( $K_M$ ), which is the substrate concentration at which the reaction rate is half of  $V_{max}$ . The saturation state is due the fact that the active site of enzyme is occupied by substrate. Then, increasing the concentration of substrate, does not affect the rate of the reaction and enzymatic reaction cannot go faster.

Similar results were obtained with LPG1L, LPG1R and LPG1G and their Michaelis-Menten plots are presented in Appendix 7.

$$v_0 = \frac{V_{max} \times [S]}{K_M + [S]}$$

**Equation IV-1:** Michaelis–Menten equation.  $[S]$  is the substrate concentration,  $v_0$  and  $V_{max}$  are, respectively, the initial and the apparent maximum velocity rate, and  $K_M$  is the apparent Michaelis constant.

Here, the Michaelis–Menten plots prove that the rate of LPG1x catalysed-reactions depends on the concentration of UDP-Galf. Kinetics parameters were calculated by fitting saturation curves, which were obtained from the average of triplicate measurements, with standard Michaelis-Menten equation, using Prism 6 (GraphPad) software. Thus, the  $k_{cat}$  and  $\frac{k_{cat}}{K_M}$  were calculated, where  $k_{cat}$  is the maximum rate of the reaction at saturating substrate concentrations with  $V_{max} = k_{cat} E_T$ , where  $E_T$  corresponds to the total enzyme concentration.  $\frac{k_{cat}}{K_M}$  corresponds to the catalytic efficiency, which is often referred to as the “specificity constant”. It is usually used as a index to compare the turnover of enzyme for the same substrate. Obtained kinetics parameters are shown in Table IV-1.

	Enzyme	Apparent $K_M^a$ (mM)	$k_{cat}^a$ (min <sup>-1</sup> )	$k_{cat}/K_M^a$ (min <sup>-1</sup> mM <sup>-1</sup> )
UDP Gal $f$	LPG1	0.07 ± 0.001	30,750 ± 62	410, 000
	LPG1G	0.03 ± 0.004	12,352 ± 58	393, 655
	LPG1L	0.02 ± 0.002	5,296 ± 63	290, 800
	LPG1R	0.55 ± 0.068	636 ± 76	1, 145
	Gl $f$ T2 <sup>14</sup>	0.38 ± 0.006	430 ± 35	1, 131

**Table IV-1:** Kinetic parameters of leishmanial Gal $f$ Ts LPG1, LPG1G, LPG1L, LPG1R compared with mycobacterial Gl $f$ T2 for UDP- $\alpha$ -D-Gal $f$ . <sup>a</sup>: Me-Man $p$  was used as the acceptor

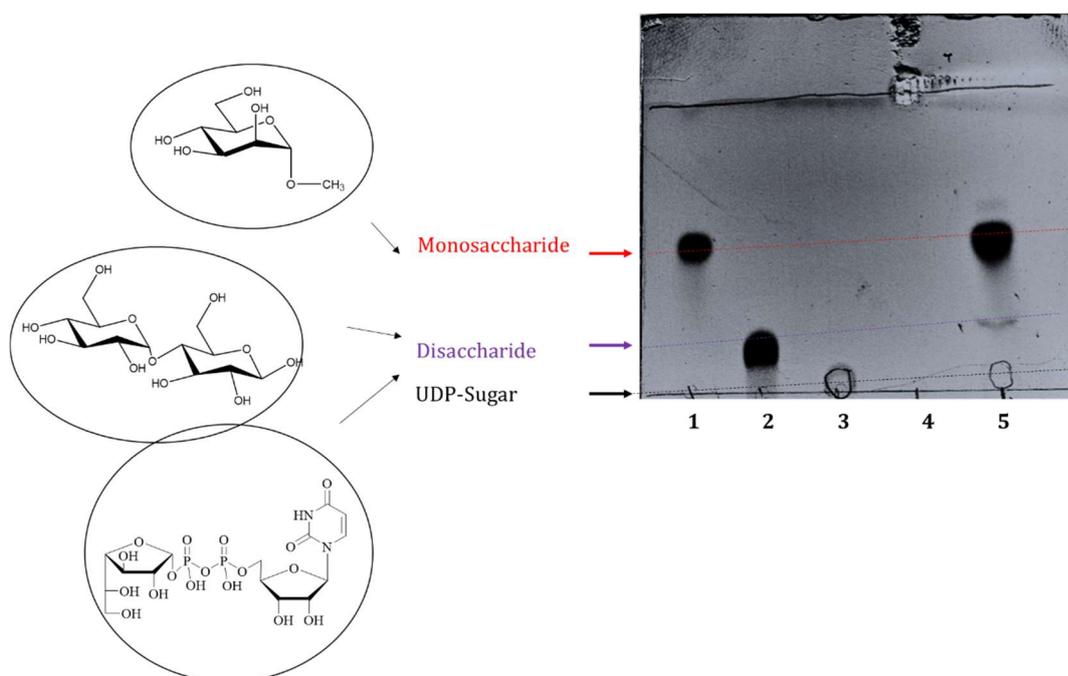
Using Me-Man $p$  as the acceptor, kinetics data was obtained (three replicate experiments) with the respective enzymes. Michaelis–Menten analysis of the data is shown in Table IV-1. All four enzymes recognized UDP-Gal $f$  with apparent  $K_M$  values ranging from 0.02 to 0.55 mM. Based on that parameter, the proteins can be clustered in two groups: LPG1, LPG1G, and LPG1L, which strongly bind to UDP-Gal $f$  with  $K_M$ 's in the low  $\mu$ M level, and LPG1R, which binds UDP-Gal $f$  more weakly. The UDP-Gal $f$   $K_M$  value for LPG1R is comparable to the one reported for mycobacterial Gl $f$ T2 (0.38 mM). Moreover, the  $k_{cat}$  values of LPG1, LPG1G and LPG1L range from 5,000 to 30,000 min<sup>-1</sup>, which are superior to the catalytic rate for LPG1R, which is at least 10-fold time lower with a value of 636 min<sup>-1</sup> (and comparable to the catalytic rate reported for Gl $f$ T2). These marked differences in the Michaelis constant and turnover rate are found also when estimating the catalytic efficiency ( $k_{cat}/K_M$ ) of LPG1, LPG1G and LPG1L yielding values higher than 300,000 min<sup>-1</sup> mM<sup>-1</sup> as opposed as 1,145 min<sup>-1</sup> mM<sup>-1</sup> and 1,131 min<sup>-1</sup> mM<sup>-1</sup> for LPG1R and Gl $f$ T2 respectively.

In summary LPG1, LPG1G and LPG1L demonstrated strong in vitro Gal $f$ Ts properties, at least 300 times higher than LPG1R and the previously reported mycobacterial Gl $f$ T2 from *M. tuberculosis*. However, in *M. tuberculosis* Gl $f$ T2 is a polymerizing enzyme that adds around 30 Gal $f$  units, linked by alternating by  $\alpha$ -(1→5) and  $\alpha$ -(1→6) glycosidic bonds.<sup>34</sup> Gl $f$ T2 belongs to the CAZY glycosyltransferase (GT) family 2,<sup>35,36</sup> which contains mainly polymerizing enzymes such as the cellulose or the chitin synthase. On the contrary, the LPG1x enzymes belong to the CAZY GT family 40, which only contains putative Gal $f$ Ts from trypanosomatids. In *Leishmania* species, Gal $f$ Ts introduce only one Gal $f$  residue into the 3-OH position of the mannosyl acceptor with  $\beta$ -selectivity for example like in LPG and GIPL.

## 2. Product characterization

Scale up reaction was done in order to obtain sufficient amount of product for its characterization. In order to follow the formation of the disaccharide, analytical thin layer chromatography (TLC) technique was used.

TLC is a technique used to separate compounds in mixture solution according to their polarity. In this case I used, paper silica gel ( $\text{SiO}_2$ ) as adsorbent to attract components, that are in the reaction mixture. Elution mixture contained ethyl acetate, methanol and water with the ratio 7/2/1. The choice of the eluent is very important to separate all the component because it affects their migration and interaction with the adsorbent by changing their solubility.<sup>37</sup> Spots, that correspond to each sugar were revealed with orcinol solution that contain 95% of ethanol, 5% of  $\text{H}_2\text{SO}_4$  and 16 $\mu\text{M}$  of orcinol.

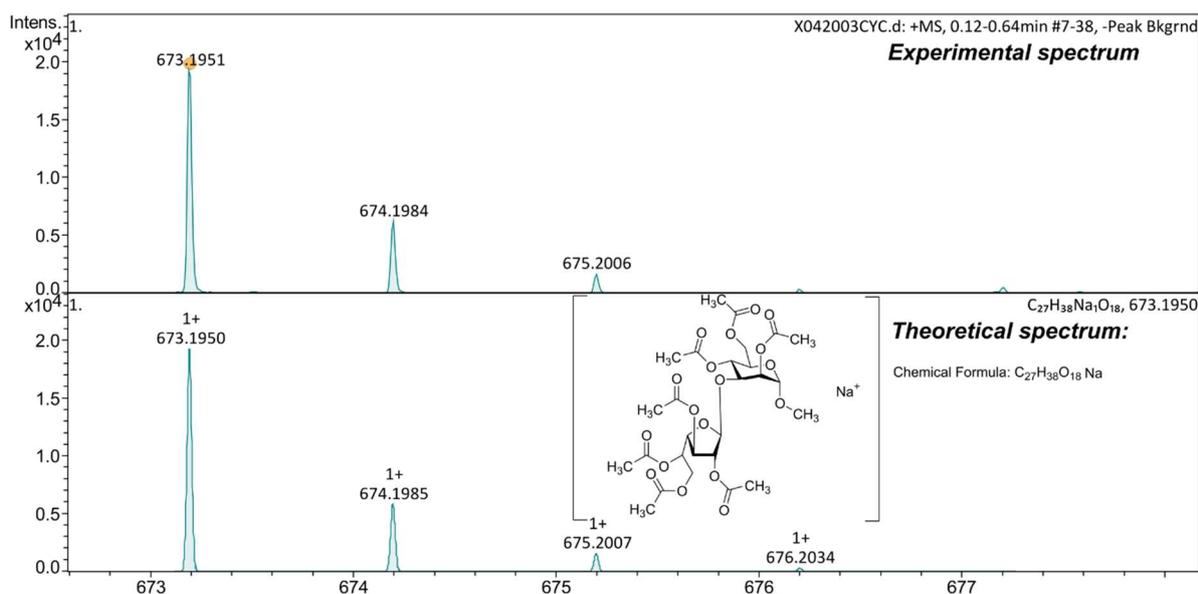


**Figure IV-11:** Typical TLC obtained with the GalT reaction and UDP-  $\alpha$ -D-Galf as donor. Lane 1: Methyl- $\alpha$ -D-mannopyranoside. Lane 2: Maltose. Lane 3: UDP- $\alpha$ -D-Glucopyranoside. Lane 4: LPG1. Lane 5: LPG1G catalysed reaction with UDP- $\alpha$ -D-galactofuranoside as donor and methyl- $\alpha$ -D-mannopyranoside as acceptor.

The Figure IV-11 shows the result of TLC obtained after migration of amount of the acceptor Me- $\alpha$ -D-mannopyranoside in the first lane, followed by an amount of maltose in lane 2, and the UDP- $\alpha$ -D-galactofuranoside in lane 3. The maltose is used as a reference of the migration of a disaccharide because we don't have the expected product, Me-Man $p$ -Galf. We estimated that maltose will migrate approximately as our product of interest. The lane 4 contains an amount of enzyme of interest LPG1. And finally, in lane 5 we put an amount of the reaction mixture, where we distinguished one band, the highest and biggest one, which migrated at the same level than the

observed band in lane 1. Consequently, this band corresponds to the acceptor, methyl- $\alpha$ -D-mannopyranoside. A second thinner band is observed in the mixture reaction that was eluted at similar level than the band, that corresponds to the disaccharide, maltose, observed in the lane 2. The UDP- $\alpha$ -D-glucopyranoside is most polar molecule in the mixture, so it is the last one to migrate as the shown in the lane 3. However, there is no band in lane 4 since there is only the enzyme LPG1. This proves that the observed band in the lane 5, that migrates at the same level than the maltose corresponds to new product. This product could probably be the result of the transfer of galactofuranose on the Me-Manp performed by LPG1. Similar results were obtained with LPG1L, LPG1R and LPG1G.

In order to make sure that this band corresponds to the product, we evaporated the mixture solution, in order to eliminate the solvent. Then, we peracetylated all the sugars, and extracted them with  $\text{CH}_2\text{Cl}_2$  in order to get rid of salts and facilitate the extraction of the product. Finally, high-resolution accurate mass measurements were performed in positive mode with an ESI source on a Q-TOF mass spectrometer.



**Figure IV-12:** HRMS analysis of peracetylated disaccharide product after LPG1-catalyzed reaction using  $\alpha$ -D-Methyl-mannoside as acceptor and UDP-Galf as sugar donor. The top panel is a zoomed region of the isolated peak corresponding to disaccharide. The bottom panel is the theoretical spectrum obtained with a chemical formula of  $\text{C}_{27}\text{H}_{38}\text{NaO}_{18}$  (Na adduct of peracetylated disaccharide). No peak was detectable in control reaction (without enzyme or one of the reaction components).

High Resolution Mass Spectrometry (HRMS) results shown in Figure IV-12 present the experimental spectra, in the top of the panel, and theoretical spectra, in the bottom panel, obtained with LPG1 catalyzed reaction in the presence of UDP-Galf as a donor and Me-manp as an acceptor. A zoom region indicating the presence of a peak at  $m/z = 673.1950$  that corresponds to the exact

mass of the sodium adducts of the peracetylated and glycosylated methyl  $\alpha$ -D-mannopyranoside product. This result confirms that LPG1 is able to transfer galactofuranose residues from an UDP- $\alpha$ -D-galactofuranoside to me- $\alpha$ -D-mannopyranoside. Similar spectrums were obtained for LPG1L, LPG1R and LPG1G.

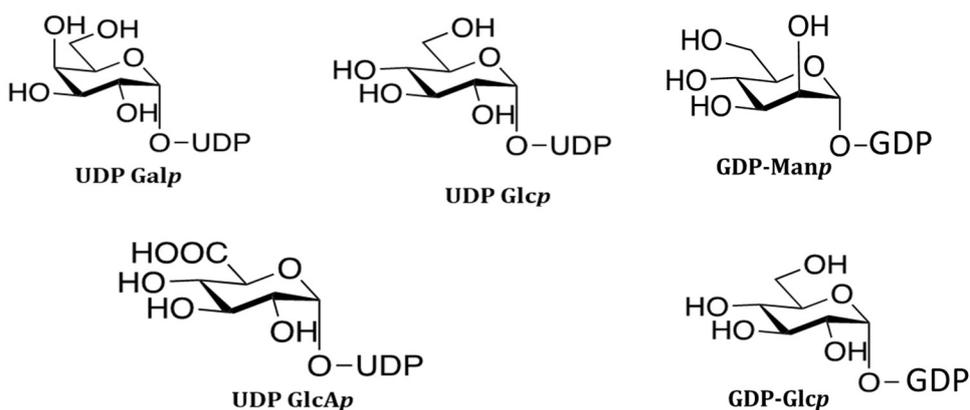
We also tried to carry out nucleic magnetic resonance (NMR) analysis to identify the type of linkage between Gal $f$  and Me-Man. Unfortunately,  $^1\text{H}$ , HMBC or HMQC NMR experiments lead to weak signals, probably due i) to the low amount of purified disaccharide or most probably, ii) to the presence of a mixture of (1-->2, 1-->3, 1-->4 and/or 1-->6) regioisomers.

## B. Substrate specificity

### 1. Donor specificity

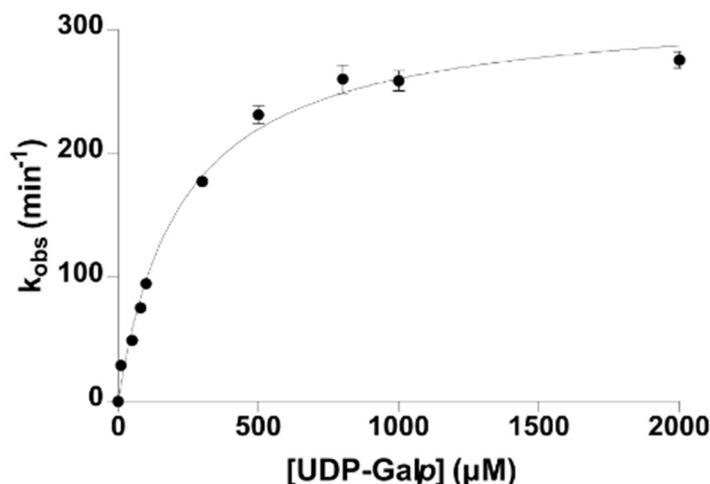
#### a) Screening assays

Some activated sugar donors; UDP-galactopyranos, UDP-glucopyranose, UDP-glucuronic acid, GDP-glucopyranose and GDP-mannopyranose are commercially available (Figure IV-13). We decided then to test them as potential donor for LPG1x galactofuranosyltransferases



**Figure IV-13:** List of NDP-pyranoses tested with LPG1x GalFTs.

The same strategy described before was used. The coupled spectrophotometric assays were performed with each of these NDP-pyranoses for each enzyme. Me-man $p$  was used as acceptor.



**Figure IV-14:** Michaelis-Menten plot that corresponds to UDP-Galp kinetics as donor substrate of LPG1. Reactions were performed in triplicate with 1μg of enzyme and 1mM of Me-Manp.

Five NDP-pyranoses were tested as donors with each of LPG1x GalT<sub>s</sub>, with Me-Manp as acceptor. Spectrophotometric assays were performed and surprisingly, we registered linear decrease of the absorbance at 340nm. It corresponds to the oxidation of NADH, thus to the transfer of pyranosyl residues on Me-Manp performed by LPG1x GalT. Therefore, we tested different concentrations of UDP-pyranoses ranging from 10 to 1000μM Figure IV-14 shows Michaelis-Menten plot of LPG1 activity with UDP-Galp as donor and Me-Manp as acceptor. Thanks to it, it was possible to determine values of  $V_{max}$  and  $K_M$ . Here, only UDP-Galp was accepted by LPG1, no activity was detected in the presence of the other NDP-donors. Similar plots were also obtained with LPG1L, LPG1R and LPG1G, which are available in appendix 7. Here again, LPG1G used only UDP-Galp as a donor, LPG1R accepted only UDP-Glcp, however, LPG1L was able to accept UDPα-D-Galp, UDP α-D-Glcp, GDP α-D-Glcp and GDP α-D-Manp. Thus, unexpectedly, all four of the enzymes were able to use UDP-pyranoses as donor substrates. Kinetics parameters of each reaction were then calculated as described previously.

	UDP-pyranose	Apparent $K_M^a$ (mM)	$k_{cat}^a$ (min <sup>-1</sup> )	$k_{cat}/K_M^a$ (min <sup>-1</sup> mM <sup>-1</sup> )
<b>LPG1</b>	UDP $\alpha$ -D-Galp	0.23 $\pm$ 0.02	320 $\pm$ 22	1,400
	UDP $\alpha$ -D-Glcp	n.d	n.d	n.d
	UDP $\alpha$ -D-GlcA	n.d	n.d	n.d
	GDP $\alpha$ -D-Glcp	n.d	n.d	n.d
	GDP $\alpha$ -D-Manp	n.d	n.d	n.d
<b>LPG1G</b>	UDP $\alpha$ -D-Galp	0.005 $\pm$ 0.001	132 $\pm$ 10	27,978
	UDP $\alpha$ -D-Glcp	n.d	n.d	n.d
	UDP $\alpha$ -D-GlcA	n.d	n.d	n.d
	GDP $\alpha$ -D-Glcp	n.d	n.d	n.d
	GDP $\alpha$ -D-Manp	n.d	n.d	n.d
<b>LPG1L</b>	UDP $\alpha$ -D-Galp	0.05 $\pm$ 0.01	468 $\pm$ 50	9,900
	UDP $\alpha$ -D-Glcp	0.006 $\pm$ 0.001	205 $\pm$ 26	34,700
	UDP $\alpha$ -D-GlcA	n.d	n.d	n.d
	GDP $\alpha$ -D-Glcp	0.043 $\pm$ 0.005	24 $\pm$ 3	500
	GDP $\alpha$ -D-Manp	0.032 $\pm$ 0.005	86 $\pm$ 6	2,860
<b>LPG1R</b>	UDP $\alpha$ -D-Galp	n.d	n.d	n.d
	UDP $\alpha$ -D-Glcp	0.038 $\pm$ 0.005	303 $\pm$ 33	7,994
	UDP $\alpha$ -D-GlcA	n.d	n.d	n.d
	GDP $\alpha$ -D-Glcp	n.d	n.d	n.d
	GDP $\alpha$ -D-Manp	n.d	n.d	n.d

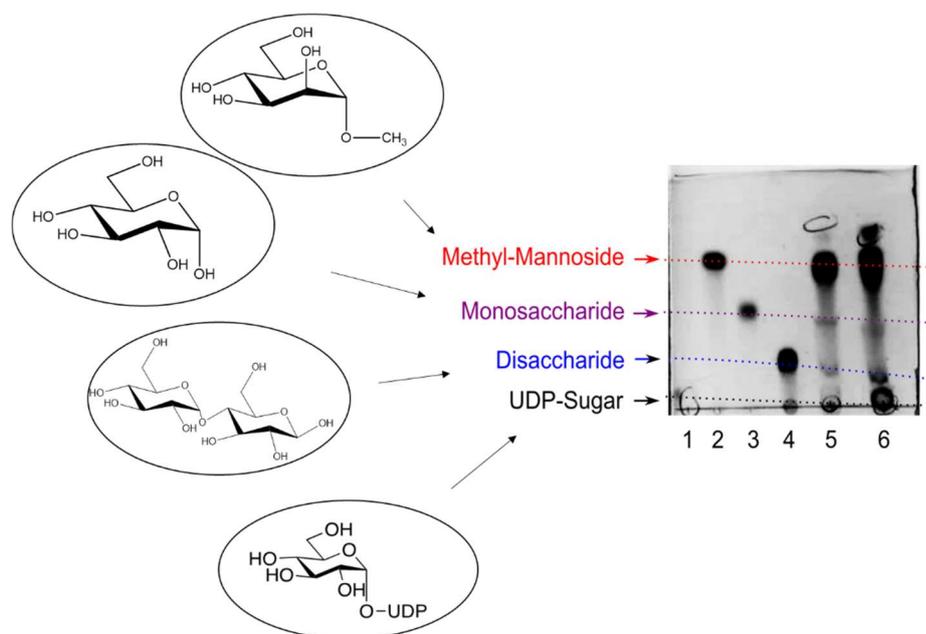
**Table IV-2:** Kinetic parameters of leishmanial GalT's LPG1, LPG1R, LPG1L and LPG1G compared for different NDP-pyranosides donors. <sup>a</sup>: Me-Manp was used as the acceptor.

Assays were performed to screen different NDP-pyranosides donors to see if they were recognized by LPG1x recombinant proteins, in the presence of Me-Manp as acceptor (Table V-2). First, none of them was able to use UDP  $\alpha$ -D-glucuronic acid as a substrate. Second, three proteins, LPG1, LPG1G and LPG1L accept UDP- $\alpha$ -D-Galp with respective  $k_{cat}/K_M$  values of 1,400 min<sup>-1</sup> mM<sup>-1</sup> and 27,978 min<sup>-1</sup> mM<sup>-1</sup>. These values are lower than those obtained with UDP-Galf. This difference is due to the  $k_{cat}$  values, which are lower than  $k_{cat}$  obtained with UDP-Galf. Indeed, the apparent  $K_M$  is still in the sub- $\mu$ M range, even as low as 5  $\mu$ M for LPG1G. LPG1R only recognized UDP- $\alpha$ -D-Glcp a similar kinetic property to the two previous GalT's for the UDP-Galp, *i.e.* a low  $k_{cat}/K_M$  (7,994

min<sup>-1</sup> mM<sup>-1</sup>). Still this is close to a seven-fold increase as compared to UDP-Galf and therefore LPG1R exhibits a better glucopyranosyltransferase than galactofuranosyltransferase activity at least *in vitro*. LPG1L was the most promiscuous enzyme in this respect as it was able to catalyse the reaction with not only UDP- $\alpha$ -D-Galp and UDP- $\alpha$ -D-Glcp but also very surprisingly with GDP- $\alpha$ -D-Manp and GDP- $\alpha$ -D-Glcp even with lower specificity. Once again, with LPG1L GalT the apparent  $K_M$  values were in the tens of millimolar range and the  $k_{cat}$  values were as low as 24 min<sup>-1</sup>. NDP-pyranoses were all recognized in a similar manner but UDP nucleotide sugars led to faster reactions than their GDP counterparts by 50 to 100-fold.

### b) Product characterization

Based on the same parameters described before, 3 mL final volume reaction were prepared containing 25  $\mu$ mol of Me- $\alpha$ -D-Mannopyranoside, 5  $\mu$ mol of NDP-pyranoside and 0.5 mg of leishmanial GalTs. After incubation during 24 hours at 37°C, TLC analysis were carried out.

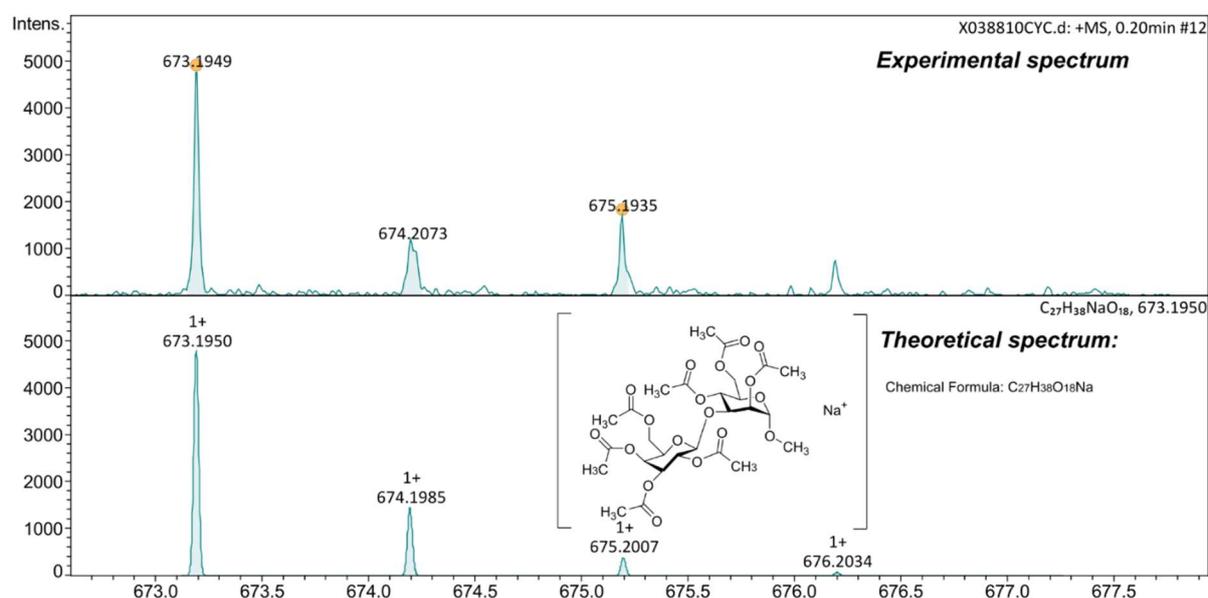


**Figure IV-15:** Typical TLC obtained with the GalT reaction and NDP-Pyranose as donor. Lane 1: UDP-sugar. Lane 2: Methyl- $\alpha$ -D-mannopyranoside. Lane 3: Monosaccharide, D-Glucose. Lane 4: Maltose. Lane 5: LPG1L catalysed reaction with UDP- $\alpha$ -D-galacropyranose as donor and methyl- $\alpha$ -D-mannopyranoside as acceptor. Lane 6: LPG1L catalysed reaction with UDP- $\alpha$ -D-glucopyranose as donor and methyl- $\alpha$ -D-mannopyranoside as acceptor.

In Figure IV-15, the lane 1 presents the spot of UDP-glucopyranose, which is the most polar molecule in the mixture reaction. Lane 2 presents the migration of Me-Manp, the less polar sugar in the mixture, to the top of the silica. It is followed by glucopyranose band, present in lane 2. The presence of glucopyranose amount will allow as to evaluate the degradation of donor of sugar, then the liberation of a monosaccharide. The third band, detected in lane 4, correspond to the maltose, used as reference of a disaccharide product. Lane 5 represents the separation of sugars

present in the reaction mixture that was incubated with LPG1L, UDP-Galp and Me-Manp. We can see the migration of three bands. The highest one is at the same level than Me-Manp, the second one, is similar to the monosaccharide and the last one, corresponds to the disaccharide product. Similar results are observed in lane 6 which represents the analysis of LPG1L catalysed-reaction with UDP-Glcp and Me-Manp. This confirms that LPG1L is able to use both UDP-Glcp and UDP-Galp as donors. During these experiments, UDP-donors were not stable. Similar results were obtained with LPG1, LPG1R and LPG1G.

In similar manner used with UDP-Galf-reactions, sugars were peracetylated, extracted and analysed with HRMS.



**Figure IV-16:** HRMS analysis of peracetylated disaccharide product after LPG1-catalyzed reaction using  $\alpha$ -D-Methylmannoside as acceptor and UDP-Galp as sugar donor. The top panel is a zoomed region of the isolated peak corresponding to disaccharide. The bottom panel is the theoretical spectrum obtained with a chemical formula of  $C_{27}H_{38}NaO_{18}$  (Na adduct of peracetylated disaccharide). No peak was detectable in control reaction (without enzyme or one of the reaction components).

Here, we detected the presence of a peak at  $m/z = 673.1950$ , that corresponds to the exact mass of the sodium adducts of the peracetylated and glycosylated methyl  $\alpha$ -D-mannopyranoside product. The experimental spectrum is similar to the theoretical one. Same spectra were obtained with the four LPG1x. This confirms once again that our four leishmanial GalT's are able to accept NDP-pyranosides as donors. However, despite several trials of  $^1H$ , HMBC or HMQC NMR, we did not manage to detect signal of our products, probably due to the same reasons exposed with Galf-reactions, i) to the low amount of disaccharide or most probably, ii) to the presence of a mixture of (1 $\rightarrow$ 2, 1 $\rightarrow$ 3, 1 $\rightarrow$ 4 and/or 1 $\rightarrow$ 6) regioisomers.

With these experiments we demonstrated that the four GalT<sub>f</sub>s are versatile. They are able to transfer not only galactofuranose moieties on Me-Man<sub>p</sub>, but also act with other NDP-pyranose donors. Some GTs have been reported to be versatile. One example can be cited, *N*-acetylglucosamine β1,4-galactosyltransferase (GalT). It catalyzes the transfer of galactose moieties from UDP-Gal to GlcNAc residues. It was reported its abilities to recognize also two other donors, UDP-Glc<sub>p</sub> and UDP-GlcNAc.<sup>38</sup> However, it has never been reported that UDP-pyranose donors can be used as donors with galactofuranosyltransferases. Here, the fact that LPG1L is able to recognize various NDP-donors this makes it the most promiscuous natural and characterized GT reported to date in term of the donor. This finding also underscores this protein as a promising tool for glycorandomization, at least for transferring carbohydrate residues to α-Man<sub>p</sub>-containing acceptors.<sup>37,39</sup>

## 2. Acceptors specificity

### a) Screening of hexoses as acceptors

A screening of new acceptors was performed using the same spectrophotometric assays described before. The acceptors that were tested and showed in appendix 8 are: Monosaccharide as D-Glc, D-GlcNH<sub>2</sub>, D-GlcNAc, Me-α-D-Glc, Oct-α-D-Man and disaccharide such as D-Maltose, D-Lactose and D-Melibiose. Different NDP-pyranoses donors and UDP-Gal<sub>f</sub> were used with their corresponding enzymes. However, no activity was detected meaning that these acceptors do not fit with LPG1x GalT<sub>f</sub>s.

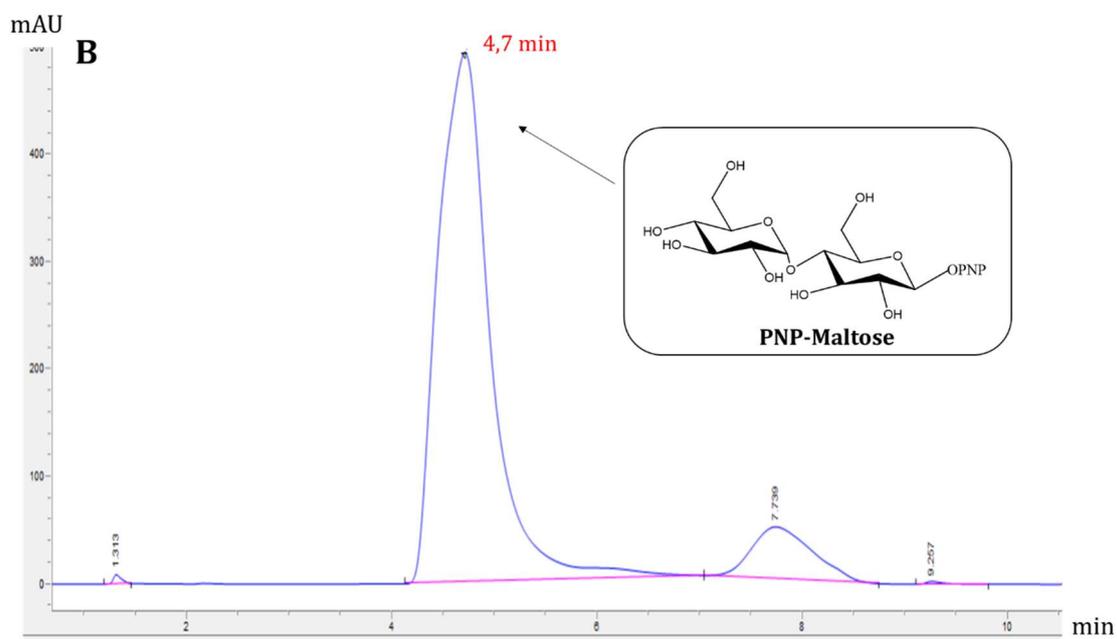
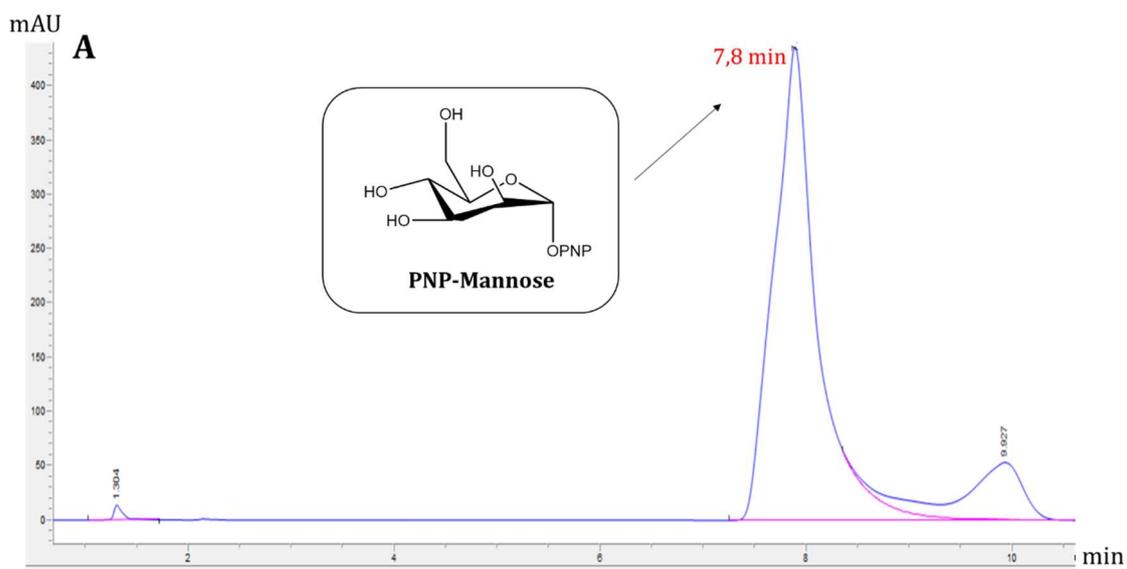
### b) Screening of PNP-sugars as acceptors

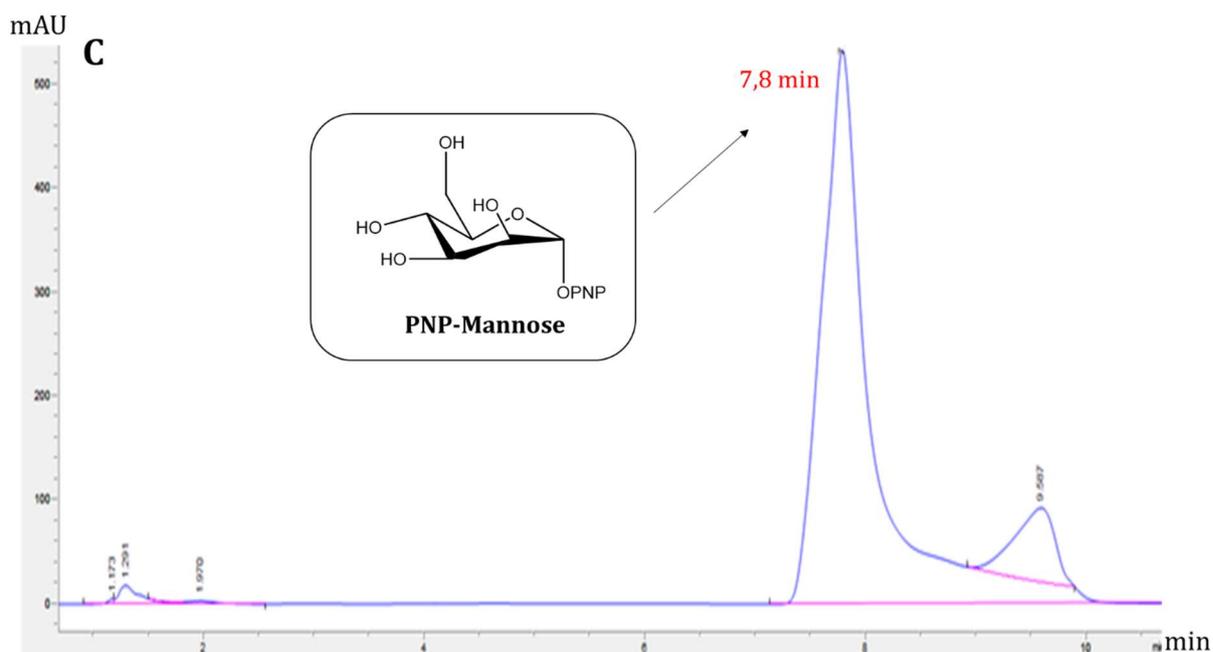
We decided to test *p*NP-sugars as potential substrates. The advantages of the use of *p*NP-sugars, is that they are inexpensive and there are easy to detect. Indeed, the *p*NP-sugars can be detected at 305nm (Figure IV-17). Most of the time, *p*NP-glycoses are used to characterize glycosidases.<sup>40,41,42</sup> They are very well recognized as substrates. So, based on this we decided to test different *p*NP-sugars, described in Appendix 9.

200μl final volumes of mixture reaction containing, GalT<sub>f</sub>, NDP-sugar donors and *p*NP-sugar as acceptor was prepared. Reactions were incubated at 37°C for 2, 4, 7 and 24 hours, then the reaction is stopped by the addition of 100 μL of Na<sub>2</sub>CO<sub>3</sub> for a final volume of 300μL. The reaction mixture was injected and analyzed by High Performance Liquid Chromatography. The purpose of this experiment is to detect PNP-glycosides used as acceptor and *p*NP-Disaccharides, which should be the product, separately.



**Figure IV- 17:** The hypothetical reaction of leishmanial GalTs with PNP-glycosides as acceptors, resulting the formation of pNP-disaccharide. pNP-glycosides are detected at 305nm.





**Figure IV-18:** HPLC analysis of reaction mixture containing 0.1 $\mu$ M of LPG1, 1mM of UDP-pyranoside and 1mM of *p*NP-Mannose. A: HPLC Trace of *p*NP-Mannose solution. B: HPLC Trace of *p*NP-Maltose solution. C: Chromatogram of the LPG1 reaction mixture.

The Figure IV-18 shows typical chromatograms of *p*NP sugars used for our enzymatic assays using C18 column (150 X 6,4mm; 3.5 $\mu$ m). The mobile phases employed for the separation of *p*NP sugars was composed of water + 0.1% TFA (solvent A) and acetonitrile + 0.1% TFA (solvent B) and the temperature was set at 30°C. In figure 16-A, we detected *p*NP-Mannose, which was used as acceptor candidate for LPG1 and its retention time is 7.8min. The retention time corresponds to the time that components spend into the column after being injected. It depends of the chemical formula of molecules. In figure 18-B the chromatograph of *p*NP-Maltose, used as reference for *p*NP-disaccharide. Its retention time is 4.7min. Nevertheless, when we injected the reaction mixture of GalT's reactions with both acceptor and donor, only the donor, *p*NP-Mannose was registered with a retention time of 7.8 min. Similar results were observed with all *p*NP-sugars for the different combinations, using the different donors corresponding to each protein. These results indicate that LPG1x do not accept *p*NP-sugars as acceptors. Thus, they are probably specific to Me- $\alpha$ -D-Mannose, used as acceptor.

### 3. Temperature and pH effect

The coupled spectrophotometric assays allowed us to identify the substrates of LPG1x and their kinetics. However, it is not adapted to test the effect of pH, temperature and ions on the activity of enzymes due to the presence of PK/LDH that need specific conditions, i.e. pH8 and 37°C for their activity. So, we decided to test another technique that was described, the phosphatase-coupled glycosyltransferase method using malachite green assay provided by Sigma Aldrich to explore

further the optimum conditions for our leishmanial LPG1x. This strategy allowed us to run different reactions with the LPG1x and their respective donors and acceptors independently and prior to the malachite green assays (Figure IV-7) Thus, it was possible to test different temperature and pH during the reaction incubation time.

Unfortunately, when we used the malachite green assay, we did not manage to quantify the released Pi. A high absorbance values were obtained that was probably due to some components in the mixture solution that interfere with this test. Moreover, the presence of magnesium that is needed for the activity of LPG1x form probably phosphate salt, which result precipitation that affect the Pi detection. Up to date, we could not do further test, but optimization of this technique should be done in order to be used with LPG1X enzymes.

### III. Crystallization of LPG1x

Getting proteins crystals is required for solving their three-dimension structures. Indeed, the characterization of proteins is not only based on the determination of their enzymatic activity and kinetic parameters but also on their structures.<sup>43</sup> In biology, structure is always closely related to function, so these studies provide insights into the molecular functioning of living systems. It is important to highlight the fact that the classification of proteins in CAZy database is based on their structural homology and activity. Over these last 30 years, the number of three-dimensional structures in the Protein Data Bank (PDB) increased. Today, there is about 140000 available structure that belong to wild type proteins, mutants, and proteins with their substrates. Up to date, only the structure of the mycobacterial galactofuranosyltransferase, GlfT2, is available on PDB accession code 4FIX. It belongs to GT2 family and uses inverting mechanism. GlfT1 was also classified in the same family. However, GfsA from *Aspergillus* and bacterial WbbI are predicted to belong to GT 31 and GT4 respectively. All of them are supposed to adopt inverting mechanism, however, they do not share significant homology of their amino acid sequences. LPG1x are expected to belong to GT40. They are not closely related to any of the other GalTs, so obtaining crystals will allow us to identify first the first time their structural configuration. It is essential for the identification of amino acid involved in their catalytical activity and the interaction with the substrate.

In order to obtain protein crystals, it is important to have good yield of pure proteins. In our cases, we were able to get around 5mg/L of culture of LPG1-MBP and LPG1R-MBP, and 10 mg/L of culture for LPG1L-MBP and LPG1G-MBP. So, the four LPG1x were purified from 2 liters of culture each time, which have been induced at 25°C for 2 hours and lysed with the protocol 1 (See pages 103-104). Proteins of interest were purified on maltose affinity chromatography followed by size exclusion chromatography using superdex™ 200 column. Finally, LPG1x were washed and concentrated several times with 50mL concentrators tube with molecular-weight cutoffs of 30K

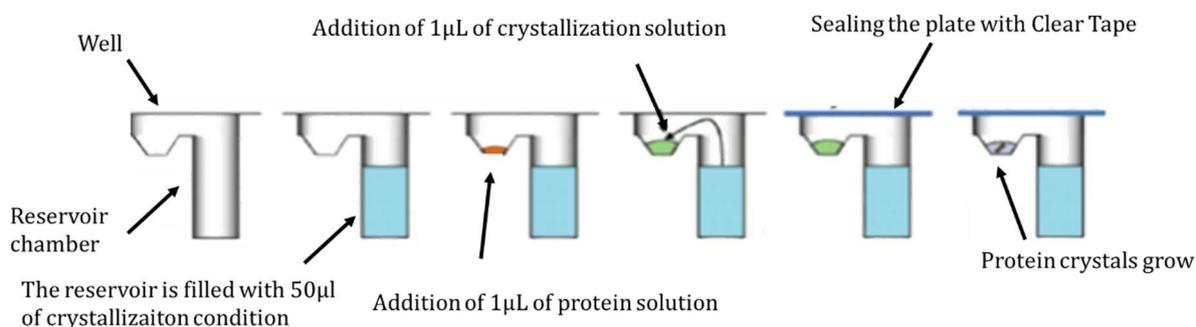
MWCO. This final step is important because it enables the desalting and exchange of superdex buffer which is highly concentrated on salt with 20 mM of tris buffer. It is essential to eliminate salts because they can easily crystallize. Once proteins were stocked in 20mM Tris buffer and pH 8, the final concentration should be ranging from 1 to 15mg/ml.

Different concentrations of the four proteins were tested:

	LPG1-MBP	LPG1L-MBP	LPG1R-MBP	LPG1G-MBP
<b>Protein concentration (mg/ml)</b>	7	2	2	2
	13	15	6	8

**Table IV- 3:** Concentration of LPG1x used for crystallogensis trials.

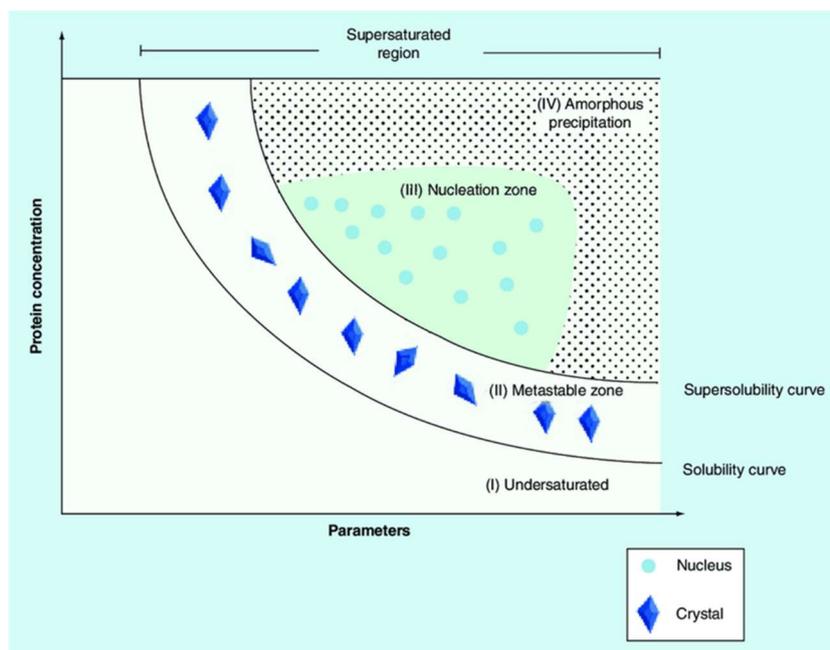
Different crystallisation conditions were screed for each condition. HTSI and HTSII crystallization kits from Jena Bioscience were used. Each kit contains 96 different conditions. First, the reservoir of specific plate was filled by 50µl of each crystallization conditions. Then, 1 µl of each protein was added to the well chamber with 1µl of crystallization condition. Finally, the whole plate was sealed with clear tape (Figure IV-19). The plate was stored either at room temperature or 4°C. Crystals will be found at the bottom of the well after hours, days or weeks of incubation.



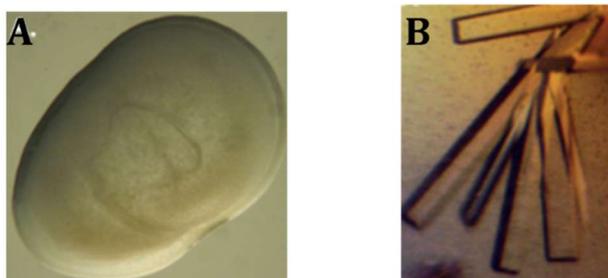
**Figure IV-19:** Different step of protein crystallogensis protocol.

Formation of protein crystals depends of many parameters. The first important one, is having high purity in a homogeneous and low polydispersity state of the protein. Also, it depends of precipitant or additive concentration, pH and temperature. In our crystallization conditions, the concentration of salt is higher than in the drop containing 1µl of protein solution and 1µl crystallization condition. This result the transfer of solvent molecule from our drop mixture to the reservoir chamber by vapour diffusion in the gas phase. The protein will reach thermodynamically unstable state, know and became supersaturated. This step is critical because during the transfer, the solubility of the protein decrease. Here, I hope that our proteins reach the “right” state that

will allow the formation of crystals. Indeed, there is two possibilities, proteins will form either crystal nuclei that finally grows into larger protein crystals, which is metastable state, or they will precipitate as amorphous proteins, which is useless of course (Figure IV-20).<sup>44,45</sup>

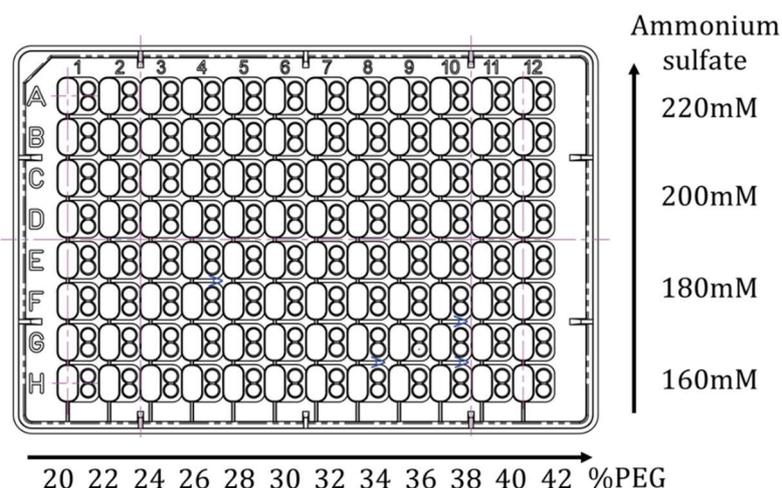


**Figure IV-20:** Different state that protein could reach during crystallization process resulted by vapor diffusion in gas state.



**Figure IV-21:** Two forms that proteins could adopt after reaching supersaturated state; (A) precipitation and amorphous state and (B) Metastable state resulting the formation of protein crystals.

In our case, after several unsuccessful trials, LPG1-MB formed crystals with 13 mg/ml of protein (Figure IV-21) Crystals appeared only few hours after incubation at 4°C. The best condition of the media that resulted in the formation of crystals contains 30%PEG 8000 and 200 mM of ammonium sulfate. We decided then to optimize this condition. New crystallization plate was used as described below in Figure IV-22.



**Figure IV-22:** Optimization strategy of crystallization condition using 20-42% of PEG and 160 – 200Mm of ammonium sulfate.

Many crystals were obtained in most of the wells (Figure IV-23). We picked some of them and freeze them in liquid nitrogen and finally analysed by X-ray diffraction at Proxima1 beamline in Synchrotron of Paris Saclay. Unfortunately, none of these proteins diffracted. At least, we can be sure that they are protein crystals and not salt crystals, because salt can diffract with specific signals.



**Figure IV-23:** LPG1 crystals stained with methylene blue.

## IV. Conclusion

Despite the natural occurrence of the galactofuranose in many pathogenic microorganisms, the pathways involved in its biosynthesis remain poorly understood. This is due, in part, to a lack of knowledge of the corresponding enzymes involved in its incorporation into glycoconjugates (mutases, transferases, transporters, hydrolases). The tedious synthesis of the required donor substrate, UDP- $\alpha$ -D-Galf, is another barrier. Indeed, only five GalT's have been described in the literature, with GlfT2 from *M. tuberculosis* being the only one fully characterized. In *L. major*, four putative genes were expected to code for functional GalT's. After obtaining pure recombinant proteins LPG1, LPG1L, LPG1R and LPG1G, with good yields of 5–10 mg/L of culture, enzymatic assays were carried out in order to characterize these putative GalT's. Lowary and co-workers described coupled spectrophometric assays that was used to characterize GlfT2. Thus, based on this, we decided to use the same strategy. Thanks to Mickaelis-Menten plot, we were able to calculate kinetics of donors of each protein. We were able to report the characterization of four leishmanial galactofuranosyltransferases, that belong to CAZy GT family 40, for the first time. We also succeeded in demonstrating their properties as GalT substrates, with  $k_{cat}/K_M$  values as much as 300-fold better than for the mycobacterial GlfT2. In addition to UDP- $\alpha$ -D-Galf, these enzymes proved in vitro to be able to use some NDP-pyranoses as substrates. Indeed, LPG1 LPG1G and LP1R are able to accept UDP-Galp, the pyranose form of UDP-Galf. LPG1L is the most versatile GalT because it interacts with UDP-Glcp, GDP-Manp and GDP-Glcp. This indicated that they are among the most promiscuous natural glycosyltransferases to date. However, the four enzymatic assays revealed that these four GalT's interact with one acceptor, Me-Manp. These unique biocatalysts also proved to be stable and robust for days and can now serve for the chemo-enzymatic incorporation of Galf moiety into complex glycoconjugates. It is also expected that the four eukaryotic GalT's from *L. major* described herein will serve as targets for the development of new drugs against leishmaniasis.

After the obtention of pure and active proteins, we decided to carry out crystallogenesis trials to get protein crystals. This experiment is indispensable to have access to their three-dimension structure. Several crystals were obtained, however none of them diffracted light during X-ray diffraction experiments carried out at Proxima 1 at Synchrotron of Paris Saclay. At this level, we were not able to determine the structure of our proteins, we need to optimize crystallization condition. According to PDB data base, 139 crystal structures of proteins fused with MBP-Tag were registered over 144211 of total structures (<https://www.rcsb.org/> September 2018). The number of MBP fusion proteins structures is growing. This is an evidence that this strategy is used more often, and it is expected that the number of such structures will continue to increase.<sup>46</sup> Yet, these results are promising.



## References

1. Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97–130 (1993).
2. Ohtsubo, K. & Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126**, 855–867 (2006).
3. Rose, N. L. *et al.* Expression, Purification, and Characterization of a Galactofuranosyltransferase Involved in Mycobacterium tuberculosis Arabinogalactan Biosynthesis. *J. Am. Chem. Soc.* **128**, 6721–6729 (2006).
4. Komachi, Y. *et al.* gfsA encodes a novel galactofuranosyltransferase involved in biosynthesis of galactofuranose antigen of O-glycan in Aspergillus nidulans and Aspergillus fumigatus. *Mol. Microbiol.* **90**, 1054–1073 (2013).
5. Rose, N. L. *et al.* Development of a coupled spectrophotometric assay for GlfT2, a bifunctional mycobacterial galactofuranosyltransferase. *Carbohydr. Res.* **343**, 2130–2139 (2008).
6. Janos, P., Kozmon, S., Tvaroska, I. & Koca, J. How Mycobacterium Tuberculosis Galactofuranosyltransferase GlfT2 Generates Alternating beta-(1-6) and beta-(1-5) linkages: QM/MM Molecular Dynamics Study of the Chemical Steps. *Chemistry (Easton)*. **2**, 7051–7059 (2018).
7. Ant, T. H., Seblova, V., Jecna, L., Beverley, S. M. & Volf, P. Leishmania major Glycosylation Mutants Require Phosphoglycans (lpg2 2) but Not Lipophosphoglycan (lpg1 2) for Survival in Permissive Sand Fly Vectors. *PLoS Negl. Trop. Dis.* **4**, 1–7 (2010).
8. Tefsen, B., Ram, A. F. J., Van Die, I. & Routier, F. H. Galactofuranose in eukaryotes: Aspects of biosynthesis and functional impact. *Glycobiology* **22**, 456–469 (2012).
9. Lee, R. *et al.* Enzymatic Synthesis of UDP-Galactofuranose and an Assay for UDP-Galactopyranose Mutase Based on High-Performance Liquid Chromatography 1. *Ann. Biochem.* **7**, 1–7 (1996).
10. Tsvetkov, Y. E. & Nikolaev, A. V. The first chemical synthesis of UDP- $\alpha$ -D-galactofuranose. *J. Chem. Soc. Perkin Trans. 1* **1**, 889–891 (2000).
11. Marlow, A. L. & Kiessling, L. L. Improved Chemical Synthesis of UDP-Galactofuranose. *Org. Lett.* **3**, 2517–2519 (2001).
12. Euzen, R., Ferrières, V. & Plusquellec, D. General one-step synthesis of free hexofuranosyl 1-phosphates using unprotected 1-thioimidoyl hexofuranosides. *J. Org. Chem.* **70**, 847–855 (2005).
13. Peltier, P., Euzen, R., Daniellou, R., Nugier-Chauvin, C. & Ferrières, V. Recent knowledge and innovations related to hexofuranosides: structure, synthesis and applications. *Carbohydr. Res.* **343**, 1897–1923 (2008).
14. Peltier, P., Guégan, J. P., Daniellou, R., Nugier-Chauvin, C. & Ferrières, V. Stereoselective chemoenzymatic synthesis of UDP-1,2-cis-furanoses from  $\alpha$ , $\beta$ -furanosyl 1-phosphates. *European J. Org. Chem.* **35**, 5988–5994 (2008).
15. Liu, Z., Zhang, J., Chen, X. & Wang, P. G. Combined biosynthetic pathway for de novo production of UDP-galactose: Catalysis with multiple enzymes immobilized on agarose beads. *ChemBioChem* **3**, 348–355 (2002).
16. Errey, J. C., Mukhopadhyay, B., Kartha, K. P. R. & Field, R. A. Flexible enzymatic and chemo-

- enzymatic approaches to a broad range of uridine-diphospho-sugars. *Chem. Commun.* **23**, 2706–2707 (2004).
17. Rose, N. L. *et al.* Development of a coupled spectrophotometric assay for GlfT2 , a bifunctional mycobacterial galactofuranosyltransferase. *Carbohydr. Res.* **343**, 2130–2139 (2008).
  18. Anumula, K. R. New high-performance liquid chromatography assay for glycosyltransferases based on derivatization with anthranilic acid and fluorescence detection. *Glycobiology* **22**, 912–917 (2012).
  19. Gosselin, S., Alhussaini, M., Streiff, M. H., Takabayashi, K. & Palcic, M. M. A continuous spectrophotometric assay for glycosyltransferases. *Anal. Biochem.* **220**, 92–97 (1994).
  20. Ainsworth, S. & MacFarlane, N. A kinetic study of rabbit muscle pyruvate kinase. *Biochem. J.* **131**, 223–36 (1973).
  21. Joseph, R. L., Watts, D. C. & Baldwin, E. Organ Specificity and Lactate -Dehydrogenase Activity. *Biochem. J.* **87**, 416–422 (1963).
  22. Specificity, N. & Kinase, O. F. P. Y. R. U. V. A. T. E. Nucleotide specificity of pyruvate kinase and phosphoenolpyruvate carboxykinase. *Biochem. Biophys. acta* **452**, 406–412 (1976).
  23. Wu, Z. L., Ethen, C. M., Prather, B., MacHacek, M. & Jiang, W. Universal phosphatase-coupled glycosyltransferase assay. *Glycobiology* **21**, 727–733 (2011).
  24. Ethen, C. M. *et al.* Nonradioactive Glycosyltransferase and Sulfotransferase Assay to Study Glycosaminoglycan Biosynthesis. *Methods Mol Biol.* **1229**, 431–441 (2015).
  25. Welner, D. H. *et al.* Plant cell wall glycosyltransferases : High- throughput recombinant expression screening and general requirements for these challenging enzymes. *PLoS One* **12**, 1–20 (2017).
  26. Sheikh, M. O. *et al.* Rapid screening of sugar-nucleotide donor specificities of putative glycosyltransferases. *Glycobiology* **27**, 206–212 (2017).
  27. Schmid, J., Heider, D., Wendel, N. J., Sperl, N. & Sieber, V. Bacterial glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products. *Front. Microbiol.* **7**, 1–7 (2016).
  28. Bird, I. M. High performance liquid chromatography: principles and clinical applications. *Br Med J* **299**, 783–787 (1989).
  29. Meek, J. L. Prediction of peptide retention times in high-pressure liquid chromatography on the basis of amino acid composition. *Proc. Natl. Acad. Sci.* **77**, 1632–1636 (1980).
  30. Petrova, O. E., Sauer, K. & Biofilm, B. High-performance liquid chromatography (HPLC)-based detection and quantitation of cellular c-di-GMP. *Methods Mol Biol.* **1657**, 33–43 (2017).
  31. Brown, P. R. Current high performance liquid chromatographic methodology in analysis of nucleotides, nucleosides, and their bases. II. *Cancer Invest.* **1**, 527–536 (1983).
  32. Galea, C., Mangelings, D. & Vander Heyden, Y. Characterization and classification of stationary phases in HPLC and SFC - a review. *Anal. Chim. Acta* **886**, 1–15 (2015).
  33. Perrett, D. Some applications of high-performance liquid chromatography in the field of nucleotides, nucleosides and bases. *Biochem. Soc. Trans.* **13**, 1067–70 (1985).
  34. Yamatsugu, K., Splain, R. A. & Kiessling, L. L. Fidelity and Promiscuity of a Mycobacterial Glycosyltransferase. *J. Am. Chem. Soc.* **138**, 9205–9211 (2016).

35. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495 (2014).
36. Engel, J., Schmalhorst, P. S., Dörk-Bousset, T., Ferrières, V. & Routier, F. H. A single UDP-galactofuranose transporter is required for galactofuranosylation in *Aspergillus fumigatus*. *J. Biol. Chem.* **284**, 33859–33868 (2009).
37. Matysik, E. *et al.* The New TLC Method for Separation and Determination of Multicomponent Mixtures of Plant Extracts. *J. Anal. Methods Chem.* **2016**, (2016).
38. Palcic, M. M. & Hindsgaul, O. in the Synthesis of Complex Carbohydrates. *Glycobiology* **1**, 205–209 (1991).
39. Chang, A., Singh, S., Phillips, G. N. & Thorson, J. S. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Curr. Opin. Biotechnol.* **22**, 800–808 (2011).
40. Michikawa, M. *et al.* Structural and biochemical characterization of glycoside hydrolase family 79  $\beta$ -glucuronidase from *Acidobacterium capsulatum*. *J. Biol. Chem.* **287**, 14069–14077 (2012).
41. Placier, G., Watzlawick, H., Rabiller, C. & Mattes, R. Evolved  $\beta$ -galactosidases from *Geobacillus stearothermophilus* with improved transgalactosylation yield for galacto-oligosaccharide production. *Appl. Environ. Microbiol.* **75**, 6312–6321 (2009).
42. Gao, J. & Wakarchuk, W. Characterization of five  $\beta$ -glycoside hydrolases from *Cellulomonas fimi* ATCC 484. *J. Bacteriol.* **196**, 4103–4110 (2014).
43. Blundell, T. L., Jhoti, H. & Abell, C. High-Throughput Crystallography for Lead Discovery in Drug Design. *Nat. Rev. Drug Discov.* **1**, 45–54 (2002).
44. Chayen, N. E. & Saridakis, E. Protein crystallization: From purified protein to diffraction-quality crystal. *Nat. Methods* **5**, 147–153 (2008).
45. Burgess, R. R. A brief practical review of size exclusion chromatography: Rules of thumb, limitations, and troubleshooting. *Protein Expr. Purif.* **150**, 81–85 (2018).
46. Waugh, D. S. Crystal structures of MBP fusion proteins. *Protein Sci.* **25**, 559–571 (2016).

***Chapter V: Characterization of  
eukaryotic LPG1x***

## Abstract

LPG1 protein was reported to be located in the membrane of Golgi apparatus of *Leishmania major*. Thus, it is likely that the GalT's are glycosylated. Indeed, as a eukaryotic species, *Leishmania* have a glycosylation machinery that ensure the posttranslational modification of proteins, which is highly important for their stability, structure and activity. It is then important to obtain similar glycosylated LPG1x and evaluate the effect of the glycans on their activities. To do that, Leishmania expression system (LEXSY) that was developed by Jena Bioscience was adopted to produce secreted and glycosylated proteins. It is based on the use of non-pathogenic *Leishmania tarentolae* as a host that use similar mechanism to express and glycosylate proteins that *Leishmania major*. Thus, in this chapter, we present the strategies used to produce glycosylated LPG1x, and the preliminary result of the glycosylated LPG1G activity.

## Résumé

La protéine LPG1 a été localisée dans la membrane de l'appareil de Golgi de *Leishmania major*. Ainsi, il est probable que les GalT's soient glycosylées. En effet, les parasites, *Leishmania*, possèdent une machinerie de glycosylation qui assurent les modifications post traductionnelles des protéines, qui sont considérés comme importantes pour leurs stabilités, leurs structures et leurs activités. Il est alors important d'obtenir des LPG1x glycosylés et évaluer l'effet des glycanes sur leurs activités. Pour ce faire, le système d'expression de *Leishmania* (LEXSY) qui a été développé par Jena Bioscience, a été adopté pour produire des protéines sécrétées et glycosylées. Il est basé sur l'utilisation de *Leishmania tarentolae*, un parasite non pathogène pour l'Homme, en tant qu'hôte utilisant le même mécanisme pour exprimer et glycosyler les protéines que *Leishmania major*. Ainsi, dans ce chapitre, nous présentons les stratégies utilisées pour produire LPG1x glycosylés, ainsi que le résultat préliminaire de l'activité LPG1G glycosylée.

I.	Recombinant proteins produced in eukaryotic cells .....	157
A.	Glycosylation of proteins.....	157
B.	<i>Leishmania tarentolae</i> as an expression host system.....	160
1.	<i>Leishmania tarentolae</i> .....	160
2.	LEXSY vector .....	161
II.	Glycosylated recombinant GalT's of <i>Leishmania</i> .....	163
A.	Cloning and transfection .....	163
1.	Cloning.....	163
2.	Transfection and selection.....	166
B.	Protein expression and characterization.....	169
1.	Protein Purification .....	169
2.	Immunodetection .....	171
3.	Glycosylation.....	173
4.	Activity characterization.....	175
III.	Conclusion .....	178

# I. Recombinant proteins produced in eukaryotic cells

## A. Glycosylation of proteins

Posttranslational modifications (PTMs) are one of the most common and complex modifications that occur during protein synthesis in eukaryotes. Despite few exceptions, secreted and membrane-associated protein expressed by eukaryotic systems are glycosylated.<sup>1,2</sup> These glycosylation phenomena are conserved through all eukaryotic species, and have important effects on three dimensional conformation, folding, stability, activity and proteolysis of proteins.<sup>3,4,5,6,7</sup> One simple example can be cited to demonstrate the importance of glycosylations; monoclonal recombinant antibodies that are key actors in immune system are widely produced for analytical, diagnosis or therapeutic use. The mABs are glycosylated, and this play important role in their stability and folding.<sup>8</sup> Moreover, deglycosylated mABs are susceptible to aggregate and present less thermal stability.<sup>9</sup>

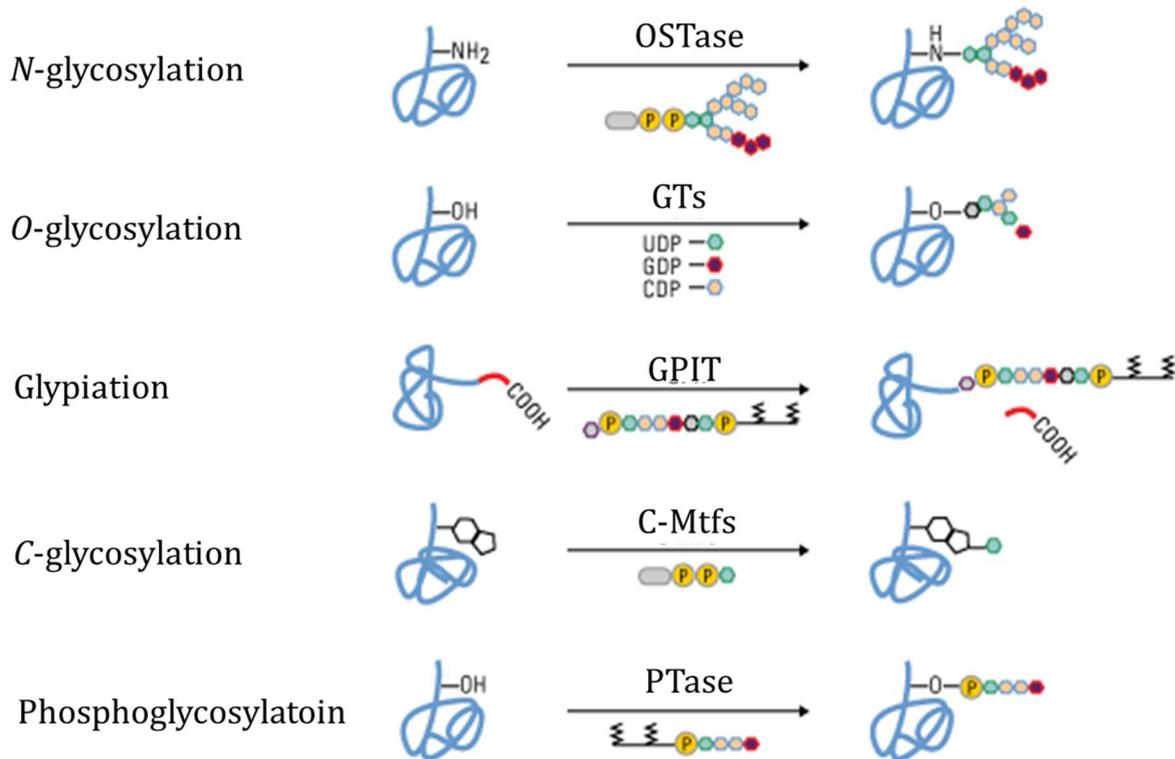
Glycans are attached to proteins in some specific sites by different glycosylation process (Figure V-1):

- *N-glycosylation* consist of the attachment of glycans to a nitrogen atom of asparagine. Specific consensus sequences *Asn-X-Ser/Thr* (Asparagine-any amino acid-Serine/Threonine) was reported to be frequently *N-glycosylated*.<sup>10</sup>
- *O-glycosylation* is characterized by the addition of only one sugar residue to the hydroxyl group of threonine or serine.<sup>6</sup>
- *C-glycosylation* consist of the attachment of sugars to a carbon of tryptophan side chain. The most characterized is *C-mannosylation* reaction which consists on the linkage between the first carbon of mannose to the second atom carbon of the indole ring of Trp. This reaction is highly described in mammalian cells and specific consensus sequence, *Trp-X-X-Trp* was identified.<sup>11</sup>
- *Phosphoglycosylation* is the enzymatic attachment of glycans to the polypeptide chain through a phosphodiester bridge. These glycosylation has been identified in many polysaccharide at the surface of *Leishmania* in lipophosphoglycan.<sup>12</sup>
- *Glypiation* is the addition of glycosylphosphatidylinositol to the *C-termini* of proteins.<sup>1</sup>

However, the same glycosylation site can accommodate different glycans due to the high number of different sugar moieties such as GlcNac, Glc and Rha can be added by *N-glycosylation*, fucose, mannose, xylose, galactose, arabinose can be attached by *O-glycosylation*, in addition to and the multitude of possible linkages and configuration *i.e.*  $\alpha$  or  $\beta$  are possible.<sup>1</sup> The presence of glycans in higher eukaryotic is considered as a sign of evolution and provide the ability to new structures without presenting changes into the genetic heritage.<sup>13</sup> This glycans generate various glycoforms

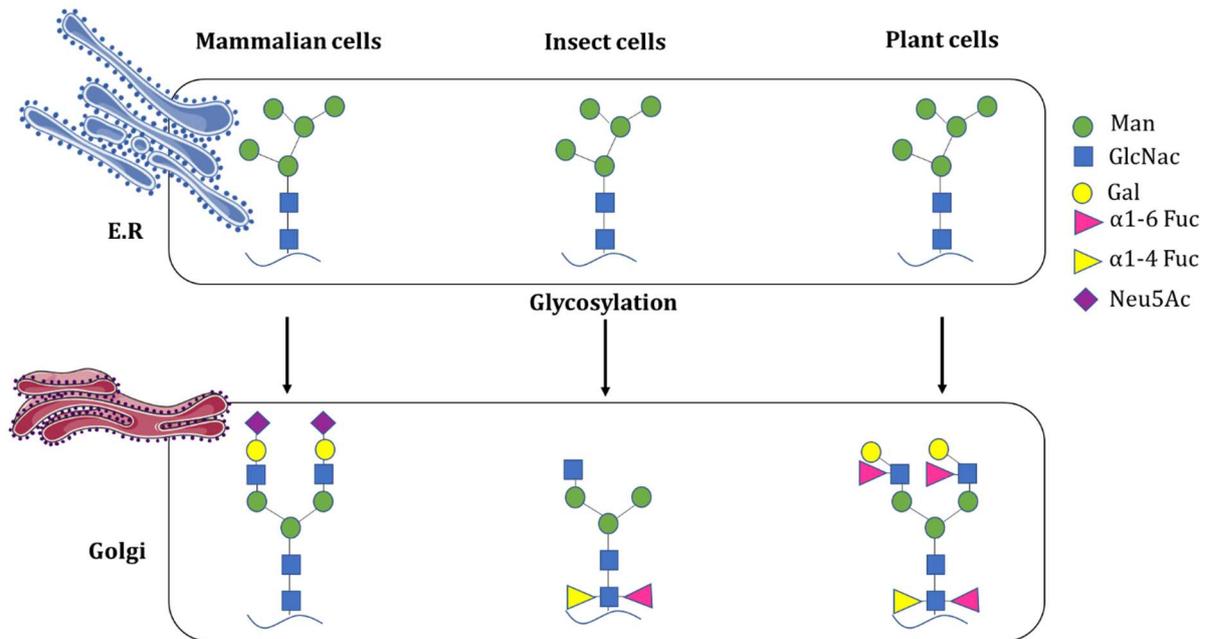
of proteins that it is referred to as site heterogeneity and can affect the activity and stability of proteins.<sup>14</sup> This heterogeneity make the study and the analysis of glycoproteins hard and arduous.

It is important to highlight the fact that glycosylation cannot be predicted from RNA template such as proteins sequences. In fact, protein glycosylation is ensured by an assembly of glycosyltransferases, glycoside hydrolases and many nucleotide transporters. This occurs during passaging the endoplasmic reticulum (ER) and Golgi network (Figure V-1).<sup>15</sup>



**Figure V-1:** Different type of glycosylation that occur in eukaryotic species. OSTase: oligosaccharyltransferases; GT: Glycosyltransferases; GPIT: glycosylphosphatidylinositol transferases; C-Mtfs: C-mannosyltransferases; PTase: Phosphosyltransferases.

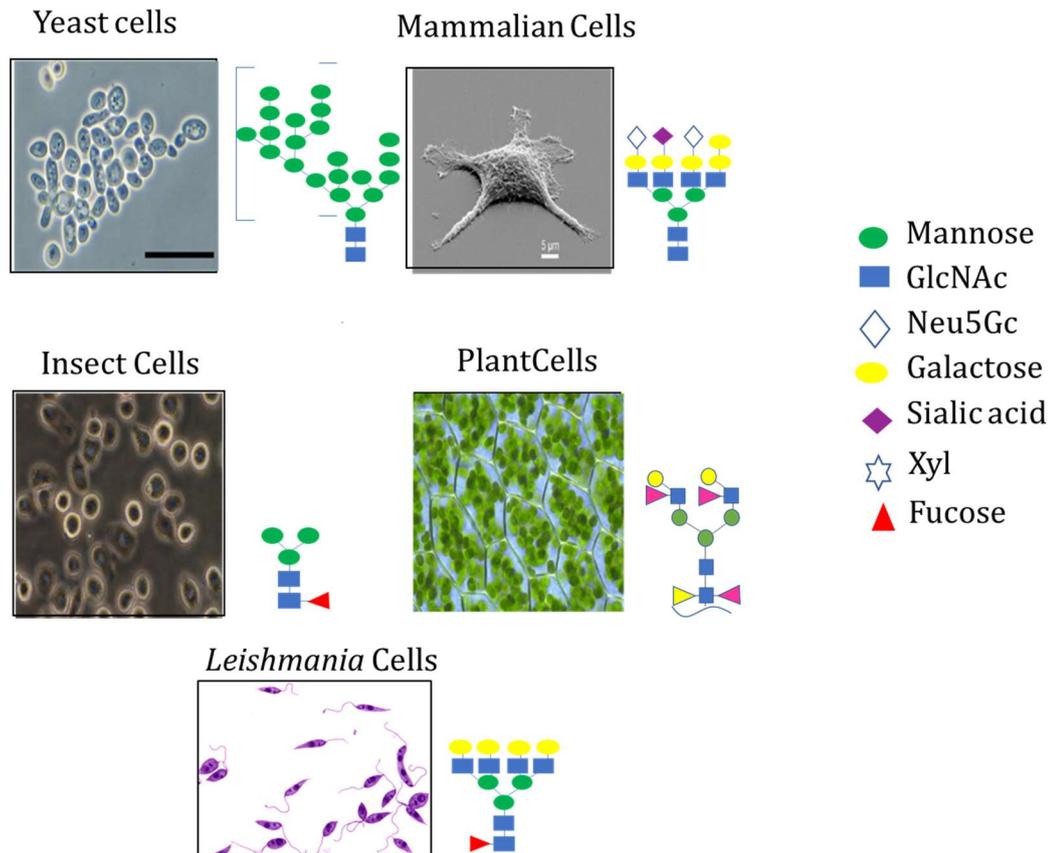
The modification that are carried out in the ER are similar and conserved between lower and higher eukaryotes. However, reaction that take place in the Golgi apparatus differs according the type of cell host.<sup>16</sup> It is then important to find expression system that can produce proteins with the appropriate glycosylation. (Figure V-2)



**Figure V-2:** N-glycosylation examples produced in E.R and Golgi apparatus of mammalian, insect and plant cells. Adapted from P.P. Jacobs and al., 2009.<sup>16</sup>

The trypanosomatidae comprehends protozoan parasite like *Trypanosoma* and *Leishmania* species, and their protein-trafficking mechanisms is similar to higher eukaryotes such as mammals and yeasts.<sup>17,18,19</sup> For example, for the GPI-anchored glycoproteins, the assembly of oligosaccharides on dolichol lipid starts in the reticulum endoplasmic, then, the nascent peptides are transferred in the lumen. Finally, the attachment and modification of oligosaccharides occurs in the Golgi apparatus.<sup>18</sup>

In our case, we are studying four galactofuranosyltransferases of *Leishmania major*. These enzymes are localized in the membrane of Golgi apparatus, so there are probably glycosylated.<sup>20</sup> It is then essential to use an appropriate expression system to produce glycosylated GalTs to assess the importance of glycosylation for the function and structure of proteins. Prokaryotic host *i.e E. coli* cannot be used because they lack glycosylation machinery. In contrast eukaryotic expression systems such as yeast, parasite, insect, plants, human and mammalian cells are widely used by academic and industrial laboratories to produce glycosylated proteins. However, the glycosylation profile of recombinant proteins is different as described in Figure V-3, this is due to the different reactions and enzymes in their Golgi apparatus.<sup>18</sup> It is then important to choose the most appropriate host to obtain similar glycoprotein than those produced by *Leishmania major*.



**Figure V-3:** Glycosylation profiles of recombinant glycoproteins produced by yeast, mammalian, plant, insect and leishmanial cells. Adapted from Amjad Hayat Khan et al., 2017.<sup>18</sup>

## B. *Leishmania tarentolae* as an expression host system

### 1. *Leishmania tarentolae*

LPG1x proteins are naturally produced by *Leishmania major*, so the best strategy might be to use the same organism to produce our recombinant proteins. Jena Bioscience developed a *Leishmania* expression system called LEXSY. It is based on the use of non-pathogenic *Leishmania* species, *L. tarentolae*. This species has been first isolated from the lizard *Tarentolae mauritanica* in 1921.<sup>21</sup> It is considered as one of the most studied *Leishmania* and its genome was totally sequenced in 2012.<sup>22</sup> This parasite infect lizards, but it is also able to enter human phagocytic cells and differentiate into amastigote-like forms. However, there is no evidence of their multiplication and division into the human cells.<sup>23</sup> It was showed that in *L. tarentolae*, a number of genes that are important for pathogenesis and expressed mostly in intracellular parasite life cycle, are either absent or more relevant during their extracellular life cycle, meaning in the insect infection stage.<sup>22</sup> Moreover, lot of genes involved in the biosynthesis of Lipophosphoglycan (LPG) structures are either absent or found in lower copy number than other *Leishmania* species such as *L. major*. According to Raymond et al., the comparison between the *L. tarentolae* genome with *L. major* show high percent of identity as showed in table V-1. In addition, they possess similar

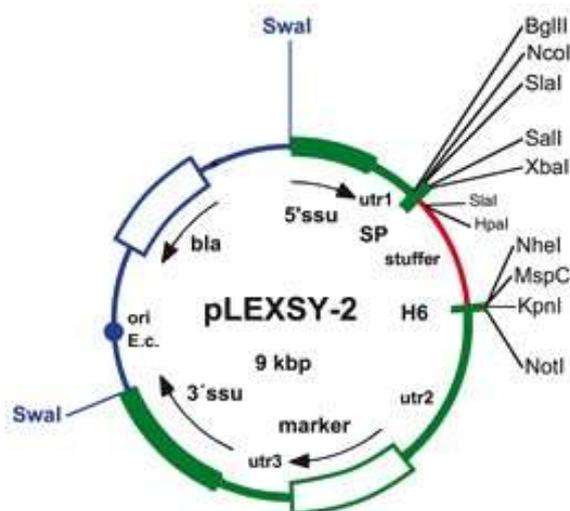
post translational modification machinery.<sup>18</sup> *L. tarentolae* cells are easy to handle and cultivated in liquid media under the promastigote form. High-cell densities of *Leishmania* are obtained.<sup>24</sup>

	<i>Leishmania tarentolae</i>	<i>Leishmania major</i>
Strain	<i>Parrot-Tar II</i>	<i>Friedlin</i>
Number of chromosomes	36	36
Genome size (bp)	30 440 719	32 816 678
Overall G+C content (%)	57.2	59.7
Coding G+C content (%)	58.4	62.5
Number of coding genes in database (time of study)	8201	8304
Number of OGs	7449	7530
Mean nucleotide identity with <i>L. tarentolae</i> (%)	-	84.9
Mean amino acid identity with <i>L. tarentolae</i> (%)	-	81.9

**Table V- 1:** Comparison between *Leishmania tarentolae* and *Leishmania major* sequenced genome. From Frédéric Raymond *et al.*, 2011.<sup>22</sup>

For these reasons, using *Leishmania tarentolae* as expression system seems very promising to produce leishmanial and glycosylated galactofuranosyltransferases.

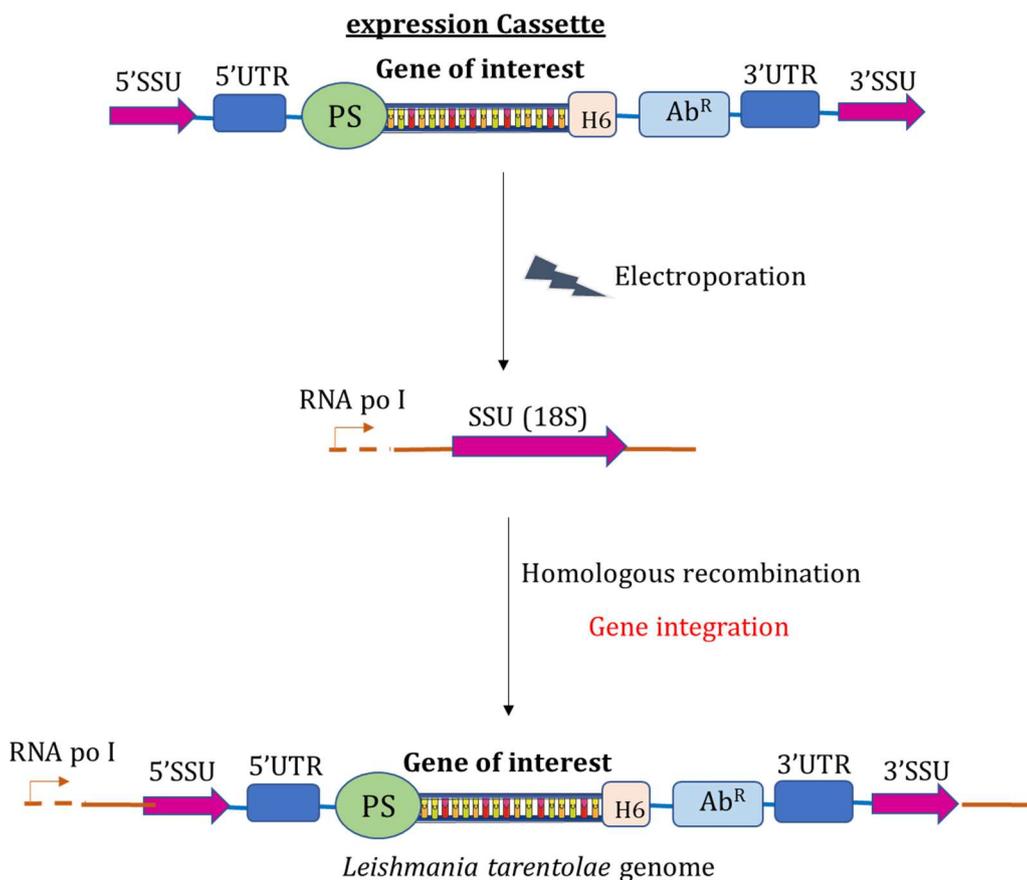
## 2. LEXSY vector



**Figure V- 4:** pLEXSY-2 plasmid cartography.

Specific expression vectors LEXSY (FigureV-4), were designed by Jena bioscience. This shuttle plasmid contains specific elements that allow the combination between the replication of the

plasmid by prokaryotic host, *E. coli* and the expression of proteins by the eukaryotic host, *L. tarentolae*. We can observe the origin of replication, in addition to the resistance maker for ampicillin for the bacterial replication and selection steps. The multi cloning site, when the gene of interest will be inserted, is flanked by two specific regions called UTR1 and UTR2 that are optimized untranslated regions that regulate the mRNA transcription of the gene of interest and the marker gene in the LEXSY host.<sup>25,26</sup> The expression cassette that contains the targeted gene, the gene marker that allows the selection of *Leishmania* and the UTRs regions, is flanked by two small subunit rRNA loci, called 5'ssu and 3'ssu. These regions are strongly transcribed by RNA polymerase I, providing high level of transcription rate of the gene of interest. They also allow the integration of the expression cassette into the genome of *Leishmania tarentolae*, in the chromosomal 18S rRNA locus (Figure V-5).<sup>27</sup> This integration is achieved by homologous recombination.<sup>28</sup> The transfection of *Leishmania tarentolae* promastigotes is performed by electroporation, and the selection of transformants is performed thanks to the antibiotic resistance genes.<sup>24</sup>



**Figure V-5:** Transfection and integration of expression cassette from LEXSY plasmid into the genome of *Leishmania tarentolae* by homologous recombination thanks to SSU regions. SP: Peptide Signal, Ab<sup>R</sup>: Antibiotic Resistance.

LEXSY vector offers the possibility to obtain either intracellular or secreted recombinant proteins as shown in Figure V-5. Indeed, depending of the strategy of cloning we can have a signal peptide of *L. Mexicana* secreted acid LMSAP1, that can be fused to our protein of interest. The secretion peptide is cleaved during the secretion of the protein.<sup>24,29</sup>

The LEXSY plamid was used to express various secreted and intracellular heterologous proteins into *Leishmania tarentolae*. Indeed, different human proteins were already produced by this system, such as erythropoietin<sup>25</sup>, laminin-332<sup>30</sup>, N-Acetyl Serotonin Methyl Transferase<sup>31</sup>, and coagulation factor VII.<sup>32</sup> However, few leishmanian proteins were also obtained such as LPG3, the *Leishmania* homolog of the mammalian endoplasmic reticulum chaperonin GRP94<sup>33</sup>, an invertase BfrA<sup>34</sup>, and LIT1 an iron transporter<sup>35</sup>.

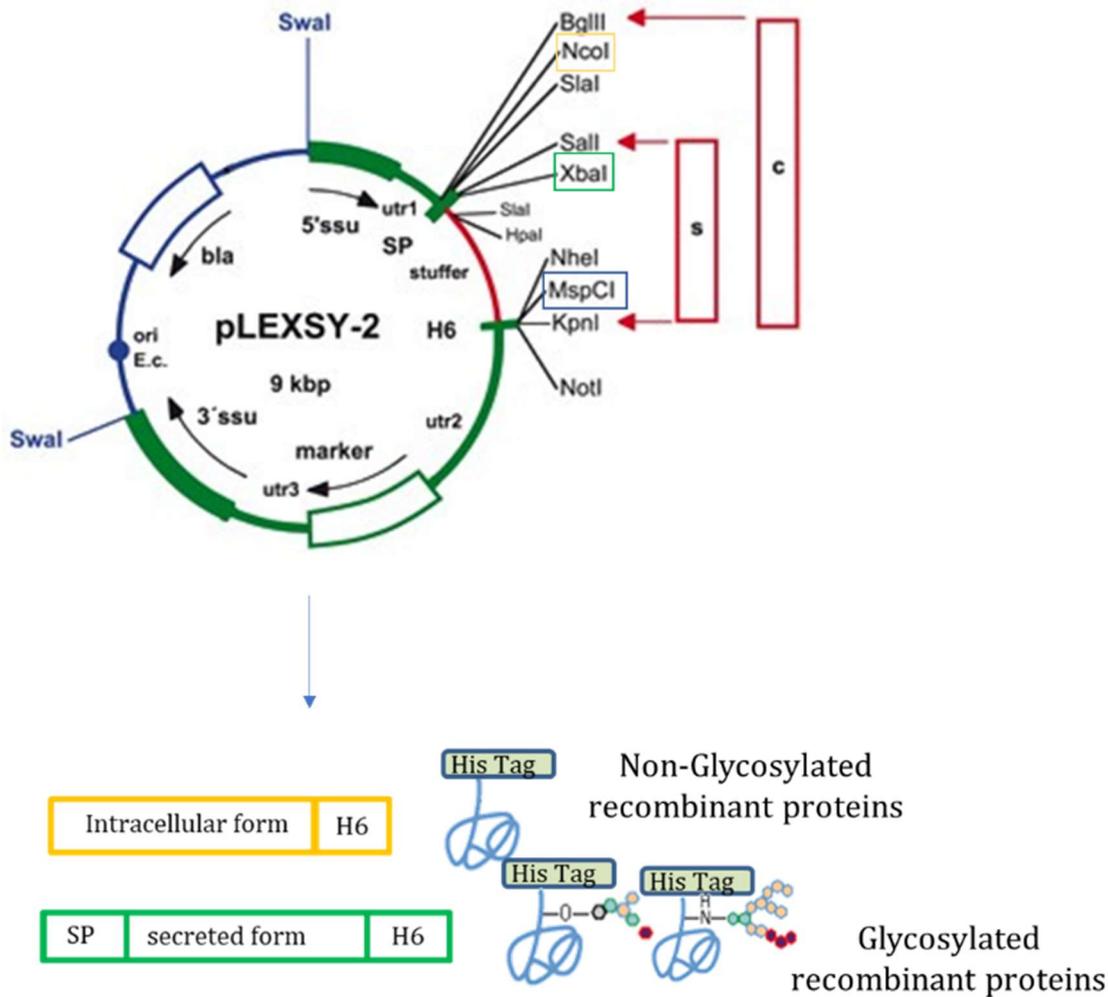
## II. Glycosylated recombinant GalTs of *Leishmania*

### A. Cloning and transfection

#### 1. Cloning

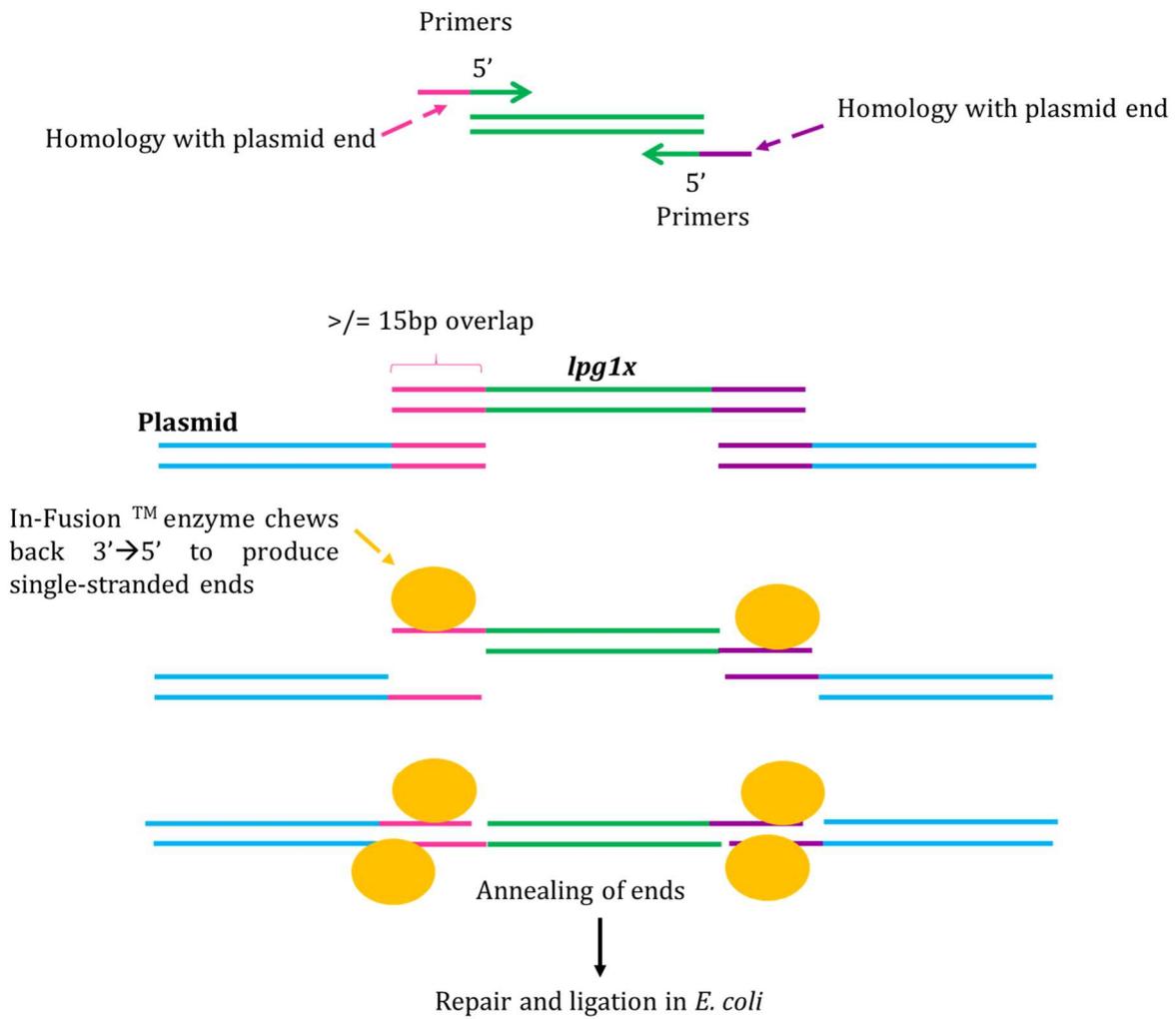
As mentioned before, *Leishmania spp* has a unique genomic organization. Indeed, their genes do not contain introns, and are transcribed as polycistronic transcripts.<sup>36</sup> For this reason, DNA fragment of *Leishmania major* was used directly as template for PCR reactions. Here again, regions that encode for transmembrane domains were removed, and two strategies of cloning were adopted, based on the cloning site. In one hand, secreted proteins were obtained when the gene of interest was inserted into LEXSY plasmid using XbaI and MspCI cloning sites. In this case, the secretion signal site was fused with the target gene, resulting in the expression of fused protein with signal secretion and the histidine tag. On the other hand, NCoI and MspCI sites were used to clone the gene of interest into LEXSY plasmid, and the secretion signal site was removed. Consequently, intracellular proteins are expressed fused only with the histidine tag. (Figure V-6)

The purpose to obtain the two forms, intracellular and secreted, for the same proteins, is to study the effect of the glycosylation on the activity. Thus, we will be able to compare between glycosylated proteins and non-glycosylated proteins produced by *L. tarentolae*, and between glycosylated proteins produced by *L. tarentolae* and/or non-glycosylated proteins produced by *E. coli*.



**Figure V-6:** The two strategies of cloning. According to the restriction site either intracellular form or secreted form of recombinant proteins are produced.

At the beginning of these experiments, a cloning protocol was chosen to insert our genes of interest into pLEXSY and first trials were carried out with *lpg1G* gene. However, only cytosolic-*lpg1G* amplification and cloning succeeded. So, different primers with additional nucleotides were tested to amplify the target genes, but the results were the same. The classical cloning was not appropriate in our hands to extract and amplify *lpg1x* genes. Consequently, In-Fusion® HD Cloning method designed by Clontech was tested. It offers the possibility to clone one or more fragments of DNA into any vector easily. Here, the primers should share 15 bases of homology with the ends of linearized plasmid.<sup>37</sup> So, they contain part of 5' end and 3' end of the desired gene and part of plasmid, pLEXSY. After gene amplification, the ligation step is performed by 3'-5' exonuclease activity of poxvirus DNA polymerase.<sup>38</sup> This enzyme has the ability to promote the pairing of two linear DNA sharing the same ends.<sup>39</sup> (Figure V-7)



**Figure V-7:** Mechanism of In-Fusion cloning reaction used to insert genes of interest into LEXSY plasmid. Adapted from Baogong Zhu et al., 2007.<sup>37</sup>

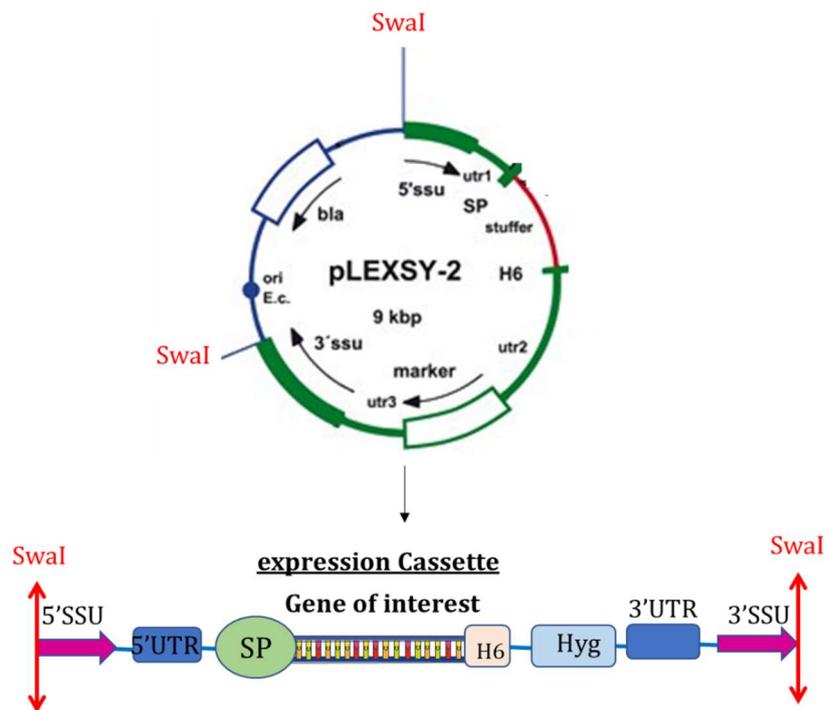
	Strategy of cloning	Type of cloning	PCR product Length (pb)
<i>lpg1</i>	Intracellular form	In-fusion	1217
	Secreted form	In-fusion	1220
<i>lpg1G</i>	Intracellular form	Classical	1696
	Secreted form	In-fusion	1707
<i>lpg1L</i>	Intracellular form	In-fusion	1677
	Secreted form	In-fusion	1679
<i>lpg1R</i>	Intracellular form	In-fusion	1293
	Secreted form	In-fusion	1287

**Table V-2:** Type of cloning used for each gene and the expected PCR product length.

Finally, all constructions were created (Table V-2) and transformed into *E. coli*. Construction were confirmed by sequencing performed by Eurofins Genomics.

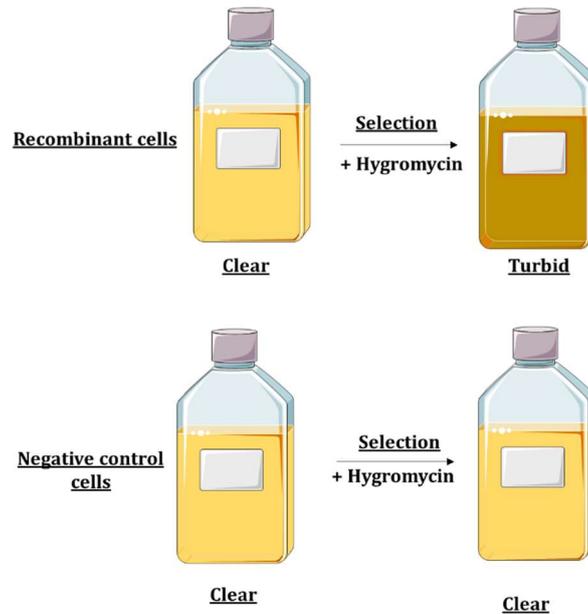
## 2. Transfection and selection

Plasmids were extracted from *E. coli* and digested with the specific restrictive enzyme, *SwaI*, in order to obtain linearized expression cassette flanked by SSU regions (Figure V-8). This linearized DNA fragments were transfected into *L. tarentolae* by electroporation. The selection of *Leishmania* that integrate genes of interest was done by the addition of hygromycin on the medium.



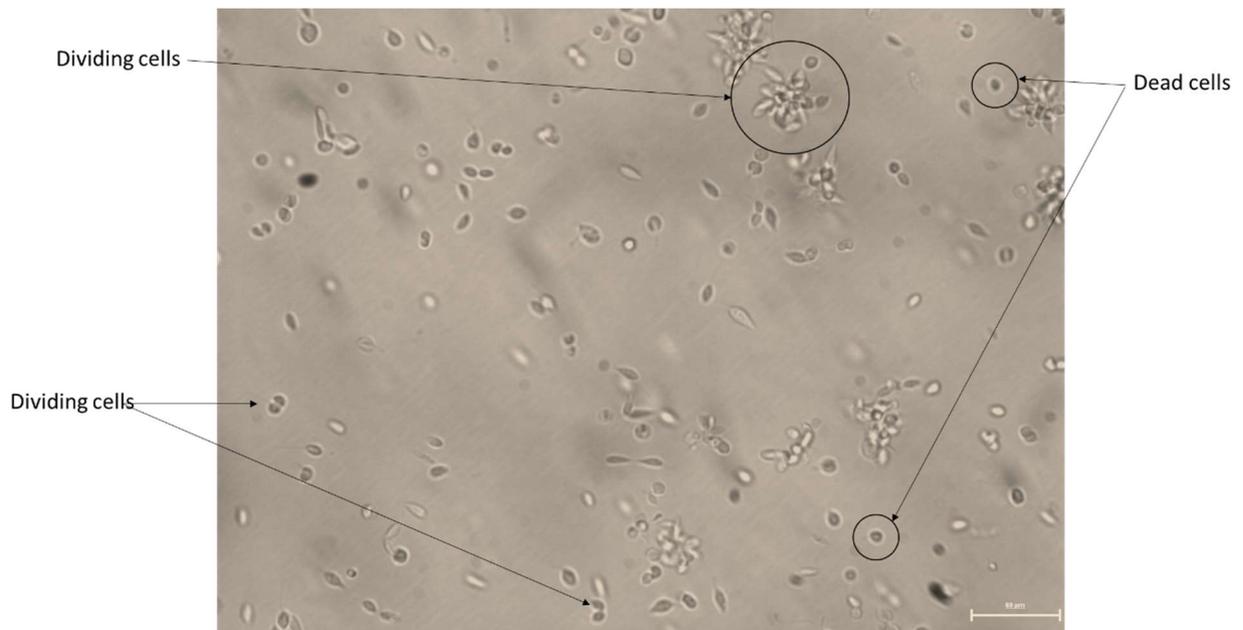
**Figure V-8:** Preparation of linearized expression cassette by enzymatic digestion of cloning constructions by *SwaI*.

*Leishmanian* cells are cultivated as suspension cultures. Visually, at the beginning of the selection, both non-transfected cells considered as negative controls and transfected cells are turbid. However, after different passage of cultures, negative-control cultures remain clear, which is explained by the death of cells, and transfected cells became turbid due to the continuous division and growth of cells. (Figure V-9)



**Figure V-9:** Culture state after selection with hygromycin antibiotic. The transfected cells remain turbid, non-transfected cells became clear.

Transfected cells were observed under microscope in order to follow the status of the cultures and control their appearance and motility. Recombinants cells were motile again, and are characterized by a drop-like shape. In the same culture, we can see dead cells, that did not integrate the gene of interest, characterized by their spherical forms without flagella (Figure V-10).



**Figure V-10:** Microscopic image (X600) of *lpg1x*-transfected cells with *lpg1G* during the selection process with hygromycin.

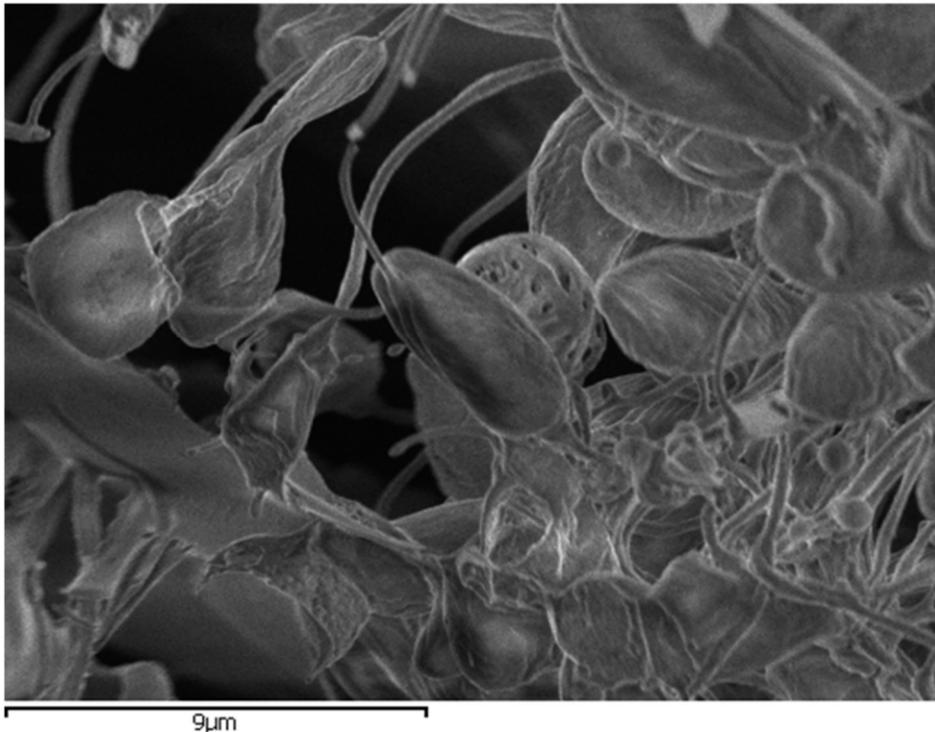
After several passage of cells, recombinant promastigotes are thinner and have lance-like shape with a long flagellum. (Figure V-11) These observations confirm that they are healthy and are resistant to antibiotic, hygromycin. Thus, the transfection and the integration of the expression cassette was successful.



**Figure V-11:** Microscopic image (X 1000) of transfected *L. tarentolae* cells with *lpg1G* after several passage showing healthy promastigotes cells with drop-like shape.

Some samples of our transfected *Leishmania* were also observed by cryo-electronic microscopy (cryo-EM) showed on Figure V-12. This technique uses a beam of electrons to create an image of the specimen. It is based on the quick freezing of the samples in hydrated form in liquid nitrogen, in order to reduce the damage of samples caused by radiations.<sup>40,41,42</sup>

B

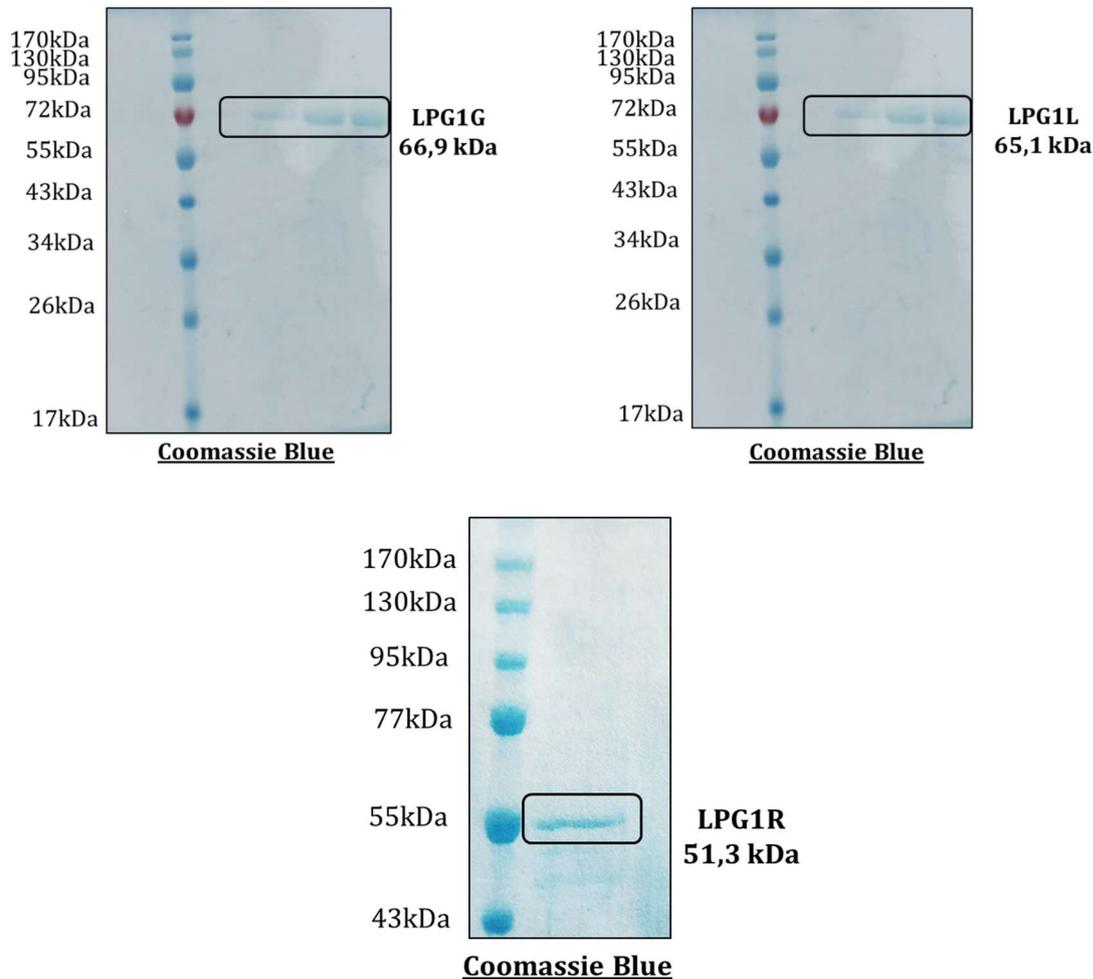


**Figure V-12:** Cryo-Electron microscopy picture showing the presence of transfected promastigotes with drop-like shape and flagellum.

## B. Protein expression and characterization

### 1. Protein Purification

The recombinant cells were cultivated. After centrifugation of 1 liter of culture, only supernatant was extracted and purified using affinity chromatography. The recombinant proteins are fused with histidine tag, so Ni-NTA column were used for their purification. Concentration of the obtained proteins were evaluated thanks to Bradford assays. A yield was obtained, around 15 -20 mg per Liter This corresponds to a higher yield than prokaryotic proteins, which is around 10mg/Liter.



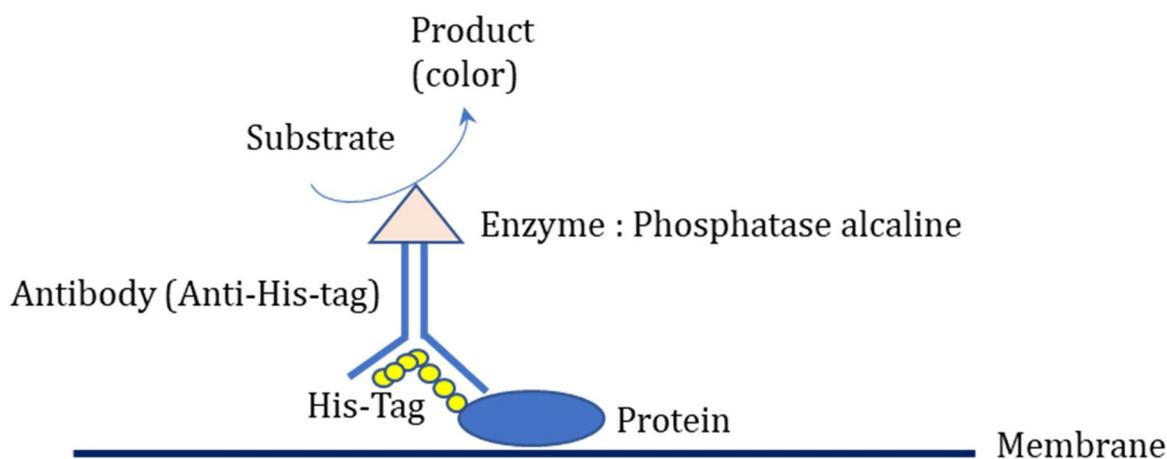
**Figure V- 13:** Evaluation of the expression and the purity of *L. major* GalT's after Ni-NTA elution step in 1-D 8% SDS-PAGE stained with Coomassie. Standard mixture of proteins' marker was used. LPG1G (67 kDa), LPG1L (62 kDa) and LPG1R (52 kDa).

Samples of each purified proteins, LPG1G, LPG1L and LPG1R were migrated on 8% SDS-PAGEs. The gel was stained with Coomassie Blue in order to evaluate the purification of glycosylated proteins and their molecular weights. In Figure V-13 three bands are visible on the three gels corresponding to the glycosylated LPG1G, LPG1L and LPG1R with their respective molecular weights, 67, 62, and 52 kDa. The presence of only one band each time indicates that these purified proteins are pure.

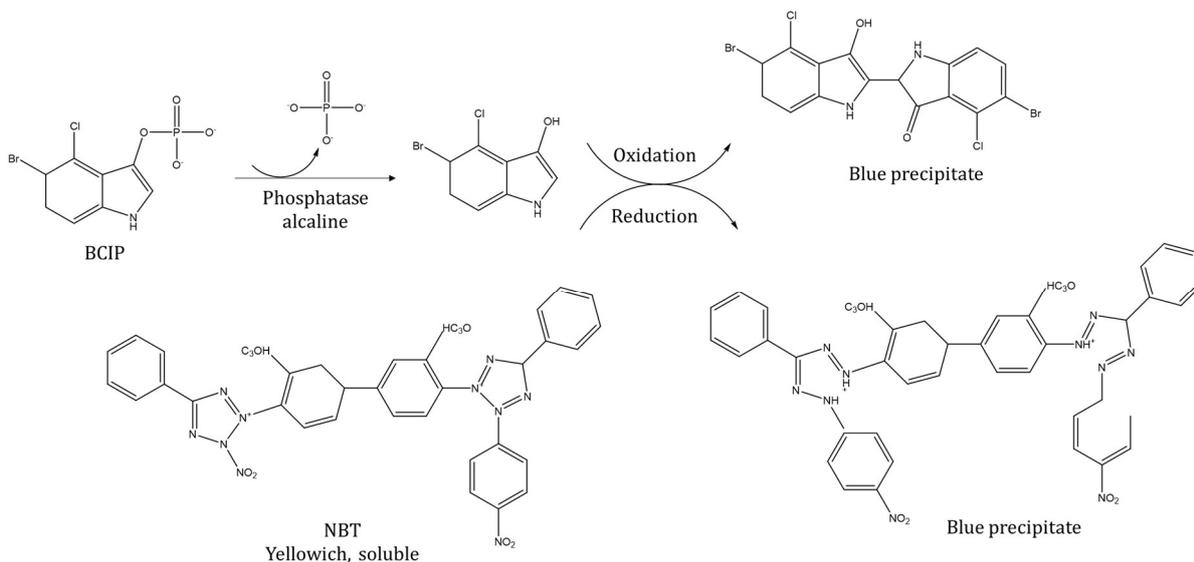
It is known that *Leishmania tarentolae* express secreted glycoproteins such as GP63, which is a 62kDa protein highly expressed by promastigotes and that plays important role in the attachment with macrophages.<sup>43</sup> It was important to make sure that these band do not correspond to one of these glycoproteins, so non transfected cells were cultivated and supernatant was purified with the same conditions. First, no peak was detected on AKTA with elution step with imidazole buffer, and analysis on SDS-PAGE revealed the absence of bands.

## 2. Immunodetection

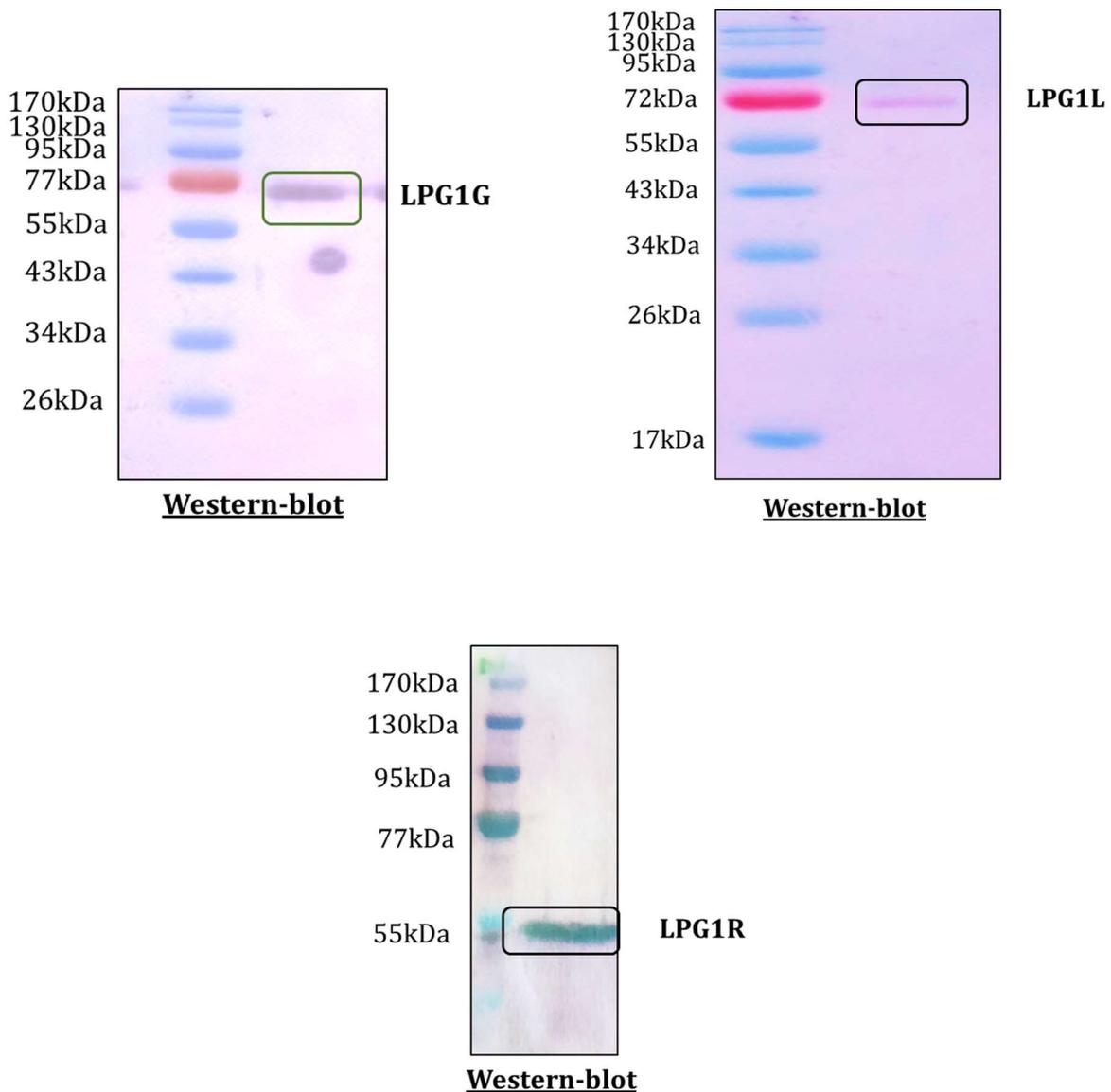
Western blot analysis was performed in order to make sure that the observed band on SDS-PAGE stained with Coomassie blue correspond to proteins of interests. Here, specific antibodies against to phosphatase alkaline that are coupled to C-terminal poly-histidine tag were used. Once the antibodies fix our proteins of interest, they are revealed with specific substrate of phosphatase alkaline resulting the apparition of bands on the membrane. Indeed, AP-substrate contain 5-bromo-4-chloro-3-indolyl phosphate/Nitroblue Tetrazolium (BCIP/NBT), which reacts with the alkaline phosphatase to form a color precipitate for detection (Figure V-14/15).



**Figure V-14:** Colorimetric detection. The antibody, which binds the antibody specific for the protein of interest, is conjugated to an enzyme; its substrate is converted to a colored precipitate.



**Figure V-15:** Reaction of phosphatase alkaline in the presence of substrate BCIP/NBT



**Figure V-16:** Identification of LPG1G (67 kDa), LPG1L (62 kDa) and LPG1R (52 kDa) by western blot technique, using antibodies against to C-terminal poly-histidine tags fused to the proteins of interest.

Figure V-16 shows the analysis of recombinant LPG1G, LPG1L and LPG1R samples by western blot. Only one band is revealed for each of these proteins. They have similar molecular weight as observed in previous SDS-PAGEs. Here, antibodies fixed specifically the His-tag present in C-termini of recombinant proteins. This confirmed that the obtained band correspond to our proteins of interests.

At this level, we can then confirm that secreted LPG1G, LPG1L and LPG1R were successfully expressed and purified with good yield.

### 3. Glycosylation

Secreted proteins should be glycosylated. However, the exact level of glycosylation cannot be calculated or predicted. But, the sites of *N*- and *O*-glycosylation can be predicted thanks to NetGGlyc 1.0 Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>) and NetOGlyc 4.0 Server (<http://www.cbs.dtu.dk/services/NetOGlyc/>). However, this prediction remains limited because glycosylation sites can vary in-vivo depending on the cells that are expressing the protein.

Name:	Sequence	Length:	596		
	MASRLVRLVLAAMLVAAAVSVDAGASLASSHQ <del>S</del> ALG <del>W</del> TREPTTSAPSGELTTTEPGEP <del>R</del> RRSMTPTS <del>A</del> ITTSAPATATTAE		80		
	QTPLPPEQTS <del>M</del> FDAPGTP <del>T</del> PAVSAASARATGAQ <del>S</del> STLM <del>E</del> GAKTETTTK <del>Q</del> NLAMKGLDR <del>W</del> MKMLPAD <del>W</del> MECIR <del>Q</del> NLQ <del>L</del> DK		160		
	HGQPMHAVTEMKDAI <del>P</del> LLITPLTGDV <del>K</del> FFPYFVCSMDVAVRYHYV <del>I</del> QNERDPDTTAVIDDLQRRFGNSGRL <del>L</del> VVRNRYNR		240		
	GYSGSMNQAF <del>E</del> WALKERTAE <del>E</del> VP <del>W</del> VFACGVD <del>A</del> IFEPGLLAKMIEV <del>V</del> Q <del>N</del> TRGDAAMLALRAEVEL <del>E</del> ERLVREG <del>N</del> YSY <del>E</del>		320		
	RWAPRGRPLKVL <del>R</del> SGYPGVPLNVRTAPLLPDRIRYTVADENRESGIVRPAELRTRFFGN <del>V</del> ATVTPVPDALGTIAVTRLA		400		
	LSAVGYFDENYF <del>P</del> AYMDDIDL <del>R</del> WRHFAYGFGALHGERNGPVTR <del>I</del> H <del>H</del> YNAANLRGSPFVDPDLEKYGTED <del>N</del> YSRR <del>A</del> FLSYI		480		
	RRSKGIYDKLKYGPRDMDGV <del>R</del> QAAQKA <del>E</del> FKYSS <del>F</del> NVSHYPADTWLDEDARGCMFHHTYNYKT <del>Q</del> NWSRPSDCSYNPRTL		560		
	EESGILGVDQLASYRSM <del>L</del> A <del>E</del> KAFAYQFKGTHHHHHH		596		
	.....		80		
	.....		160		
	.....		240		
	.....N.....		320		
	.....		400		
	.....N.....		480		
	.....N.....N.....		560		
	.....		640		
(Threshold=0.5)					
SeqName	Position	Potential	Jury agreement	N-Glyc result	
Sequence	315	NYSY	0.6402	(9/9)	++
Sequence	470	NYSR	0.5986	(7/9)	+
Sequence	516	NVSH	0.6108	(6/9)	+
Sequence	546	NWSR	0.5129	(5/9)	+

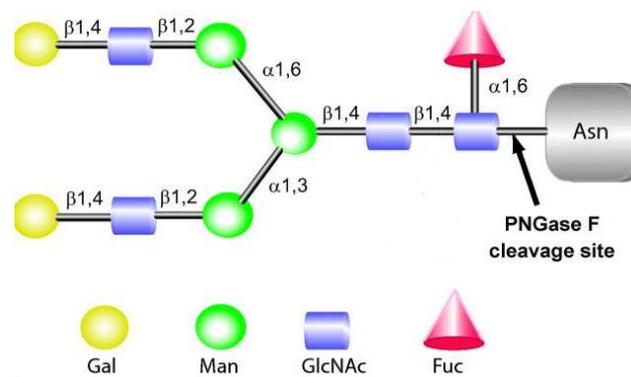
**Figure V-17:** Predictions for *N*-Glycosylation sites in specific sequence in secreted and recombinant LPG1G amino acid sequence. Asn-Xaa-Ser/Thr sequences (including Asn-Pro-Ser/Thr) are shown in blue. Asparagine predicted to be *N*-glycosylated are shown in red.

#seqname	source	feature	start	end	score	strand	frame	comment
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	3	3	0.122737	.	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	20	20	0.0956727	.	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	26	26	0.335303	.	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	29	29	0.5	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	30	30	0.682966	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	33	33	0.515574	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	38	38	0.662962	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	42	42	0.899269	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	43	43	0.953311	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	44	44	0.951045	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	47	47	0.829287	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	51	51	0.701886	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	52	52	0.89276	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	53	53	0.896143	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	61	61	0.921949	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	63	63	0.703449	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	65	65	0.749063	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	66	66	0.795687	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	69	69	0.942121	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	70	70	0.911296	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	71	71	0.881336	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	75	75	0.781385	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	77	77	0.846321	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	78	78	0.827261	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	82	82	0.892965	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	89	89	0.795071	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	90	90	0.698239	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	97	97	0.5	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	99	99	0.916566	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	103	103	0.97457	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	106	106	0.916202	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	110	110	0.774461	.	.	#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	115	115	0.343597	.	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	116	116	0.82403	.	.	#POSITIVE

**Figure V-18:** Predictions for *O*-Glycosylation sites in specific sequence in secreted and recombinant LPG1G amino acid sequence.

Only the sites with scores higher than 0.5 are predicted as glycosylated, and according to Figure V-17 and V-18, 4 *N*-glycosylation sites and around 30 *O*-glycosylated sites have been predicted.

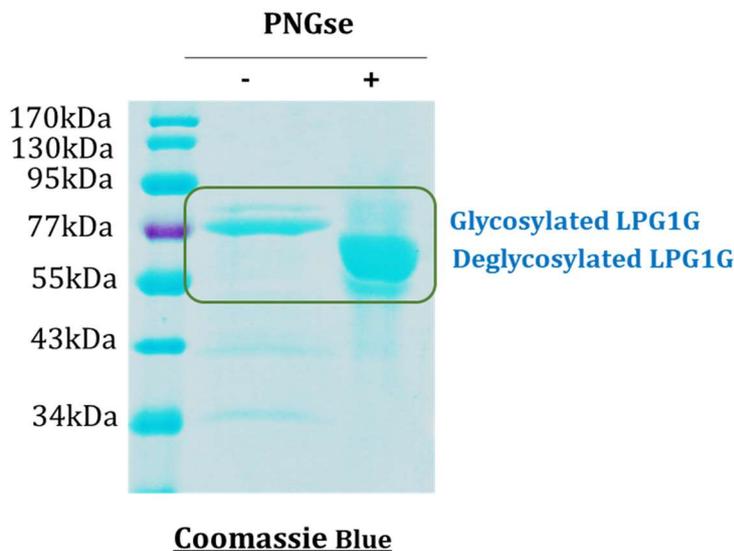
PNGase, an enzyme that catalyzes the cleavage (the removal) of *N*-linked oligosaccharides attached to glycoproteins was used (Figure V-19).



**Figure V-19:** Removal of *N*-linked oligosaccharides by PNGase.

However, the other type of linked glycans were not removed. There is no available enzyme able to cleave the entire *O*-glycans from the protein's backbone. An exoglycosidase can be used as an alternative enzyme, to trim *O*-glycans oligosaccharides, and the left short cores could be removed

by some *O*-glycosidases such Endo- $\alpha$ -*N*-acetylgalactosaminidase, which hydrolyzes the serine or threonine-linked unsubstituted *O*-glycan core [Gal- $\beta$ (1 $\rightarrow$ 3)-GalNAc].<sup>40,41</sup>



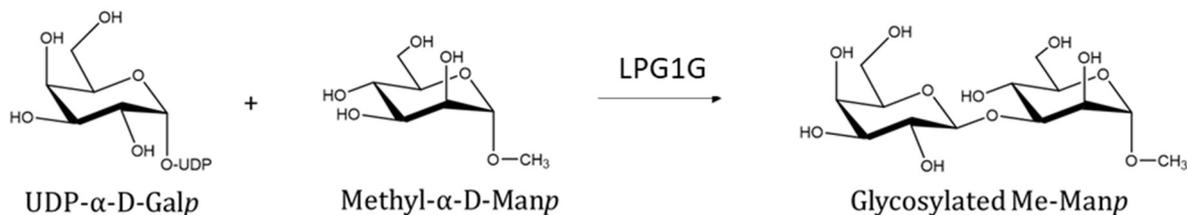
**Figure V-20:** Evaluation of deglycosylation of LPG1G due to PNGase activity on 1-D 8% SDS-PAGE stained with Coomassie with standard mixture marker proteins.

An amount of LPG1G was incubated with PNGase, then the mixture migrated on SDS-PAGE and revealed with Coomassie Blue as showed in Figure V-20. In the absence of PNGase, there is one band around 72kDa which is higher than the expected weight, 67kDa. In the presence of PNGase, there is a band with lower molecular weight. The observed shift is due to the activity of PNGase. The difference in the molecular weight is explained by the cleavage of the *N*-linked oligosaccharide. This confirms that secreted LPG1G is glycosylated. We cannot say how many sites are *N*-glycosylated, or if they heavily glycosylated or not. A mass spectrometry analysis was performed to evaluate the molecular weight of the glycosylated LPG1G, but was not successful. Moreover, there is probably different glycoforms of LPG1G due to the site heterogeneity. Site-directed mutagenesis can be performed to modify predicted *N*-glycosylation sites. This could help us to confirm if they are glycosylated as predicted or not. Such experiments were carried out on human gastric lipase by Wicker-Planquart C et al., Their purpose was to study the effect of some *N*-glycosylation on the activity of the enzyme.<sup>44</sup> It is quite difficult to cleave glycans from glycoproteins, that is why, we decided to study the activity of glycosylated enzymes first.

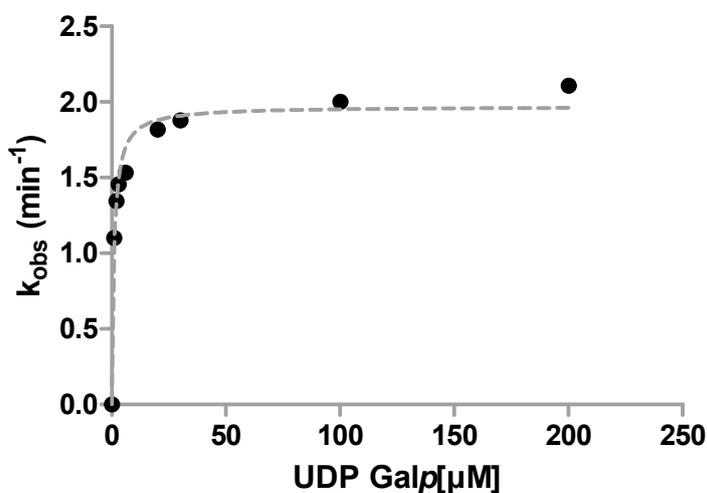
#### 4. Activity characterization

Preliminary trials to evaluate the activity of glycosylated recombinant proteins were carried out with LPG1G. Here due to the lack of UDP-Galactofuranose and its difficult synthesis, we decided to use UDP-Galp as donor and Me-Manp as acceptor. Based on results obtained with prokaryotic

proteins, LPG1G produced by *E. coli* was able to transfer galactopyranosyl moieties on Me-Manp. So we wanted to see if glycosylated LPG1G has the same activity or not (Figure V-21).



**Figure V-21:** Expected enzymatic reaction catalysed by glycosylated LPG1G. The enzyme could be able to use UDP-Galp as donor and Me-Manp as acceptor to form disaccharide, glycosylated Me-Manp.



**Figure V- 22:** Michaelis-Menten plot that corresponds to UDP-Galp kinetics as donor substrate of glycosylated LPG1G.

Spectrophotometric assays using the tri-enzymatic assays described before in chapter 2 were performed. Figure V-22 shows the Michaelis Menten plot, that was obtained with LPG1G using different concentration of UDP-Galp ranging from 1 to 200 $\mu$ M as donor and 1mM Me-Manp as acceptor. We can see that LPG1G reaches saturation state in the presence of 50 $\mu$ M of UDP-Galp. Nevertheless, this preliminary result indicates that eukaryotic LPG1G is active and is also able to transfer galactopyranose on Me-Manp. This experiment was performed only once. So, it is important to repeat the same experiments at least twice. Nevertheless, we were able to estimate  $k_{cat}$  and  $k_{cat}/K_M$ .

Enzyme	Apparent $K_M^a$ (mM)	$k_{cat}^a$ (min <sup>-1</sup> )	$k_{cat}/K_M^a$ (min <sup>-1</sup> mM <sup>-1</sup> )
Eukaryotic LPG1G	0.001	3.1	3,100
Prokaryotic LPG1G	0.005 ± 0.001	132 ± 10	27,978

**Table V-3:** Kinetic parameters of eukaryotic GalT LPG1G compared with prokaryotic GlT2 for UDP- $\alpha$ -D-Galp. a Me-Manp was used as the acceptor.

$k_{cat}$  and  $k_{cat}/K_M$  of both prokaryotic and eukaryotic LPG1G for UDP-Galp are presented in Table V-3.  $k_{cat}/K_M$  values of eukaryotic LPG1G (3100 min<sup>-1</sup> mM<sup>-1</sup>) is 10-fold lower than LPG1G expressed by *E. coli* (27,978 min<sup>-1</sup> mM<sup>-1</sup>). There is a decrease of turnover of the enzyme but not a complete loss of its activity. Glycans attached to glycoproteins can be heterogenous, which lead to the presence of various glycoforms of proteins. In some glycoforms, glycosylation can affect their properties, which affects their activity. Aldo Ceriotti et al.,<sup>45</sup> suggested that *N*-glycosylation affect directly the interaction of nascent polypeptides with solvent or chaperonin; which could modify the final structure of the glycoproteins.

Glycosylation can shield some regions of the active site of the enzyme. Thus, it can modulate the specificity of the enzyme toward substrate. A similar phenomenon was reported by Pi Liu et al., 2015. They highlighted the importance of *N*-glycosylation on the activity of the enzyme. Indeed, they showed that the glycosylation of only one site could block the entrance of substrate into the binding pocket, which could explain its activity.<sup>46</sup> Moreover, Daniellou and co-worker have investigated the activity of the invertase BfrA of the GH32 family in the *Leishmania* genus. They expressed BfrA using *Leishmania tarentolae* and proved that it is not active when it is glycosylated.<sup>34</sup>

Since LPG1 was located in the membrane of the Golgi apparatus of *Leishmania major*, they are expected to be glycosylated *in vivo*.<sup>20</sup> However, this has never been confirmed. We can then suggest that the presence of glycans could make LPG1x proteins more versatile or more specific toward substrates. Indeed, we demonstrated in the previous chapter, that the non-glycosylated LPG1, LPG1R, LPG1L and LPG1G are all able to transfer Galf on Me-Manp residues. These disaccharides are also connected in the LPG and GIPLs, so we can suggest that the four proteins could be involved in their biosynthesis. But, the knock out genes of *lpg1*, *lpg1R* and *lpg1L* that were performed by Zhang et al., demonstrated that only LPG1 is involved in the biosynthesis of LPGs but none of them seem to be responsible of the addition of galactofuranose in GIPLs structures.

This could be then explained by the presence of glycans on the proteins that modulate the activity of the four enzymes *in-vivo*.

### III. Conclusion

In this chapter, *Leishmania tarentolae*, a non-pathogenic parasite, that infect only lizard was used as an expression host to produce the four GalTs of *Leishmania major*. The purpose of the use of these cells is the availability of post-translational modification machinery that will provide glycosylated proteins similar to those produce by the pathogenic species, *Leishmania major*. Specific plasmid, pLEXY was used to clone the four LPG1x proteins into *L. tarentolae*. According to the strategy of cloning, both intracellular and secreted form of each of LPG1x can be produced. This enabled the obtention of non-glycosylated (intracellular) and glycosylated (secreted) forms. This could allow us to study the impact of the glycosylation on the stability and the activity of the same eukaryotic enzyme. Nevertheless, due to the lack of time, we decided to produce and study only the secreted proteins. The four genes, *lpg1*, *lpg1L*, *lpg1R* and *lpg1G* were cloned into pLEXY and transfected into *L. tarentolae*. However, we only confirmed the production of pure LPG1G LPG1L and LPG1R. A good yield ~ 15 to 20mg/L of culture was obtained compared to 10 mg of LPG1x produced by 1L of *E. coli* culture. We were able to demonstrate that LPG1G is at least *N*-glycosylated, however, *O*-glycosylations were not studied. But, more experiments need to be done to analyze and have more details about the glycosylation sites, in addition to the glycosylation profile. Spectrophotometric assays were carried out to assess the activity of LPG1G using Me- $\alpha$ -D-Mannose as acceptor and UDP-Galactopyranose as donor. Preliminary results suggested that glycosylated LPG1G is also active and able to transfer galactopyranosyl residues on Me-Manp. However, its  $k_{cat}/K_M$  value is 10-fold lower than LPG1G expressed by *E. coli*. This suggest that the presence of glycans affects the activity of the enzymes. Moreover, the presence of *N*-glycans could modulate the entrance of substrate to the active site of the enzyme, leading to the low activity. Further experiments should be carried out with UDP-galactofuranose and other substrates for the four LPG1x to evaluate their activity and specificity.

## References

1. Spiro, R. G. Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, (2002).
2. Chandler, K. B. & Costello, C. E. Glycomics and glycoproteomics of membrane proteins and cell-surface receptors: Present trends and future opportunities. *Electrophoresis* **37**, 1407–1419 (2016).
3. Shental-Bechor, D. & Levy, Y. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci.* **105**, 8256–8261 (2008).
4. Parekh, R. B. Effects of glycosylation on protein function. *Curr. Opin. Struct. Biol.* **1**, 750–754 (1991).
5. Ceriotti, A., Duranti, M. & Bollini, R. Effects of N-glycosylation on the folding and structure of plant proteins. *J. Exp. Bot.* **49**, 1091–1103 (1998).
6. Zauner, G. *et al.* Protein O-glycosylation analysis. *Biol. Chem.* **393**, 687–708 (2012).
7. Goettig, P. Effects of glycosylation on the enzymatic activity and mechanisms of proteases. *Int. J. Mol. Sci.* **17**, 1–24 (2016).
8. Higel, F., Seidl, A., Sörgel, F. & Friess, W. N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur. J. Pharm. Biopharm.* **100**, 94–100 (2016).
9. Zheng, K., Bantog, C. & Bayer, R. The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs* **3**, 567–576 (2011).
10. Mellquist, J. L., Kasturi, L., Spitalnik, S. L. & Shakin-Eshleman, S. H. The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. *Biochemistry* **37**, 6833–6837 (1998).
11. Shcherbakova, A., Tiemann, B., Buettner, F. F. R. & Bakker, H. Distinct C- mannosylation of netrin receptor thrombospondin type 1 repeats by mammalian DPY19L1 and DPY19L3. *Proc. Natl. Acad. Sci.* **114**, 2574–2579 (2017).
12. Haynes, P. A. Phosphoglycosylation: A new structural class of glycosylation? *Glycobiology* **8**, 1–5 (1998).
13. Lauc, G., Krištić, J. & Zoldoš, V. Glycans - the third revolution in evolution. *Front. Genet.* **5**, 1–7 (2014).
14. An, H. J., Froehlich, J. W. & Lebrilla, C. B. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr. Opin. Chem. Biol.* **13**, 421–426 (2009).
15. Stanley, P. Golgi glycosylation. *Cold Spring Harb. Perspect. Biol.* **3**, 1–13 (2011).
16. Jacobs, P. P. & Callewaert, N. K. N-glycosylation Engineering of Biopharmaceutical Expression Systems. *Curr. Mol. Med.* **9**, 774–800 (2009).
17. Flora C. Y. Luk, Terezina M. Johnson, and C. J. B. N-linked glycosylation of proteins in the protozoan parasite *Toxoplasma gondii*. *Mol Biochem Parasitol* **6**, 2166–2171 (2008).
18. Khan, A. H., Bayat, H., Rajabibazl, M., Sabri, S. & Rahimpour, A. Humanizing glycosylation pathways in eukaryotic expression systems. *World J. Microbiol. Biotechnol.* **33**, (2017).
19. Samuelson, J. *et al.* The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. *Proc. Natl. Acad. Sci.* **102**, 1548–

- 1553 (2005).
20. Zhang, K., Barron, T., Turco, S. J. & Beverley, S. M. The LPG1 gene family of *Leishmania major*. *Mol. Biochem. Parasitol.* **136**, 11–23 (2004).
  21. Jones, J. D. *Leishmania tarentolae*: An alternative approach to the production of monoclonal antibodies to treat emerging viral infections. *Infect. Dis. Poverty* **4**, 1–5 (2015).
  22. Raymond, F. *et al.* Genome sequencing of the lizard parasite *Leishmania tarentolae* reveals loss of genes associated to the intracellular stage of human pathogenic species. *Nucleic Acids Res.* **40**, 1131–1147 (2012).
  23. Taylor, V. M. *et al.* *Leishmania tarentolae*: Utility as an in vitro model for screening of antileishmanial agents. *Exp. Parasitol.* **126**, 471–475 (2010).
  24. Basile, G. & Peticca, M. Recombinant Protein Expression in *Leishmania tarentolae*. *Mol. Biotechnol.* **43**, 273–278 (2009).
  25. Breitling, R. *et al.* Non-pathogenic trypanosomatid protozoa as a platform for protein research and production. *Protein Expr. Purif.* **25**, 209–218 (2002).
  26. Clayton, C. E. Genetic manipulation of Kinetoplastida. *BioL* **93**, 53 (1978).
  27. Mißlitz, A., Mottram, J. C., Overath, P. & Aebischer, T. Targeted integration into a rRNA locus results in uniform and high level expression of transgenes in *Leishmania amastigotes*. *Mol. Biochem. Parasitol.* **107**, 251–261 (2000).
  28. Zheng, H. & Wilson, J. H. Gene targeting in normal and amplified cell lines. *Nature* **344**, 170–173 (1990).
  29. Klatt, S. & Konthur, Z. Secretory signal peptide modification for optimized antibody-fragment expression-secretion in *Leishmania tarentolae*. *Microb. Cell Fact.* **11**, 1–10 (2012).
  30. Basak, A., Shervani, N. J., Mbikay, M. & Kolajova, M. Recombinant proprotein convertase 4 (PC4) from *Leishmania tarentolae* expression system: Purification, biochemical study and inhibitor design. *Protein Expr. Purif.* **60**, 117–126 (2008).
  31. Ben-Abdallah, M. *et al.* Production of soluble, active acetyl serotonin methyl transferase in *Leishmania tarentolae*. *Protein Expr. Purif.* **75**, 114–118 (2011).
  32. Mirzaahmadi, S. *et al.* Expression of recombinant human coagulation factor VII by the lizard *Leishmania* expression system. *J. Biomed. Biotechnol.* **2011**, 8 (2011).
  33. Pirdel, L. *et al.* Cloning and expression of *Leishmania infantum* LPG3 gene by the lizard *Leishmania* expression system. *Avicenna J. Med. Biotechnol.* **4**, 186–192 (2012).
  34. Belaz, S. *et al.* Identification, biochemical characterization, and in-vivo expression of the intracellular invertase BfrA from the pathogenic parasite *Leishmania major*. *Carbohydr. Res.* **415**, 31–38 (2015).
  35. Jacques, I., Andrews, N. W. & Huynh, C. Functional characterization of LIT1, the *Leishmania amazonensis* ferrous iron transporter. *Mol. Biochem. Parasitol.* **170**, 28–36 (2010).
  36. Kazemi, B. Genomic Organization of *Leishmania* Species. *Iran. J Parasitol* **6**, 1–18 (2011).
  37. Zhu, B., Cai, G., Hall, E. O. & Freeman, G. J. In-Fusion™ assembly: Seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *Biotechniques* **43**, 354–359 (2007).
  38. Hamilton, M. D., Nuara, A. A., Gammon, D. B., Buller, R. M. & Evans, D. H. Duplex strand

- joining reactions catalyzed by vaccinia virus DNA polymerase. *Nucleic Acids Res.* **35**, 143–151 (2007).
39. Marsischky, G. & LaBaer, J. Many paths to many clones: A comparative look at high-throughput cloning methods. *Genome Res.* **14**, 2020–2028 (2004).
  40. Costello, M. J. Cryo-electron microscopy of biological samples. *Ultrastruct. Pathol.* **30**, 361–371 (2006).
  41. Milne, J. L. S. *et al.* Cryo-electron microscopy - A primer for the non-microscopist. *FEBS J.* **280**, 28–45 (2013).
  42. Carroni, M. & Saibil, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* **95**, 78–85 (2016).
  43. Olivier, M., Atayde, V. D., Isnard, A., Hassani, K. & Shio, M. T. Leishmania virulence factors: Focus on the metalloprotease GP63. *Microbes Infect.* **14**, 1377–1389 (2012).
  44. Wicker-Planquart, C., Cnaan, S., Rivière, M. & Dupuis, L. Site-directed removal of N-glycosylation sites in human gastric lipase. *Eur. J. Biochem.* **262**, 644–651 (1999).
  45. Losfeld, M. E. *et al.* Influence of protein/glycan interaction on site-specific glycan heterogeneity. *FASEB J.* **31**, 4623–4635 (2017).
  46. Liu, P., Wang, Z., Zhang, L., Li, D. & Lin, J. The mechanism by which 146-N-glycan affects the active site of neuraminidase. *PLoS One* **10**, 1–16 (2015).

***Chapter VI: Conclusion and  
perspectives***

Glycosylation is the core of the glycobiology field, and for the scientific community, it is considered as the most complex posttranslational modification that occur during the production of proteins. They are not generated from a template, and their prediction remain an ongoing challenge. Different biopolymers can be glycosylated to give rise to various glycoconjugates such as glycolipids and glycoproteins that play a crucial role in the communication, growth, development of healthy or cancerous cells, mediate the host-pathogen interaction and they are considered as essential elements for the virulence and infection mechanisms. Today it is well known that specific enzymes, called glycosyltransferases, are responsible of the biosynthesis of these glycoconjugates, and they are considered as an interesting target for the development of drugs and novel diagnosis tools. However, our knowledge about these GTs still limited, and many remain poorly studied.

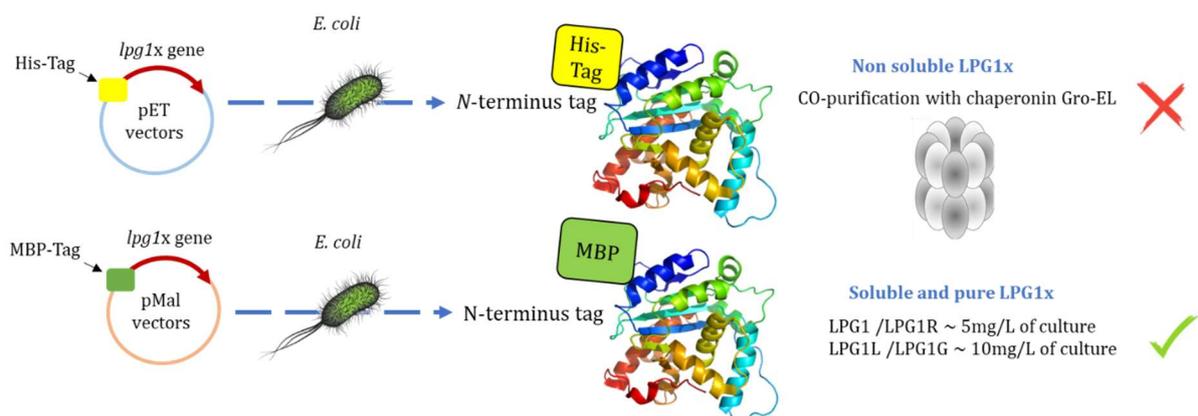
In this thesis, we have studied the galactofuranosyltransferases (Gal $f$ Ts), enzymes that catalyse the incorporation of a rare monosaccharide, galactofuranose (Gal $f$ ), onto glycoconjugates. Gal $f$  is found mainly in different pathogenic species such as *Aspergillus*, *Trypanosoma*, *Mycobacterium*, *Leishmania*, and is absent in mammalian cells. It is worth mentioning that our knowledge about these Gal $f$ Ts is very limited. Indeed, only four galactofuranosyltransferases have been identified, GfsA of *Aspergillus*, WbbI of *E. coli*, GlfT1 and GlfT2 of *Mycobacterium tuberculosis*, but only mycobacterial GlfT2 has been fully characterized. For this, we were interested particularly in galactofuranosyltransferases of *Leishmania* that have never been expressed before. *Leishmania* parasite is the causative agent of leishmaniosis diseases that are classified by the world health organism as a neglected tropical disease and are endemic specifically in the third-world countries. Current drugs, which are commercially available, are either too expensive or toxic. In addition, drug-resistant phenomena have emerged that consist of one of the greatest current threats to human health. Consequently, there is an emergency for novel strategies with improved tolerance and that are effective toward leishmaniosis. Among the leishmanian glycoconjugates, there are lipophosphoglycans and glycosylinositol phospholipids that were proved to be major actors in the infection and survival mechanisms. Thus, the Gal $f$ -containing glycoconjugates attracted a lot of attention. It is then essential to study the Gal $f$ Ts that are involved in their biosynthesis.

Four genes ***lpg1***, ***lpg1L***, ***lpg1R***, and ***lpg1G*** have been identified in the genome of *Leishmania major*. The first three genes are presented in one copy at the loci LMJF\_25\_0010 (*lpg1*), LMJF\_33\_0300 (*lpg1L*), LMJF\_26\_0550 (*lpg1R*), but *lpg1G* exist in identical three copies at different loci LMJF\_32\_3990, LMJF\_05\_1230 and LMJF19\_650. They were predicted to encode for putative transmembrane Gal $f$ Ts that were located in the Golgi apparatus of the parasite. In addition, they do not share any significant homology with the other Gal $f$ Ts. Preliminary results based on RTqPCR experiments showed that the four genes are transcribed in different promastigote form of species of *Leishmania*, *L. major*, *L. donovani* and *L. infantum* with an increase of their expression according

the time. Moreover, these parasites seem to need a high expression of *lpg1* gene compared to the others. This could suggest that the upregulation of this gene is essential for the promastigote form. This could be correlated with the fact that *lpg1* is involved in the biosynthesis of LPGs that are highly expressed in the promastigote stage.

This work was driven by three objectives: i) Clone, overexpress and purify the four LPG1x of *Leishmania major* using *E. coli* as an expression host. ii) Characterize the non-glycosylated enzymes, evaluate their activity and identify their substrates and obtention of crystal proteins, iii) Clone, express and purify the four glycosylated GalT's using *Leishmania tarentolae* as an expression host, and to assess their activity and study the effect of glycosylation.

Firstly, to obtain the proteins of interest, and facilitate their extraction and purification, the DNA regions that encode for the transmembrane domains that were localized only in N-terminus of proteins and distant from their catalytic site were removed. Then the genes were cloned into different plasmids and transformed into *E. coli*. As a result, none of pET vectors were appropriate to produce pure and soluble. Moreover, co-purification with chaperonin GroEL suggested that the proteins of interest are not well folded during their production. Different strategies were adopted to get rid of the chaperonin but they were not efficient. Consequently, other vector, pMAL-c2X, was used and finally allowed the obtention of pure and soluble LPG1x fused with MBP-Tag with a good yield, 5mg/L of culture for LPG1 and LPG1R and 10 mg/L of culture for LPG1G and LPG1L. In addition, the removal of the MBP-Tag is possible thanks to the activity of the factor X in case we want to study only the protein of interest. (Figure VI-1) These experiments proved that the choice of expression vector is very important for the protein production and that the bacteria *E. coli* can be used to produce these eukaryotic proteins.



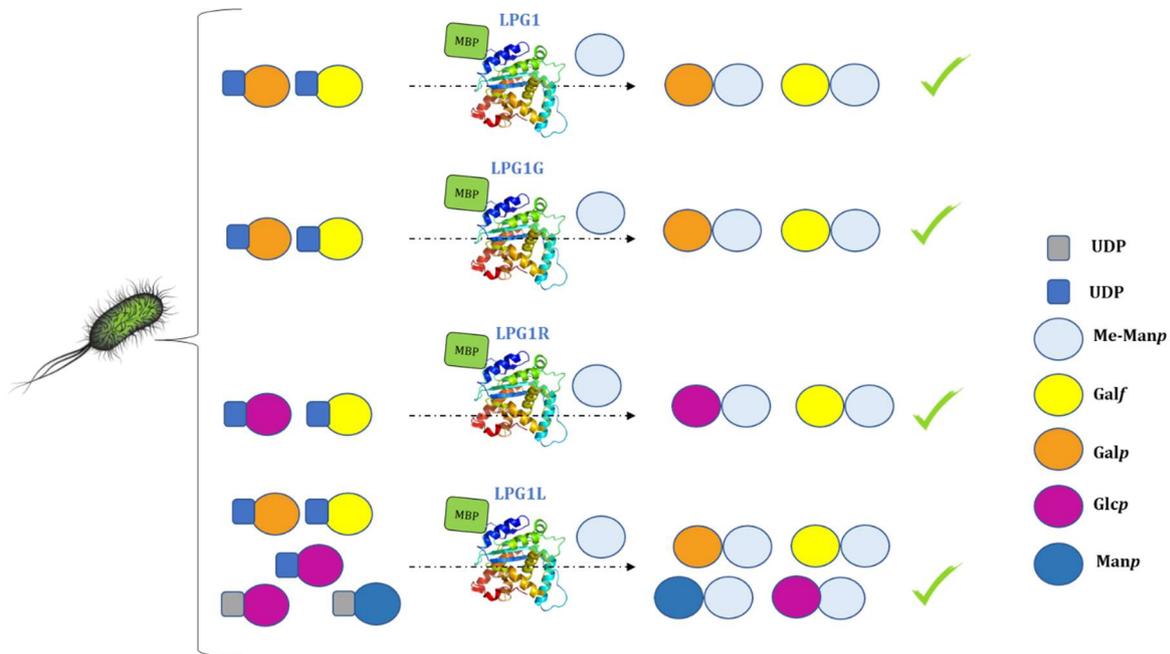
**Figure VI-1:** Summary of LPG1x production using pET and pMAL vectors.

For the **second objective**, the activity of the prokaryotic LPG1x proteins was assessed. The same spectrophotometric assays that were developed by Lowary and co-workers to evaluate the activity of mycobacterial GlfT2 was adopted. All the four enzymes LPG1, LPG1R, LPG1L and LPG1G recognized UDP-Galf as donor (Figure VI-2). They demonstrated **good galactofuranosyl-transferase activity** with apparent  $K_M$  values ranging from 20 to 550  $\mu\text{M}$ . It was revealed that LPG1R binds UDP-Galf more weakly and that its UDP-Galf  $K_M$  value is comparable to the one reported for mycobacterial GlfT2 (0.38 mM). In addition, the catalytic rate of LPG1 (30,750  $\text{min}^{-1}$ ), LPG1L (12,352  $\text{min}^{-1}$ ), and LP1G (5,296  $\text{min}^{-1}$ ) are at least 10-fold time higher than LPG1R with a value of 636  $\text{min}^{-1}$  and which is also comparable to the catalytic rate reported for GlfT2. The four LPG1x presented different kinetic parameters and are the first enzymes that belong to GT 40 family that have been characterized. Indeed, GlfT2 belongs to the GT family 2 and is a polymerizing enzyme that adds around 30 Galf units linked by alternating by  $\alpha$ -(1 $\rightarrow$ 5) and  $\alpha$ -(1 $\rightarrow$ 6) glycosidic bonds.<sup>1</sup> However, LPG1x belong to the family GT 40 and is predicted to attach only one galactofuranose molecules to the 3-OH position of the acceptor with  $\beta$ -selectivity. During this experiment, more than 30 different acceptors were tested but only Me-Manp was recognized by the four leishmanian GlfTs. It is important to mention that these the Galf residues that are described in the GIPLs and LPGs structures are attached to mannose. This could suggest that the four enzymes could be involved in the addition of Galf residues onto these glycoconjugates. However, single (*lpg1*-), double (*lpg1L*-/*lpg1R*-) and triple (*lpg1*-*lpg1L*-/*lpg1R*-) genes knockout studies have been performed and reported that only *lpg1* gene is involved of LPG biosynthesis, but *lpg1R* and *lpg1L* have no effect on LGPs or GIPLs synthesis. LPG1G can be responsible of the addition of Galf molecules to mannose residues on GIPLs. However, to confirm this hypothesis, it is essential to knock out the gene *lpg1G* and also generate multiple knock out (*lpg1*-*lpg1L*-/*lpg1R*- *lpg1G*-) parasites.

Surprisingly, the four LPG1x were also able to use UDP-pyranoses as donor substrates. LPG1, LPG1G and LPG1R were the less promiscuous as they were only able to recognize only one UDP-pyranose with a  $k_{\text{cat}}/K_M$  ranging from 1400 to 27.897  $\text{min}^{-1} \text{mM}^{-1}$ . Nevertheless, LPG1R  $k_{\text{cat}}/K_M$  for the UDP-Galp is 10-fold higher than the value registered with UDP-Galf. Thus, LPG1R exhibits a better glucopyranosyltransferase than galactofuranosyltransferase activity at least *in vitro*. While LPG1L seems to be the most promiscuous because it recognized UDP- $\alpha$ -D-Galp and UDP- $\alpha$ -D-Glcp in addition to GDP- $\alpha$ -D-Manp and GDP- $\alpha$ -D-Glcp even with lower specificity (Figure VI-2)

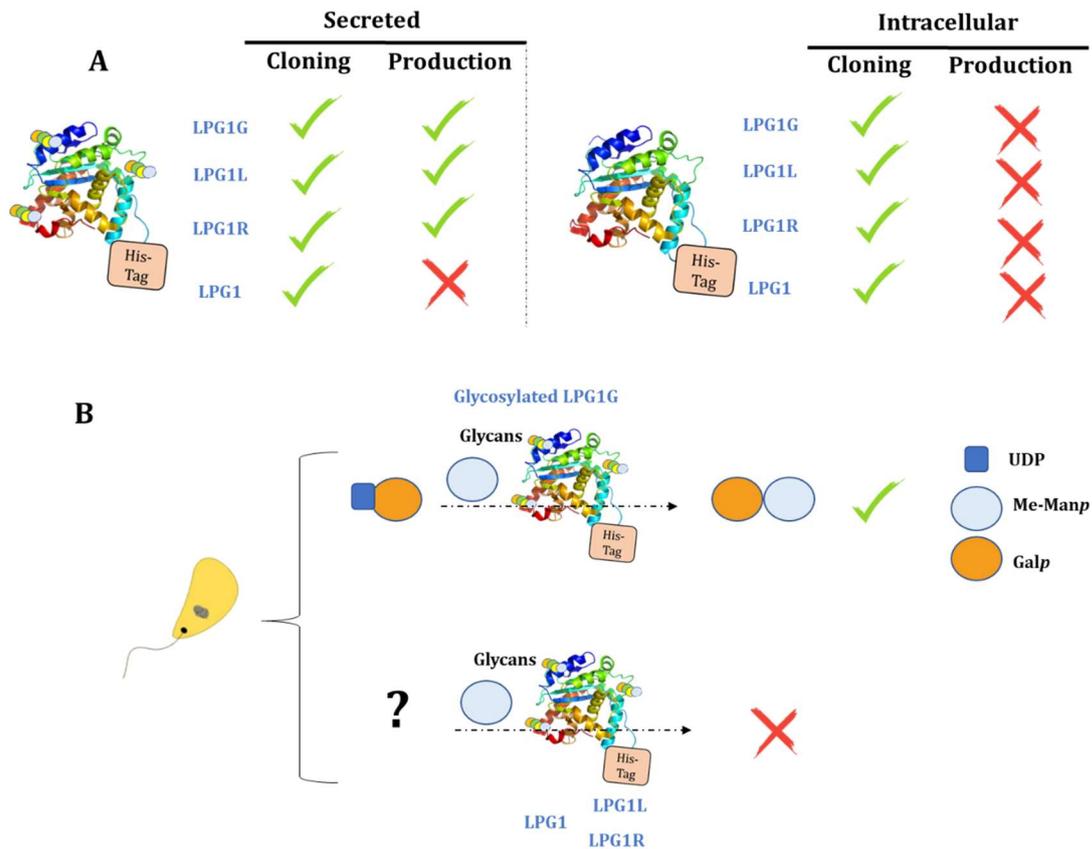
Since LPG1x were revealed to be active, crystallization trials were performed in order to obtain protein crystals. Only LPG1 crystals were obtained with 30%PEG 8000 and 200Mm of ammonium sulfate, however, X-ray diffraction was not successful.

This work provided the first enzymatic characterization of eukaryotic GalTs that belong to the CAZY GT family 40. Moreover, these enzymes were produced by prokaryotic host, and despite the absence of glycosylation, they displayed a good activity.



**Figure VI-2:** Summary of the substrate recognized by the four LPG1x produced by *E. coli*

The **third aim of this project** was to clone and express the same four LPG1x in eukaryotic host, using the non-pathogenic *Leishmania tarentolae*. **The four genes were successfully cloned** into specific plasmid pLEXY that was designed by *Jena Bioscience*, and two strategies of cloning were adopted in order to obtain secreted and glycosylated or intracellular and non-glycosylated forms. This would allow us to evaluate the effect of the glycosylation on the stability and the activity of these enzymes and to compare also with the prokaryotic LPG1x. During this thesis, only secreted LPG1G, LPG1L and LPG1R were produced and purified. In addition, we confirmed that the presence of *N*-glycans on the LPG1G added during the posttranslational modifications pathways that occur in the endoplasmic reticulum and Golgi apparatus of *Leishmania tarentolae*. However, we did not manage to validate the presence of other type of glycosylation such as *O*-glycosylation, C-glycosylation, glypiation and phosphoglycosylation. The activity of LPG1G was evaluated and preliminary results showed that this **eukaryotic enzyme is active and able to recognize UDP-Galp as donor and Me-Manp as an acceptor**. The  $k_{cat}/K_M$  values for UDP-Galp was  $3300 \text{ min}^{-1} \text{ mM}^{-1}$  which is 10-fold time lower than the catalytic efficiency of the prokaryotic LPG1G for the same donor. This difference could be caused by the presence of glycans. Nevertheless, we did not manage to do further experiments, and we did not assess the activity of the other secreted GalTs, LPG1, LPG1R and LPG1L (Figure VI-3).



**Figure VI-3:** Summary of the glycosylated and non-glycosylated LPG1x produced by *Leishmania tarentolae* (A) and their activity (B).

This thesis **provided a first expression of eukaryotic glycosylated and non-glycosylated form** of galactofuranosyltransferase that belong to the **GT40 family**. This will help enriching our knowledge about GTs involved in the transfer of rare sugars.

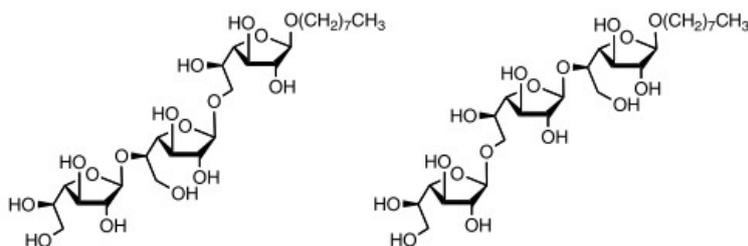
All the results obtained during this thesis open several research perspectives:

- **Short term objectives**
  - *Substrate specificity*

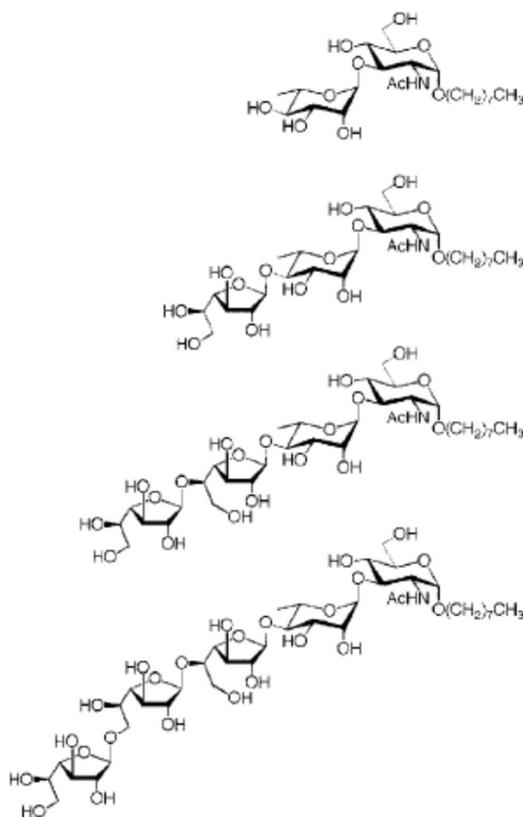
First, it will be interesting to test the activity of prokaryotic GalT's but after removing the MBP-Tag that is characterized with a high molecular weight ~ 42kDa. Even if the fused proteins proved to be active, it is interesting to see if we could obtain better affinity toward substrates. We should also produce and purify the proteins that have never been expressed during this thesis such as the secreted LPG1, the intracellular forms of the four LPG1x, in addition to de-glycosylated forms. Then, evaluating their activities in triplicates using the same tri-enzymatic assays. This will give a more precise idea and about their activities, and their kinetic parameters and glycosylation effect.

Then, the Methyl-  $\alpha$ -D-Mannose is recognized as an acceptor by the four LPG1x. However, we are convinced that we can find better acceptor if we use similar polysaccharides to those found in the

LPGs and GIPLs structures. Indeed, galactofuranose residues are attached to at least two consecutive mannose molecules, so, it will be interested to test different parts of the polysaccharide chains containing mannose residues in the terminus regions to identify the minimal structure necessary for the activity. With a similar aim in mind, Lowary and co-workers synthesized different portion of mycobacterial arabinogalactan to evaluate the activity and affinity of the galactofuranosyltransferases *GlT1* and *GlT2*. This allowed them identified the dual activity and real substrate of *GlT1* and *GlT2*<sup>2</sup> (Figure VI-4/5).



**Figure VI-4:** Reported trisaccharide substrates for *GlT2*.



**Figure VI-5:** Structures of the synthetic acceptors tested for *GlT1*

Moreover, this strategy was adopted by Adrian P. Higson et al., who synthesized fragments of phosphoglycan portion of LPGs expressed on the surface of *Leishmania mexicana* and *Leishmania major* in order to test them as substrate for the elongating  $\alpha$ -D-Mannopyranosylphosphate Transferase of *Leishmania*.<sup>3</sup> They were able to create a bank of leishmanian phosphorylated linear

or branched polysaccharides fragments as showed in Figure IV-8, and evaluate their affinity. A collaboration could be proposed to test their product with our galactofuranosyltransferases.

➤ *Active site identification*

It will also be interesting to identify the amino acid that bind the substrate in the active site. To do that, we can optimize the conditions to obtain protein crystals that are able to diffract. Thus, we can reveal for the first time the three-dimensional structure of the leishmanian Gal/Ts alone or with co-crystallization with ligands. This will help also the disclosing of the mechanism that these enzymes use to transfer galactofuranose molecules onto acceptor. In addition, mutants can be generated by site directed mutagenesis method to identify the amino-acid that are involved in the recognition and transfer of the donor. This will help identifying their exact role and the effect of their absence on the activity of the enzyme.

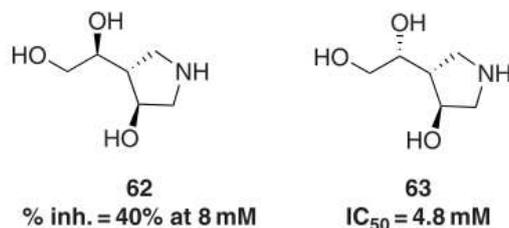
- Long term objective

➤ *Biological role of LPG1x*

Mutant strains of *Lieshmania* lacking the genes of *lpg1*, *lpg1R*, *lpg1L* and *lpg1G* could be considered. They can be used to infect mammalian cells and study their biological role in the cell interaction and infection process.

➤ *Inhibitors synthesis*

LPG1x are an interesting therapeutic target for the development of novel drugs against leishmaniosis. Thus, it is important to identify potent inhibitors to inactivate these leishmanial Gal/Ts. Analogues of enzymes donors and acceptors or transition state are the most common used GTs inhibitors.<sup>4</sup> Glycoscientists believe that analogues that mimic the transition state are very a valuable and are expected to fix the enzyme more strongly than the natural substrate or analogues in the ground state.<sup>5</sup> Bisubstrate analogues that are constituted of the glycosyl donor and the acceptor in some specific arrangement that simulate the transition state complex are very attractive because they resulted higher specificity for some GTs.<sup>6</sup> One of the structural modification that was very studied lately for carbohydrate mimics is based on the replacement of the endocyclic oxygen atom with a nitrogen atom, which produces an iminosugars.<sup>7</sup> The amino groups are protonated under physiological conditions and can interact with anionic groups in the active site of the enzymes. Consequently, they are considered as attractive targets as promising and potent potential analogues for the transition states. Indeed, they have been tested as inhibitors for different glycosidase and GTs.<sup>8,9</sup> As an example, we can cite the two iminosugars showed in Figure VI-6 that were synthetized and were found to be moderate inhibitors of mycobacterial Gl/T2 with IC50 values in the millimolar range.<sup>10</sup> Similar strategies can then be adopted for our leishmanial Gal/T.

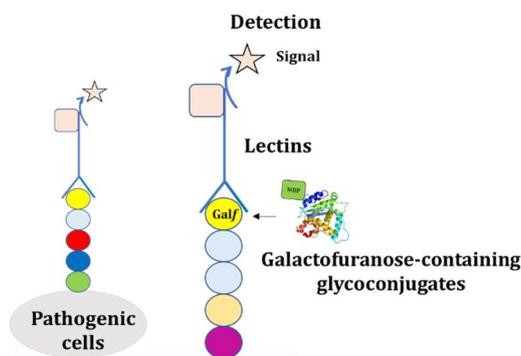


**Figure VI-6:** Imminosugars tested as substrate with GlfT<sup>10</sup>

➤ *Biocatalyst tools to synthesize galactofuranose-containing glycoconjugates*

In addition to their therapeutic potential, glycosyltransferases are powerful biocatalyst.<sup>11</sup> Indeed, it is common to use sugars for the pharmacological properties and the activity of compounds.<sup>12</sup> However, the synthesis of oligosaccharide can be laborious and can take a lot of time.<sup>13</sup> Chemical tools are often used with multiple protection, deprotection and purification steps, the yield is not always efficient and some combinations and linkages between monosaccharides are not easy to create.<sup>14,15</sup> Here, comes the role of glycosyltransferases. They offer a simplest protocol to obtain various glycoconjugates. Nowadays, thanks to chemoenzymatic methods, it is possible to combine the high efficiency and selectivity of enzymatic methods with the flexibility of chemical synthesis.<sup>16,17</sup> Thus, these galactofuranosyltransferases could be used to synthesize not only galactofuranose-containing glycoconjugates but different kind of polysaccharides thanks to their versatility, especially LPG1L.

Finally, these GalfTs could be used to synthesize similar galactofuranose-containing glycoconjugates to those expressed on the surface of pathogenic species to be used for the development of diagnosis tools. This could be designed in association with lectins that recognize Galf residues. One soluble human Intelectin was identified by Shoutaro Tsuji et al., that detect the galactofuranose present on carbohydrate chains of bacterial cell wall (Figure VI-7).<sup>18</sup>



**Figure VI- 7:** Detection strategies of galactofuranose-containing glycoconjugates synthesized by GalfTs



Les glycosylations sont le coeur de la glycobiologie, et sont considérées par la communauté scientifique comme les modifications post-traductionnelles les plus complexes qui affectent les protéines. La glycosylation ne résulte pas d'une expression et traduction d'une matrice d'ADN ou d'ARN, ce qui rend leur prédiction une tâche difficile. Des différents biopolymères peuvent être glycosylés pour donner naissance à des glycoconjugués comme les glycolipides et les glycoprotéines. Ces derniers jouent un rôle très crucial dans la communication et le développement des cellules saines ou cancéreuses, ainsi que dans les interactions entre les cellules hôtes et les pathogènes. Aujourd'hui, il est bien connu que certaines enzymes, nommées glycosyltransférases (GTs), sont responsables de la biosynthèse de ces glycoconjugués et sont considérées comme des cibles intéressantes pour le développement de nouveaux médicaments ainsi que de moyens d'analyses innovants. Cependant, ces GTs restent très peu étudiés.

Au cours de cette thèse, nous avons étudié des galactofuranosyltransférases (Gal $f$ Ts) qui sont des enzymes qui catalysent l'incorporation de monosaccharide rare, le galactofuranose (Gal $f$ ), au niveau des chaînes polysaccharidiques des glycoconjugués. Le Gal $f$  est essentiellement présent dans différentes espèces pathogènes tels que, *Aspergillus*, *Trypanosoma*, *Mycobacterium* et *Leishmania*, cependant, il est absent chez les cellules mammifères. Il est important de noter que nos connaissances concernant ces Gal $f$ Ts sont très limitées. En effet, seulement quatre Gal $f$ Ts ont été identifiés et étudiés, GfsA d'*Aspergillus*, Wbb1 d'*E. coli*, GlfT1 et GlfT2 de *Mycobacterium tuberculosis*, mais uniquement GlfT2 a été entièrement caractérisée. Nous nous sommes donc particulièrement intéressés aux Galactofuranosyltransférases de *Leishmania* qui n'ont jamais été exprimés au paravent. *Leishmania* est un parasite qui cause la leishmaniose qui est classée par l'Organisation mondiale de la santé comme étant une maladie tropicale négligée, et qui est spécifiquement endémique dans les pays du tiers monde.

Les médicaments actuels qui sont disponibles sur le marché pour traiter la leishmaniose, sont soit trop chers, soit toxiques. De plus, les leishmanies ont développé une résistance à certains médicaments constituant ainsi une des plus grandes menaces actuelles envers la santé humaine. Par conséquent, il est donc urgent de trouver de nouvelles stratégies efficaces pour lutter contre cette maladie. Le parasite *Leishmania* exprime à sa surface certains glycoconjugués tels que les lipophosphoglycanes et les glycosylinositols-phospholipides qui se sont révélés être des acteurs majeurs dans les mécanismes d'infection et de survie de ces pathogènes. Ces glycoconjugués contiennent du Gal $f$ , cette particularité a incité beaucoup d'attention, spécialement envers les Gal $f$ T. En effet, elles sont considérées comme des cibles thérapeutiques innovantes, d'où la nécessité de les étudier.

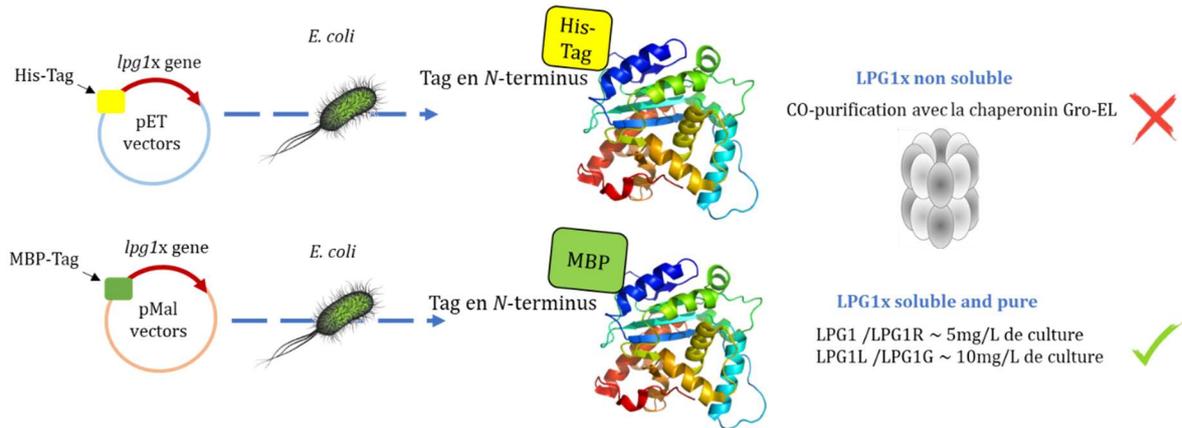
Quatre gènes *lpg1*, *lpg1L*, *lpg1R* et *lpg1G* ont été identifiés dans le génome de *Leishmania major*. Les trois premiers gènes sont présents en une seule copie aux loci suivants : LMJF\_25\_0010 (*lpg1*),

LMJF\_33\_0300 (*lpg1L*), LMJF\_26\_0550 (*lpg1R*), mais *lpg1G* existe en trois copies identiques à des loci différents LMJF\_32\_3990, LMJF\_05\_1230 and LMJF19\_650. Ces gènes devraient coder pour des potentiels GalTs transmembranaires qui seraient situés au niveau de l'appareil de Golgi des leishmanies. De plus, leurs séquences peptidiques ne partagent aucune homologie significative avec les autres GalTs. Des résultats préliminaires issus des expériences de RTqPCR ont démontré que ces quatre gènes sont transcrits chez différentes espèces de *Leishmania*, *L. major*, *L. donovani* et *L. infantum*, qui ont été cultivées sous la forme des promastigote. De plus, leurs expressions augmentent en fonction du temps, et il semblerait que ces parasites nécessitent une expression élevée du gène *lpg1* par rapport aux autres. Ces observations pourraient suggérer que cette régulation d'expression de gène, notamment *lpg1*, est essentielle pour le développement et la survie de la forme promastigote. Ceci pourrait être corrélé avec l'implication du gène *lpg1* dans la biosynthèse des LPGs qui sont fortement exprimés chez les promastigotes.

Ce travail avait trois objectifs : i) Cloner, surexprimer et purifier les quatre LPG1x de *Leishmania major* en utilisant *E. coli* comme hôte d'expression. ii) Caractériser ces enzymes sous la forme non glycosylées, évaluer leurs activités, identifier leurs substrats et obtenir des cristaux de protéines, iii) Cloner, exprimer et purifier les quatre GalT sous la forme glycosylée en utilisant *Leishmania tarentolae* comme hôte d'expression, évaluer leurs activités et en évaluer l'effet de la glycosylation.

Premièrement, pour obtenir les protéines d'intérêt et faciliter leurs extractions et leurs purifications, nous avons clivé les régions d'ADNs codants pour les domaines transmembranaires qui sont localisés au niveau des extrémité *N*-terminales. Ensuite, les gènes ont été clonés dans différents plasmides pET et transformés dans *E. coli*. Cependant, aucun de ces vecteurs n'étaient appropriés pour donner des protéines pures et solubles. De plus, leur co-purification avec la chaperonne GroEL a laissé suggérer que les protéines obtenues ne se sont pas bien repliées. Par conséquent, un autre vecteur, pMAL-C2X, a été testé et qui a finalement permis l'obtention des quatre LPG1x pures et solubles fusionné avec un bon rendement, 5 mg / L de culture pour LPG1 et LPG1R et 10 mg / L de culture pour LPG1G et LPG1L. La particularité de ce plasmide consiste dans l'élimination du Tag-MBP qui est envisageable grâce à l'activité du facteur Xa, dans le cas où nous souhaiterions étudier uniquement la protéine d'intérêt (Figure VI-1).

Ces expériences ont prouvé l'importance du choix du vecteur d'expression pour la production de protéines solubles et que la bactérie *E. coli* pouvait être utilisée pour produire des protéines eucaryotes.



**Figure VI- 8:** Un résumé des stratégies utilisées pour produire les LPG1x en utilisant les vecteurs pET and pMAL.

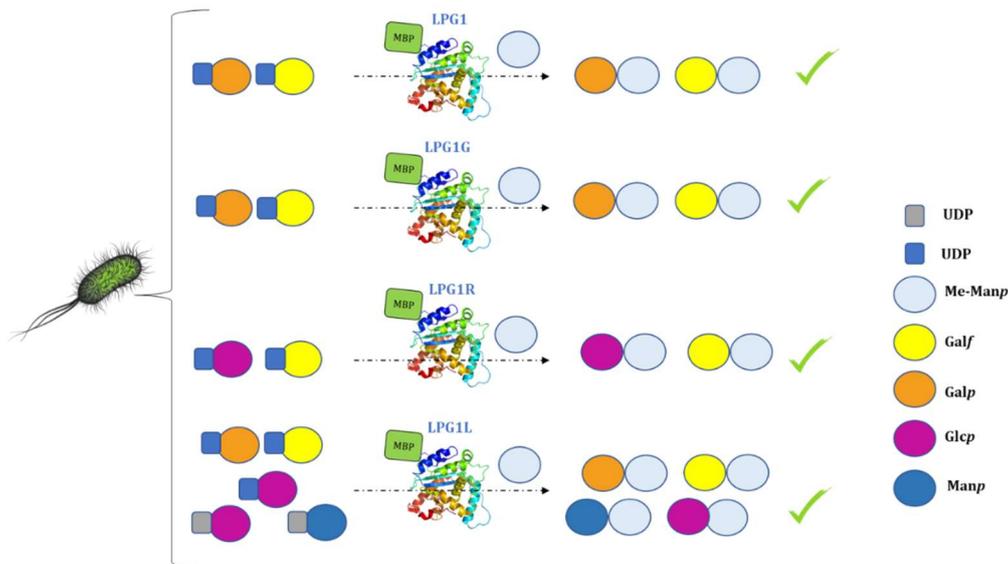
Concernant le deuxième objectif, l'activité des protéines procaryotes LPG1x a été étudiée. Des analyses spectrophotométriques similaires que celles utilisées par Lowary et ses collaborateurs pour évaluer l'activité de la mycobactérie *GlfT2* ont été utilisées. Les quatre enzymes LPG1, LPG1R, LPG1L et LPG1G ont reconnu l'UDP-Galf en tant que donneur (Figure VI-2). Nous avons démontré une bonne activité galactofuranosyl-transférase avec des valeurs de  $K_M$  apparentes allant de 20 à 550  $\mu$ M. Nous avons aussi révélé que LPG1R se lie à l'UDP-Galf avec moins d'affinité et que sa valeur  $K_M$  est comparable à celle enregistrée pour *GlfT2* de *Mycobacterium tuberculosis* (0,38 mM). De plus, les vitesses catalytiques de LPG1 (30 750  $\text{min}^{-1}$ ), LPG1L (12 352  $\text{min}^{-1}$ ) et LP1G (5 296  $\text{min}^{-1}$ ) sont au moins 10 fois supérieures à celles de LPG1R qui a une valeur de 636  $\text{min}^{-1}$  qui est également comparable au taux catalytique rapporté pour *GlfT2*. Les quatre LPG1x ont présenté différents paramètres cinétiques et sont considérées comme les premières enzymes caractérisées appartenant à la famille GT 40. En effet, *GlfT2* appartient à la famille GT 2 et est connue comme étant une enzyme polymérase qui incorpore successivement environ 30 unités de Galf en alternant des liaisons glycosidiques de type  $\alpha$ - (1  $\rightarrow$  5) et  $\alpha$ - (1  $\rightarrow$  6).<sup>1</sup> Alors que, LPG1x appartiennent à la famille GT 40 et il est attendu que ces enzymes ajoutent uniquement une seule molécule de galactofuranose qui soit liée à la position 3-OH de l'accepteur avec une sélectivité  $\beta$ . Au cours de ces expériences, plus de 30 différents accepteurs ont été testés, mais seul le Me-Manp a été reconnu par les quatre *GalfT* de leishmanie. Il est important de mentionner que les glycoconjugués GIPLs et LPGs contiennent dans leurs structures des résidus Galfs qui sont liés au mannose. Cela pourrait suggérer que les quatre enzymes pourraient être impliquées dans l'addition de résidus de Galf sur ces glycoconjugués. Cependant, des études antérieures basées sur l'inactivation de gènes simples (*lpg1-*), doubles (*lpg1L-* / *lpg1R-*) ou triples (*lpg1 - lpg1L-* / *lpg1R-*) ont été menées et ont démontré que seul le gène *lpg1* joue un rôle dans la biosynthèse des LPGs. Alors qu'aucun des gènes *lpg1R* ou *lpg1L* ne semble avoir un effet

sur la biosynthèse des glycoconjugués LGP et GIPL. Le gène *lpg1G* quant à lui n'a jamais été inactivé, mais nous pouvons supposer qu'il puisse être responsable de l'addition de molécules de Galf sur le mannose au niveau des structures GIPLs. Cependant, pour confirmer cette hypothèse, il est essentiel d'inactiver le gène *lpg1G* et de générer plusieurs souches de parasites qui soient génétiquement modifiés (*lpg1-lpg1L-* / *lpg1R-lpg1G-*).

De manière surprenante, les quatre LPG1x ont également pu utiliser des UDP-pyranoses comme substrats donneurs. LPG1, LPG1G et LPG1R ont été les moins prometteurs, car ils n'étaient capables de reconnaître qu'un seul UDP-pyranose avec un  $k_{cat} / K_M$  allant de 1400 à 27,897 min<sup>-1</sup> mM<sup>-1</sup>. Néanmoins, les valeurs du ratio  $k_{cat} / K_M$  enregistré par LPG1R pour UDP-Galp est 10 fois plus élevé que la valeur enregistrée avec UDP-Galf. Ainsi, LPG1R présente une meilleure activité glucopyranosyl-transférase que celle de galactofuranosyl-transférase, du moins en condition *in vitro*. Alors que LPG1L semble être la plus prometteuse parce qu'elle reconnaît l'UDP- $\alpha$ -D-Galp et UDP- $\alpha$ -D-Glcp en plus de GDP- $\alpha$ -D-Manp et GDP- $\alpha$ -D-Glcp avec une spécificité moins importante que pour UDP-Galf (Figure VI-2).

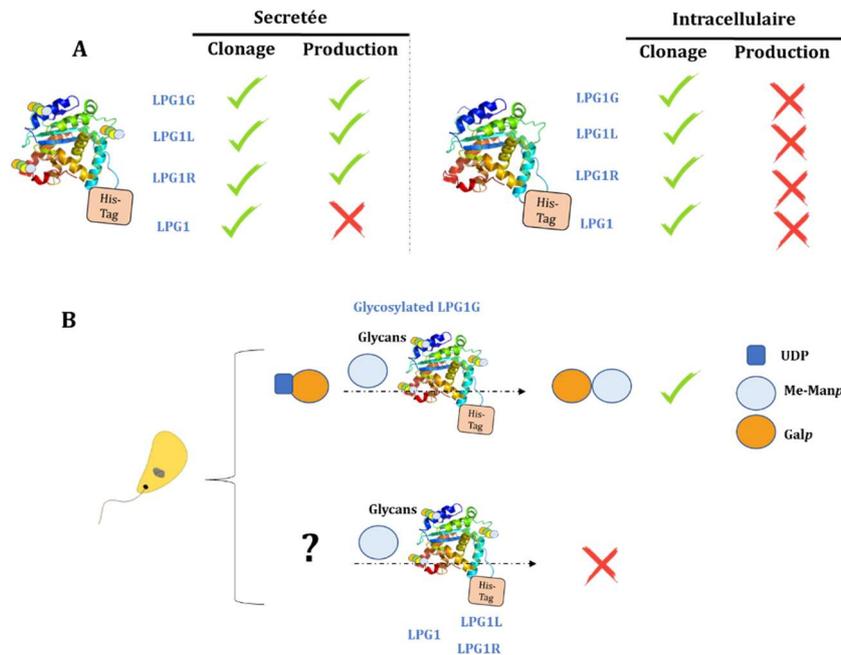
Étant donné que les enzymes LPG1x se sont révélées actives, des essais de cristallisations ont été donc réalisés afin d'obtenir des cristaux de protéines. Seuls des cristaux de LPG1 contenant 30% de PEG 8000 et 200 mM de sulfate d'ammonium ont été obtenus. Cependant, la diffraction des rayons X n'a pas été concluante.

Ce travail a fourni une première caractérisation enzymatique de GalfT eucaryotes appartenant à la famille CAZY GT 40. De plus, ces enzymes ont été produites par un hôte procaryote et, malgré l'absence de glycosylation, elles ont présenté une bonne activité.



**Figure VI- 9:** Un résumé des substrats qui sont reconnu par les quatre protéines LPG1x qui sont produites par *E. coli*.

Le troisième objectif de ce projet était de cloner et d'exprimer ces mêmes LPG1x dans un hôte eucaryote, en utilisant *Leishmania tarentolae*, une souche non pathogène pour l'homme, comme hôte d'expression. Les quatre gènes ont été clonés avec succès dans un plasmide spécifique pLEXSY conçu par Jena Bioscience. Deux stratégies de clonage ont été adoptées afin d'obtenir des formes sécrétées et glycosylées ou intracellulaires et non glycosylées. Cela nous permettrait d'évaluer l'effet de la glycosylation sur la stabilité et l'activité de ces enzymes et de comparer également avec les résultats obtenus par les LPG1x procaryote. Au cours de cette thèse, seuls les protéines LPG1G, LPG1L et LPG1R sous la forme sécrétée ont été produites et purifiées. Ensuite, nous avons pu confirmer la présence des *N*-glycanes au niveau des LPG1G. Elles ont donc été ajoutées au cours des voies de modification post-traductionnelles qui ont été effectuées au niveau du réticulum endoplasmique et l'appareil de Golgi de *Leishmania tarentolae*. Cependant, nous n'avons pas réussi à valider la présence d'autres types de glycosylation tels que la *O*-glycosylation, la *C*-glycosylation, la glypiation ou la phosphoglycosylation. L'activité de LPG1G a été évaluée et des résultats préliminaires ont montré que cette enzyme eucaryote est active et capable de reconnaître l'UDP-Galp en tant que donneur et le Me-Manp en tant qu'accepteur. Les valeurs de  $k_{cat}/K_M$  pour UDP-Galp étaient de  $3300 \text{ min}^{-1} \text{ mM}^{-1}$ , soit 10 fois moins que l'efficacité catalytique du procaryote LPG1G pour le même donneur. Cette différence pourrait être causée par la présence de glycanes. Néanmoins, nous n'avons pas réussi à faire d'autres expériences pour évaluer l'activité des autres GalT's, LPG1, LPG1R et LPG1L de forme sécrétée ni les formes intracellulaires (Figure VI-3).



**Figure VI- 10:** Résumé des protéines glycosylées et non glycosylées qui ont été obtenues (A) ainsi que leur activités (B).

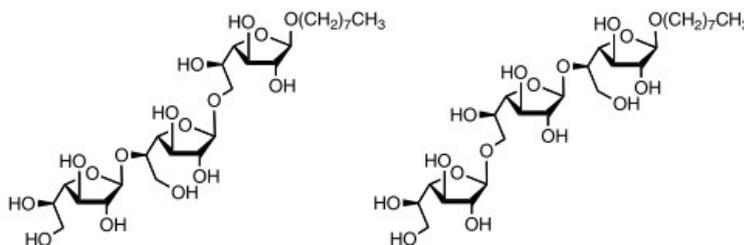
Cette thèse reporte les premières expériences d'expression des galactofuranosyltransférases de *Leishmania major* appartenant à la famille GT40 sous la forme glycosylée et non glycosylée. Cette étude peut contribuer à enrichir nos connaissances sur les GTs qui sont impliquées dans le transfert de sucres rares.

Tous les résultats obtenus au cours de cette thèse ouvrent plusieurs perspectives de recherches :

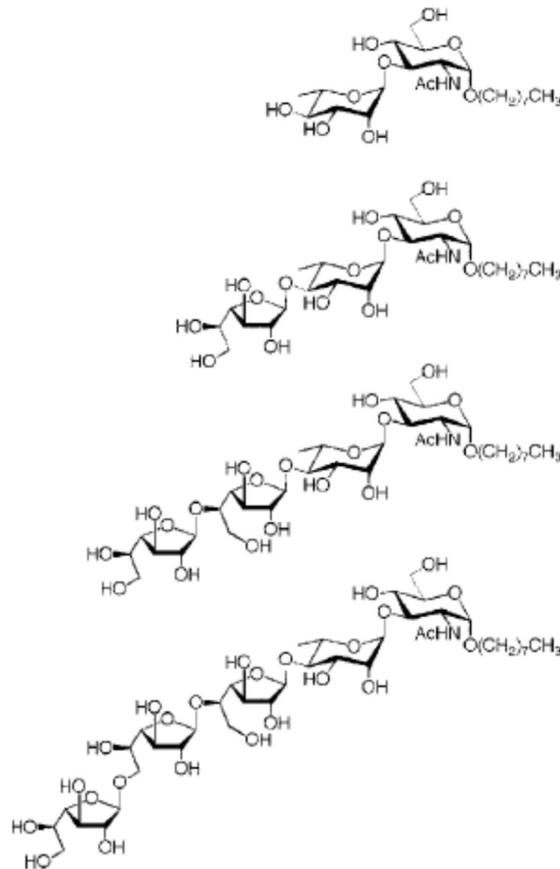
- Objectives à court termes
  - Spécificité du substrat

Tout d'abord, il serait intéressant de tester l'activité des GalT procaryotes sans la présence du Tag-MBP qui est caractérisé par un poids moléculaire élevé, d'environ 42 kDa. En effet, même si les protéines recombinantes qui sont fusionnées avec ce gros Tag se sont avérées actives, il est intéressant de voir si nous pourrions obtenir une meilleure affinité pour les substrats uniquement avec nos protéines d'intérêt. Nous devrions également produire et purifier les protéines qui n'ont pas pu être exprimées au cours de cette thèse, telles que la LPG1 sécrétée, les formes intracellulaires des quatre LPG1x, en plus des formes eucaryotes dé-glycosylées. Ensuite, leurs activités devraient être évaluée en utilisant les mêmes essais tri-enzymatiques. Cela donnera une idée plus précise concernant leurs activités, leurs paramètres cinétiques, ainsi que l'effet de la glycosylation.

Le méthyl- $\alpha$ -D-Mannose est reconnu comme accepteur par les quatre LPG1x, cependant, nous sommes convaincus que nous pouvons trouver un meilleur accepteur si nous utilisons des polysaccharides similaires à ceux trouvés dans les structures des LPGs et des GIPLs. En effet, le galactofuranose est attaché à au moins deux molécules de mannose consécutives, il sera donc intéressant de tester différentes parties de ces chaînes polysaccharidiques contenant des résidus de mannose dans les régions terminales pour identifier la structure minimale nécessaire à l'activité de nos enzymes. Dans un but similaire, Lowary et ses collaborateurs ont synthétisé différentes portions d'arabinogalactane mycobactérien afin d'évaluer l'activité et l'affinité des galactofuranosyltransférases GlfT1 et GlfT2 qui leur a permis d'identifier la double activité ainsi que le vrai substrat de leur GalTs<sup>2</sup> (Figure VI-4/5).



**Figure VI- 11:** Les substrats tri-saccharidiques de GlfT2.



**Figure VI- 12:** Structures des accepteurs qui sont reconnues pour GlfT1

De plus, cette stratégie a été adoptée par Adrian P. Higson et al. Ils ont synthétisé des fragments de la partie phosphoglycane des LPG qui sont exprimés à la surface de *Leishmania mexicana* et de *Leishmania major* afin de les tester en tant que substrat de la  $\alpha$ -D-Mannopyranosylphosphate Transférase of *Leishmania*.<sup>3</sup> Ils ont pu créer une banque de fragments de polysaccharides linéaires ou ramifiés et phosphorylés de leishmanies et évaluer leur affinité. Une collaboration pourrait être proposée pour tester leur produit avec nos galactofuranosyltransférases.

➤ *Identification du site actif*

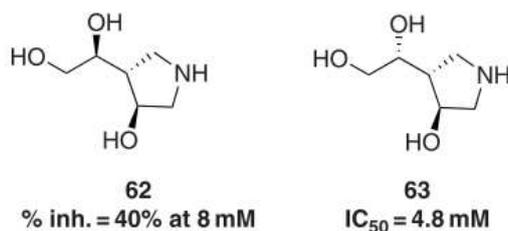
Il serait également intéressant d'identifier les acides aminés qui se lient aux substrats au niveau du site actif de l'enzyme. Pour ce faire, nous pouvons optimiser les conditions pour obtenir des cristaux de protéines capables de diffracter. Ainsi, nous pouvons révéler pour la première fois la structure tridimensionnelle des GalfT de *Leishmania* seuls ou avec une co-cristallisation avec des ligands tel que l'UDP-Galf. Cela contribuera également à dévoiler le mécanisme que ces enzymes utilisent pour transférer les molécules de galactofuranose sur l'accepteur. De plus, des mutants peuvent être générés avec une méthode de mutagenèse dirigée pour identifier les acides aminés impliqués dans la reconnaissance et le transfert du donneur. Cela aidera à identifier leur rôle exact et l'effet de leur absence sur l'activité de l'enzyme.

- Objectives à long terme
  - *Role biologique des LPGx*

Des souches mutantes de *Leishmania* dépourvues des gènes *lpg1*, *lpg1R*, *lpg1L* et *lpg1G* pourraient être envisagées. Ils peuvent être utilisés pour infecter des cellules de mammifère et étudier leurs rôles biologiques dans l'interaction cellulaire avec les cellules immunitaires ainsi le processus d'infection.

- *Synthèse d'inhibiteurs*

LPG1x est une cible thérapeutique intéressante pour le développement de nouveaux médicaments contre la leishmaniose. Il est donc important d'identifier de puissants inhibiteurs qui nous permet d'inactiver ces GalT de leishmanies. Des analogues de donneurs et accepteurs d'enzymes ou des molécules présent pendant l'état de transition sont considérés comme des inhibiteurs de GTs les plus utilisés.<sup>4</sup> Les glyco biologistes et les glycochimistes estiment que les analogues mimant l'état de transition sont très prometteurs car ils devraient fixer l'enzyme avec une meilleur affinité que le substrat naturel ou les analogues qui sont présent à l'état fondamental.<sup>5</sup> Les analogues bi-substrats qui sont constitués du donneur et de l'accepteur qui se trouvent dans un arrangement spécifique simulant le complexe d'états de transition sont très attractifs car ils présentent une spécificité plus élevée pour certains GTs.<sup>6</sup> L'une des modifications structurales récemment étudiées pour les molécules mimant des glucides est basée sur le remplacement de l'atome d'oxygène endocyclique par un atome d'azote, formant ainsi un iminosucre.<sup>7</sup> Les groupes amino sont protonés dans des conditions physiologiques et peuvent interagir avec des groupes anioniques présents dans le site actif des enzymes. Par conséquent, ils sont considérés comme des cibles attrayantes en tant qu'analogues potentiellement puissants pour les états en transition. En effet, ces molécules ont été testés en tant qu'inhibiteurs de différentes glycosidases et de GTs.<sup>8,9</sup> A titre d'exemple, nous pouvons citer les deux iminosucres montrés à la figure VI-6 qui ont été synthétisés et qui se sont révélés être des inhibiteurs modérés de la GlfT2 mycobactérienne avec des valeurs de IC<sub>50</sub> dans la gamme des millimolaires.<sup>10</sup> Des stratégies similaires peuvent donc être adoptées pour nos GalT de Leishmanie.

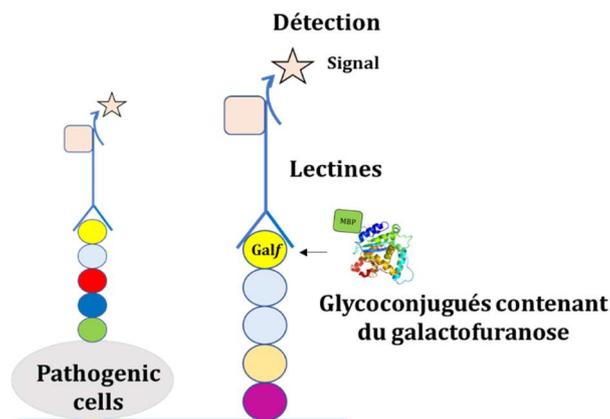


**Figure VI- 13:** Des exemples d'imminosucres testés comme inhibiteurs pour GlfT2<sup>10</sup>

- *Des outils de biocatalyse pour synthétiser des glycoconjugués contenant des galactofuranoses.*

En plus de leur potentiel thérapeutique, les glycosyltransférases sont de puissants biocatalyseurs.<sup>11</sup> En effet, il est courant d'utiliser des sucres pour les propriétés pharmacologiques.<sup>12</sup> Cependant, la synthèse des oligosaccharides peut être laborieuse et peut prendre beaucoup de temps.<sup>13</sup> Les méthodes chimiques sont souvent caractérisés par plusieurs étapes de protection, de déprotection, de purification et un rendement qui n'est pas toujours élevé. De plus, certaines combinaisons et liaisons entre monosaccharides ne sont pas faciles à créer.<sup>14,15</sup> C'est donc ici qu'intervient le rôle des glycosyltransférases car ils offrent un protocole simple pour obtenir divers glycoconjugués. De nos jours, grâce aux méthodes chimio-enzymatiques, il est possible de combiner l'efficacité et la sélectivité des enzymes avec la flexibilité de la synthèse chimique.<sup>16,17</sup> Ainsi, ces galactofuranosyltransférases pourraient être utilisées pour synthétiser non seulement des glycoconjugués contenant du galactofuranose, mais aussi différents types de polysaccharides grâce à leur polyvalence, en particulier la LPG1L.

Enfin, ces Gal $f$ T pourraient être utilisées pour synthétiser des glycoconjugués similaires à ceux contenant du galactofuranose et exprimées à la surface de différentes espèces pathogènes, afin de développer des outils de diagnostic. Cela pourrait être conçu en association avec des lectines qui reconnaissent les résidus de Gal $f$ . En effet, une Intélectine humaine soluble a été récemment identifiée par Shoutaro Tsuji et al., qui est capable de détecter le galactofuranose présent sur les chaînes glucidiques de la paroi cellulaire bactérienne (Figure VI-7).<sup>18</sup>



**Figure VI- 14:** Stratégie de détection des glycoconjugués contenant du Gal $f$  présents à la surface des espèces pathogènes grâce aux Gal $f$ Ts.



## References

1. Rose, N. L. *et al.* Development of a coupled spectrophotometric assay for GlfT2 , a bifunctional mycobacterial galactofuranosyltransferase. *Carbohydr. Res.* **343**, 2130–2139 (2008).
2. Belánová, M. *et al.* Galactosyl transferases in mycobacterial cell wall synthesis. *J. Bacteriol.* **190**, 1141–1145 (2008).
3. Higson, A. P. *et al.* Synthetic fragments of antigenic lipophosphoglycans from *Leishmania major* and *Leishmania mexicana* and their use for characterisation of the *Leishmania* elongating  $\alpha$ -D-mannopyranosylphosphate transferase. *Chem. - A Eur. J.* **11**, 2019–2030 (2005).
4. Lisa J. Whalen, William A. Greenberg, M. L. M. and C.-H. W. in *Iminosugars: From synthesis to therapeutic applications* 153–176 (2008).
5. Kraur, J. How do enzymes work ? *Science.* **152**, 41–43 (1988).
6. Waldscheck, B., Streiff, M., Notz, W., Kinzy, W. & Schmidt, R. R. A(1-3)-Galactosyltransferase Inhibition Based on a New Type of Disubstrate Analogue. *Angew. Chemie - Int. Ed.* **40**, 4007–4011 (2001).
7. GünterLegler. Glycoside Hydrolases: Mechanistic Information from Studies with Reversible and Irreversible Inhibitors. *Adv Carbohydr Chem Biochem.* **48**, 319–84 (1990).
8. Masson, G., Compain, P. & Martin, O. R. A new, stereocontrolled approach to iminosugar C-glycosides from L-sorbose. *Org. Lett.* **2**, 2971–2974 (2000).
9. Saotome, C., Wong, C. H. & Kanie, O. Combinatorial library of five-membered iminocyclitol and the inhibitory activities against glyco-enzymes. *Chem. Biol.* **8**, 1061–1070 (2001).
10. Sylvaine Cren, a Sudagar S. Gurcha, b Alexander J. Blake, a G. S. B. and & Thomas, N. R. Synthesis and biological evaluation of new inhibitors of UDP- Galf transferase—a key enzyme in. *Org. Biomol. Chem.* 2418–2420 (2004).
11. Palcic, M. M. Glycosyltransferases as biocatalysts. *Curr. Opin. Chem. Biol.* **15**, 226–233 (2011).
12. Solá, R. J. & Griebenow, K. A. I. Effects of Glycosylation on the Stability of Protein Pharmaceuticals. *Biochemistry* **98**, 1223–1245 (2010).

13. Gandolfi-Donadío, L., Santos, M., De Lederkremer, R. M. & Gallo-Rodriguez, C. Synthesis of arabinofuranose branched galactofuran tetrasaccharides, constituents of mycobacterial arabinogalactan. *Org. Biomol. Chem.* **9**, 2085–2097 (2011).
14. Lepenies, B., Yin, J. & Seeberger, P. H. Applications of synthetic carbohydrates to chemical biology. *Curr. Opin. Chem. Biol.* **14**, 404–411 (2010).
15. Saddam Muthana, Hongzhi Cao, and X. C. Recent Progress in Chemical and Chemoenzymatic Synthesis of Carbohydrates. *Curr. Protoc. Mol. Biol.* **13**, 1–23 (2014).
16. Drepper, T. *et al.* Novel biocatalysts for white biotechnology. *Biotechnol. J.* **1**, 777–786 (2006).
17. Richards, M. R. & Lowary, T. L. Chemistry and biology of galactofuranose-containing polysaccharides. *ChemBioChem* **10**, 1920 – 1938 (2009).
18. Tsuji, S. *et al.* Human Intelectin Is a Novel Soluble Lectin That Recognizes Galactofuranose in Carbohydrate Chains of Bacterial Cell Wall. *J. Biol. Chem.* **276**, 23456–23463 (2001).

This work was already published through articles, poster and oral communications.

### **Publications**

- **Ati, J.**, Lafite, P. & Daniellou, R. Enzymatic synthesis of glycosides: from natural O- and N-glycosides to rare C- and S-glycosides. *Beilstein J. Org. Chem.* 13, 1857–1865 (2017).
- **Ati, J.**, Colas, C., Lafite, P., Zheng, RB., Lowary, T L. & Daniellou R. The LGP1x family from *Leishmania major* is constituted of rare eukaryotic galactofuranosyltransferases with unprecedented catalytic properties, *Sci. Report.* 1<sup>st</sup> submission in June, currently under 2<sup>nd</sup> revision.

### **Oral communications**

- **Ati, J.**, SENICAR.M., Lafite, P. & Daniellou, R.  
The maker, the baker and the shapeshifter  
*French group of glycosciences -Mai 2018* – Nouan le fuzelier (FRA)
- **Ati, J.**, Manno.M., Lafite, P. & Daniellou, R.  
Glycosylate Me Lesxy !  
*Glycoserpin workshop -October 2017* – Orleans (FRA)
- **Ati, J.**, Lafite, P. & Daniellou, R.  
Galactofuranosyltransferases of *Leishmania*?  
*EnzBio – French Society of Biochemistry and Molecular Biology (SFBBM) - May 2017* – Le Croisic (FRA)

### **Poster communications**

- **Ati, J.**, Colas, C., Lafite, P., Zheng, RB., Lowary, TL., and Daniellou, R.  
LPG1x family from *Leishmania major*: four galactofuranosyltransferases with unexpected catalytic properties.  
*International Carbohydrate Symposium – ICS - July 2018* – Lisboa (Por)
- **Ati, J.**, Lafite, P. & Daniellou, R.  
Galactofuranosyltransferases of *Leishmania*: Novel Therapeutic Targets.  
*Structural Glycoscience workshop - June 2016* – Grenoble (FRA)
- **Ati, J.**, Lafite, P., Daniellou, R.  
Galactofuranosyltransferases of *Leishmania*: Novel Therapeutic Targets.  
*Biotechnocentre - October 2016* – Seillac (FRA)

### **Workshops**

- *Structural Glycosciences* : **June 2016** – Institute of Structural Biology at Grenoble (Fra)
- *Crystallography of proteins: from collected data to density map*: **June 2017** – Biologic station of Roscoff – University of Pierre Marie Curie – CNRS- (Fra)
- *PhD and entrepreneurship*: **June 2018** – Pépite France - The 1st network of “doctorant-entrepreneurs” (Fra)



## ***Chapter VII: Experimental part***

I.	DNA Cloning.....	208
A.	PCR and cloning strategies .....	208
B.	Vectors digestion and transformation in DH5 $\alpha$ cells.....	211
C.	Plasmid DNA purification, analysis and sequencing.....	212
II.	Expression and purification of proteins .....	212
A.	Prokaryotic proteins .....	212
1.	Transformation in and overexpression of proteins .....	212
2.	Lysis of cells .....	213
3.	Protein purification.....	214
B.	Eukaryotic proteins .....	214
1.	Culture of <i>Leishmania</i> .....	214
2.	<i>Leishmania</i> transfection.....	215
a)	DNA preparation .....	215
b)	Transfection and selection.....	215
c)	Purification of eukaryotic recombinant proteins .....	216
III.	Analysis of recombinant proteins.....	216
A.	Bradford protein assays.....	216
B.	Tryptic digestion of proteins .....	216
C.	Western Blot.....	216
1.	solutions .....	216
2.	Transfer .....	217
3.	Immunodetection .....	217
4.	Detection .....	217
IV.	Enzyme activity assays and kinetics parameters determination .....	217
A.	Malachite green assay.....	217
B.	HPLC detection .....	218
C.	Spectrophotometric assays .....	218
D.	Galactofuranosyltransferase preparative reactions .....	219
E.	Separation and TLC analysis.....	219
F.	Acetylation of sugars and analysis .....	219
V.	Cristallogenesi s of proteins.....	219

## I. DNA Cloning

### A. PCR and cloning strategies

The four genes *lpg1*, *lpg1L*, *lpg1R* and *lpg1G* were amplified by PCR from *Leishmania major* genomic DNA (provided by the Pr. Françoise ROUTIER from the Hannover Medical School) using Phusion high-fidelity DNA polymerase (Thermo fisher™), nucleotides (Thermo fisher™) and designed primers as described below synthesized by Eurofins Genomics.

In order to obtain prokaryotic proteins expressed in *E. coli* strains, PCR products were directly inserted into linearized pET (Novagen®) and pMAL (NEW ENGLAND BioLabs®*Inc*) vectors using T4 DNA Ligase (Rapid DNA ligation kit, Thermo fisher™). For this classical cloning, specific restrictions sites sequences were added at the 5' end of the primers (Table VII-1).

Gene	Primers (For prokaryotic proteins)	Plasmid
<i>lpg1</i>	lpg1-Fwd1 5'-TTGGATCCCGCTCGGGCACAGAGACCT-3'	pET-28(a)-1His/2His
	lpg1-Rev1 5'-GGAAGCTTTAGCTAGGATCAACAGCAAAG-3'	pET-32(a)-1His/pMal
	lpg1-Fwd2 5'-TTGGATCCCGCTCGGGCACAGAGACCT-3'	pET-28(a)-2His
<i>lpg1L</i>	lpg1L-Fwd1 5'-TTGGATCCGACGTAACACAGCCCAC-3'	pET-28(a)-1His/2His
	lpg1L-Rev1 5'-TTGCGGCCGCTTAAGGGCTGACAGC-3'	pET-32(a)-1His
	lpg1L-Fwd2 5'-TTGGATCCGACGTAACACAGCCCAC-3'	pET-28(a)-2His
	lpg1L-Fwd1 5'-TTGGATCCGACGTAACACAGCCCAC-3'	pMal
<i>lpg1R</i>	lpg1R-Fwd1 5'-TTGGATCCGACGCGGGCGGTGGGAG-3'	pET-28(a)-1His/2His
	lpg1R-Rev1 5'-TTAAGCTTCTACTTTTCGCCAATCCGGCTC-3'	pET-32(a)-1His/pMal
	lpg1R-Fwd2 5'-TTGGATCCGACGCGGGCGGTGGGAG-3'	pET 28(a)-2His
<i>lpg1G</i>	lpg1G-Fwd1 5'-TTGAATTCGCGCTCGGATGGACAAG-3'	pET-28(a)-1His/2His
	lpg1G-Rev1 5'-TTAAGCTTTCATTGGTAAGCAAAC-3'	pET-32(a)-1His
	lpg1G-Fwd2 5'-TTGAATTCGCGCTCGGATGGACAAG-3'	pET 28(a)-2His

**Table VII- 1:** List of primers used to clone *lpg1x* genes into pET and pMal vectors by classical cloning (Enzyme restriction sites)

However, to get eukaryotic proteins expressed by *L. tarentolae*, both classical and In-fusion cloning strategies have been performed. PCR products were annealed to linearized pLEXY-Hyg2 plasmid based on recombinational cloning strategy using cloning In-Fusion® HD Cloning Kit User (Clontech). In this case, the 5' end of the primers contain nucleotides that are homologous to the end of the linearized DNA plasmid fragment. And, the 3' end contain specific sequence of the target gene. Also, according to the chosen cloning site, target DNA can be inserted into pLEXY-Hyg2 vector in a way that proteins could be expressed either cytosolically or secreted in the medium.

For LPG1R, first PCR was made with primers that amplify the gene region plus 200 bases from either side. This PCR product was used as template to amplify only the gene target with the primers, detailed below (Table VII-2), in order to amplify the genes.

Gene	Primers ( For eukaryotic proteins)		Strategy of cloning
<i>lpg1</i>	<i>lpg1</i> -S-Fwd	5'-CTGGCGCCTCTCTAGAU <b>AATCGCTCGGGCACAGAGAC</b> -3'	Secreted form
	<i>lpg1</i> -S-C-Rev	5'-GTGGGTACCCTTAAG <b>GCTAGGATCAACAGCAAAGTC</b> -3'	
	<i>lpg1</i> -C-Fwd	5'-GATCTGCCATGGCC <b>AATCGCTCGGGCACAGAGA</b> -3'	Intracellular form
	<i>lpg1</i> -S-C-Rev	5'-GTGGGTACCCTTAAG <b>GCTAGGATCAACAGCAAAGTC</b> -3'	
<i>lpg1L</i>	<i>lpg1L</i> -S-Fwd	5'-CTGGCGCCTCTCTAGAU <b>TCCTTAAGCGACGTAACACAGCCC</b> -3'	Secreted form
	<i>lpg1L</i> -S-C-Rev	5'-GTGGGTACCCTTAAG <b>AGGGCTGACAGCCTGCAGAATAAC</b> -3'	
	<i>lpg1L</i> -C-Fwd	5'-AGATCTGCCATGGCC <b>TCCTTAAGCGACGTAACACAGCCC</b> -3'	Intracellular form
	<i>lpg1L</i> -S-C-Rev	5'-GTGGGTACCCTTAAG <b>AGGGCTGACAGCCTGCAGAATAAC</b> -3'	
<i>lpg1R</i>	<i>lpg1R</i> -Fwd -200	5'-TCTCCGAGTTTGTGTGTGCGC-3'	PCR1 (+200 bases from either side)
	<i>lpg1R</i> -Rev +200	5'-GCTTCTGCGCGTGCCTTC-3'	
	<i>lpg1R</i> -S-Fwd	5'-GGCGCCTCTCTAG <b>ACGGCGGGTGGGAG</b> -3'	Secreted form
	<i>lpg1R</i> -S-C-Rev	5'-GTGGGTACCCTTAAG <b>CTTTCGCCAATCCGGCTCTGTT</b> -3'	
	<i>lpg1R</i> -C-Fwd	5'-CAGATCTGCCA <b>TGCCTGACGCGGGTG</b> -3'	Intracellular form
	<i>lpg1R</i> -S-C-Rev	5'-GTGGGTACCCTTAAG <b>CTTTCGCCAATCCGGCTCTGTT</b> -3'	
<i>lpg1G</i>	<i>lpg1G</i> -S-Fwd	5'-CTGGCGCCTCTCTAG <b>CCAGCTCCCATCAATCGG</b> -3'	Secreted form
	<i>lpg1G</i> -S-Rev	5'-GGTGGGTACCCTTAAA <b>TTGGTAAGCAAACGCTTCTCC</b> -3'	
	<i>Lpg1G</i> -C-Fwd	5'-TTA <b>CCATGG</b> CGCTCGGATG-3'	Intracellular form
	<i>lpg1G</i> -S-C-Rev	5'-TTT <b>TCTTAAG</b> TTGGTAAGCAAACGCTTCTCC-3'	

**Table VII-2:** List of primers used to clone *lpg1x* genes into pLexsy-Hyg2 vector by classical cloning (Enzymatic restriction sites inserted for classical cloning) and In-fusion cloning strategies (Fragment specific to the target gene for In-fusion cloning).

Two PCR programs were used according to primers:

A- with classical primers for classical cloning (Table VII-3)

- 1- 98°C: 5min
- 2- 98°C:10sec
- 3- Ta: Annealing Time (Detailed for each couple of primers below)
- 4- 72°C: 1.30min
- 5- 72°C: 10min

Step 2 to 4 were repeated for 35 cycles.

B- With In-fusion Primers for In-fusion cloning (Table VII-4)

- 1- 98°C: 5min
- 2- 98°C:10sec
- 3- Ta: Annealing Time (Detailed for each couple of primers below)
- 4- 72°C: 2min
- 5- 98°C:10sec
- 6- 72°C: 2min
- 7- 72°C: 10min

Step 2 to 4 were repeated for 6 cycles, then step 5 to 6 were repeated for 30 cycles.

Gene	Primers	DNA (ng)	"Ta" (°C)	Annealing time (scd)
<i>lpg1</i>	lpg1-Fwd1 lpg1-Rev1	0.3	70	10
	lpg1-Fwd1 lpg1-Rev2	0.3	70	10
<i>lpg1L</i>	lpg1L-Fwd1 lpg1L-Rev1	0.2	68	30
	lpg1L-Fwd1 lpg1L-Rev2	0.2	68	30
	lpg1L-Fwd1 lpg1L-Rev3	1.7	67	30
<i>lpg1R</i>	lpg1R-Fwd1 lpg1R-Rev1	0.3	68	10
	lpg1R-Fwd1 lpg1R-Rev2	0.2	68	30
<i>lpg1G</i>	lpg1G-Fwd1 lpg1G-Rev1	0.3	70	30
	lpg1G-Fwd1 lpg1G-Rev2	0.2	65	30

**Table VII-3:** PCR conditions to amplify genes the prokaryotic strategy cloning

Gene	Primers	DNA (ng)	"Ta" (°C)	Annealing time (scd)
<i>lpg1</i>	lpg1-S-Fwd lpg1-S-C-Rev	0.37	67	30
	lpg1-C-Fwd lpg1-S-C-Rev	0.37	68.7	30
<i>lpg1L</i>	lpg1L-S-Fwd lpg1-S-C-Rev	0.35	70	30
	lpg1L-C-Fwd lpg1L-S-C-Rev	0.35	70.6	30
<i>lpg1R</i>	lpg1R-Fwd -200 lpg1R-Rev +200	0.4	71	30
	lpg1R-S-Fwd1 lpg1R-S-C-Rev1	0.24	72	30
	lpg1R-C-Fwd1 lpg1R-S-C-Rev1	0.24	72	30
<i>lpg1G</i>	lpg1G-S-Fwd3 lpg1G-S-Rev3	2.4	63.6	30
	Lpg1G-C-Fwd 2 lpg1G-S-C-Rev2	2.4	72	30

**Table VII-4:** PCR conditions to amplify genes the Eukaryotic strategy cloning

### B. Vectors digestion and transformation in DH5 $\alpha$ cells

Commercial pET, pMAL and pLEXY-hyg2 vectors were first transformed into *E. coli* DH5 $\alpha$  using heat shock procedure. Only one clone was incubated on 5ml LB Broth media overnight at 37°C with the appropriate antibiotics as described below with the concentration of 34 $\mu$ g/ml for kanamycin and 50 $\mu$ g/ml for ampicillin. Then, vectors were extracted and purified using GeneJET™ Plasmid Miniprep Kit (*Thermo scientific™*). After DNA quantification by  $\mu$ drop (*Thermo scientific™*), DNA plasmids were digested with the appropriate restrictive enzymes (*Thermo scientific™*) as described below (Table VII-5) according to their respective protocols.

Gene Strategy of cloning	Vectors	Antibiotic	Restrictive enzyme (Forward)	Restrictive enzyme (Reverse)
<i>lpg1</i>	pET-28(a)	Kanamycin	BamHI	HindIII
	pET-32(a)	Ampicillin	BamHI	HindIII
	pMAL	Ampicillin	BamHI	HindIII
<i>lpg1L</i>	pET-28(a)	Kanamycin	BamHI	NotI
	pET-32(a)	Ampicillin	BamHI	NotI
	pMAL	Ampicillin	BamHI	XbaI
<i>lpg1R</i>	pET-28(a)	Kanamycin	BamHI	HindIII
	pET-32(a)	Ampicillin	BamHI	HindIII
	pMAL	Ampicillin	BamHI	HindIII
<i>lpg1G</i>	pET-28(a)	Kanamycin	BamHI	HindIII
	pET-32(a)	Ampicillin	BamHI	HindIII
	pMAL	Ampicillin	BamHI	HindIII
Secreted form	pLEXSYS-Hyg2	Ampicillin	XbaI	AflII
cytosolic form	pLEXSYS-Hyg2	Ampicillin	NcoI	AflII

**Table VII-5:** List of restriction enzyme used for cloning.

### C. Plasmid DNA purification, analysis and sequencing

Transformed DH5 $\alpha$  colonies were selected and analyzed by colony PCR using commercial primers (*T7/ T7term* for pET vectors, *MalE /pTrc His Rev* for pMAL-C2X vectors, and A264/p1442 for pLEXSYS-Hyg 2) in addition to the designed primers for cloning. PCR results were analyzed through agarose electrophoresis. Potential positive clones were grown in LB Broth medium with the resistance appropriate antibiotic. Then, plasmids were extracted, purified and digested by specific restrictive enzymes. Finally, positive clones were sent for sequencing performed by Eurofins Genomics (<https://www.eurofinsgenomics.eu>). Results were finally analyzed with CLC sequence viewer software (<https://www.qiagenbioinformatics.com/products/clc-sequence-viewer/>).

## II. Expression and purification of proteins

### A. Prokaryotic proteins

#### 1. Transformation in and overexpression of proteins

On day 1, pure DNA plasmid (50-100ng) were mixed with 50 $\mu$ l of chemically competent *E. coli* cells. The mixture was incubated on ice for 20min. Then, heat shock was performed by placing the cells on 42 $^{\circ}$ C water bath for exactly 45 seconds followed by 2min incubation on ice. After that,

450µl of LB Broth media was added and cells were incubated at 37°C for one hour and 140 rpm. After 1min centrifugation at 1000 rpm, 350 µl of supernatant was removed and pellet was resuspended and plated on LB agar plate with the appropriate antibiotics as described below (Table VII-6). Plates were incubated at 37°C overnight. the following day, a single colony was selected and cultivated in 5 ml LB Broth medium containing antibiotics at 37°C, 140 rpm overnight. On the third day, 60ml of LB Broth media was inoculated with 600 µL of grown bacteria and incubated overnight at 37°C and 140 rpm. On day 4, 1liter of LB Broth media was inoculated with 10 ml of grown bacteria and incubated in the morning at 37°C / 250 rpm until optical density at 600 nm reaches 0.6. Then, induction was started by the addition of 200 µM final concentration of IPTG at the desired temperature (20, 25, 30 or 37°C) for the wanted period of 2, 4, 7 or 20 hours of induction.

	<i>BL21 (DE3)</i>	<i>C43 (DE3)</i>	<i>Rosetta (DE3)</i>
<i>pET-28a(+)</i> Vectors	Kanamycin (30µg/ml)		Kanamycin (34 µg/ml) /Chloramphenicol (30µg/ml)
<i>pET-32a(+)</i> Vectors	Ampicillin (50µg/ml)		Ampicillin (50 µg/ml) /Chloramphenicol (30µg/ml)
<i>pMAL</i> Vectors			

**Table VII-6:** List of antibiotics used with each expression vector and bacteria strain.

## 2. Lysis of cells

After the induction step, cells were first harvested by centrifugation, resuspended in 50 ml of lysis buffer.

1. Lysis buffer 1: Tris HCl 50 mM pH8, NaCl 200 mM, 0.1mg/mL lysozyme and 1mM PMSF containing detergent (Chaps 0.1%, Tween 0.1% or Triton 0.1%).
2. Lysis Buffer 2: Tris HCl 50 mM pH8, NaCl 200 mM, 1mg/mL lysozyme, 1mM PMSF, 5% glycerol and 1mM DTT.

After 20 min of incubation at 4°C under stirring in lysis buffer, cells have undergone 3 heat/shock step followed by sonication step as follows: 50% life cycle, 30sec on /30sec off, this was repeated 6 times. After that, mixtures were centrifuged at 32000xg for 20min. Supernatant were filtrated with 0.45µm filter.

Supernatants were filtrated with 0.45µm filter.

### 3. Protein purification

Recombinant proteins tagged with His-tag or MBP-tag were purified manually or on AKTA FPLC Purifier (*GE Healthcare*) using 1ml HiTrap™ (*Thermo Scientific™*) or 1ml MBPTrap HP™ (*GE Healthcare*) columns. First, HiTrap™ and MBPTrap HP™ columns were equilibrated respectively with buffer his-tag A (Tris HCl 50 mM pH8, NaCl 200 mM, Imidazole 10 mM) and MBP-tag A (Tris HCl 20 mM pH8, NaCl 200 mM, EDTA 1 mM pH7, DTT 1 mM). Second, the washing step was done with 10 volume column of buffer his-tag A or MBP-tag A. Finally, proteins were eluted with imidazole gradient (from 10 to 500mM) or Maltose gradient (1 to 10 mM) using buffer his-tag B (Tris HCl 50 mM pH8, NaCl 200 mM, Imidazole 500 mM) and MBP-tag B (Tris HCl 20 mM, NaCl 200 mM, EDTA 1 mM pH7, DTT 1 mM, maltose 10 mM). Size Exclusion chromatography was performed using Buffer C (Tris HCl 50 mM pH8, NaCl 200 mM) as eluent. Purity of proteins were assessed by SDS PAGE followed by blue Coomassie staining.

## B. Eukaryotic proteins

### 1. Culture of *Leishmania*

M199-5X medium ingredient were first prepared separately as described below:

- M199 – 5X, final volume 500 mL
  - 40.15 g of Gibco medium M199 + Hanks + L-glutamine without NaHCO<sub>3</sub>
  - 0.875 g of NaCHO<sub>3</sub>
  - H<sub>2</sub>O
  - pH settled at 7.2
  - Filtrated with 0.22 μm filter, autoclaved and stored at 4°C
- Hepes free acid 1M pH 7.5, sterilized with 0.22μm filter, autoclaved and stored at 4°C
- Adenin solution 10 mM, final volume 50 ml
  - 0.86 g of Adenin minimum 99% (*Sigma Aldrich*®)
  - 2.5 mL of Hepes 1M pH 7.5
  - H<sub>2</sub>O
  - Filtrated with 0.22 μm filter and stored at 4°C
- Hemin 0.25%, final volume 5 ml
  - 12 mg of Hemin
  - 2.5 mL of Triethanolamin (*Sigma Aldrich*®)
  - H<sub>2</sub>O
  - Filtrated with 0.22μm filter and stored at 4°C
- Biotin 0.1%, final volume 5 mL
  - 5 mg of Biotin (*Sigma Aldrich*®)

- 4.75 ml of ethanol 100%
- H<sub>2</sub>O
- Filtrated with 0.22 µm filter and stored at -20°C

M199-1X medium can be then prepared and ready to use as described below (final volume 1L):

- 336 mL of filtered with 0.22µm filter and autoclaved ionized water
- 100 mL of M199- 5X medium
- 7.5 mL of 1M Hepes pH7.5
- 5 mL of 10Mm Adenin
- 1 mL of 0.25% Hemin
- 0.5 mL of 0.1% Biotin
- 50 mL of decomplexed SVF
- 2.5 mL of Penicillin (10000 U/ml) /Streptomycin (10 mg/ml) (*Sigma Aldrich™*)

*Leishmania* cells were cultured in ventilated tissue culture (TC) flasks at 26°C in the dark and aerated condition (NO CO<sub>2</sub> incubator is required). The flasks were positioned upright with successive dilution in order to have 1.4 – 2 OD<sub>600 nm</sub> which corresponds to 6 – 8 10<sup>7</sup> cells/mL.

## 2. *Leishmania* transfection

### a) DNA preparation

Digestion of pLEXSY-Hyg2 expression plasmid containing the gene of interest was done with restrictive enzyme SwiI. Two fragments were generated. The first and unwanted one is a 2.9kbp fragment representing the *E. coli* part; The second and larger fragment represents the linearized expression cassette with the wanted gene. This part cassette was isolated by *Gel Extraction Kit* (*Thermo Scientific™*).

### b) Transfection and selection

After several passages of the inoculation culture from a glycerol stock,

Day1: *Leishmania* pre-culture was prepared and diluted 1:20 in 10mL M199 medium supplemented with PenS/strep. Then it was incubated in upright ventilated TC flask at 26°C in the dark.

Day 3: The pre-culture was diluted 1:10 into 10 mL in the same medium and incubated in flat ventilated TC flask overnight at 26°C in the dark.

Day 4: Density of cells reached 1.4 at 600 nm representing 6.10<sup>7</sup> cells/mL, the cells were vital and had droplike shape on microscope. Cells were centrifuged at 2000g for 3 min, then ½ volume of supernatant was removed. The pellet was resuspended in remaining medium to get 10<sup>8</sup> cells/mL and put on wet ice for 10 min.

Tubes with 1-10 $\mu$ g transforming DNA in maximum 50 $\mu$ l of water and electroporation cuvette (2nm) were placed on wet ice in parallel. Then, 350  $\mu$ L pre-chilled cells were added to the tube with DNA and transferred to the electroporation cuvette on wet ice. Electroporation step was performed at 450V, 450 $\mu$ F and monitor pulse time (5-6 msec). After that, electroporation cuvettes were put back on ice for exactly 10min. Electroporated cells were transferred into 10mL of medium in ventilated TC flask. Finally, flasks were incubated as static suspension culture overnight at 26°C.

Day 5: Selection starts with the addition of hygromycin (50  $\mu$ g/mL) in the medium.

### c) Purification of eukaryotic recombinant proteins

100mL of recombinant selected clones was prepared and incubated for 2 days at 26°C in the dark. Cells were centrifuged and supernatant was filtrated with 0.24  $\mu$ m filter and purified on 1ml HiTrap™ (*Thermo Scientific*™) using the same purification conditions described before.

## III. Analysis of recombinant proteins

### A. Bradford protein assays

BSA stock solution (1mg/ml) was diluted to prepare standard calibrators in concentration ranging from 0 to 0.5mg/mL. Purified proteins were also diluted 2, 10 and 20 times. 10  $\mu$ L of each sample, calibrators and proteins of interest were deposited in 96-well microplate in triplicate. Then 200  $\mu$ L of Bradford (Bio-Rad) reagent diluted 5 times were added. After 5 min of incubation at room temperature, absorbances were measured at 595 nm. Protein samples concentrations were calculated according to an experimental calibration curve.

### B. Tryptic digestion of proteins

SDS-PAGE was destained and protein bands of interest were cut using scalpel. Proteins were digested by Trypsin enzymes. Then, digested fragments were extracted from the gel. This was made according to the commercial In-Gel Tryptic Digestion Kit (*Thermo Scientific*™). Cleaved peptides were sent to mass spectrometry.

### C. Western Blot

#### 1. solutions

- ✓ Transfer Buffer (wet)
  - Tris- HCl 20 mM
  - Glycine 150 mM
  - Methanol 20%
- ✓ TBS
  - Tris-HCl 20 mM pH 7.5

- NaCl 140 mM
- ✓ TBS +TWEEN 0.1%
  - 500mL TBS
  - 500 $\mu$ L Tween
- ✓ Blocking buffer (3% BSA)
  - 10mL TBS
  - 300mg BSA
- ✓ Antibody dilution
  - 10mL TBS + Tween
  - 10 $\mu$ L Antibody

## 2. Transfer

The PVDF membrane was activated with methanol for 1 min and rinsed with transfer buffer before preparing the stack. The gel and two filter papers were also soaked in transfer buffer for 5min. The “sandwich” was built first with filter paper, then the membrane, followed by the gel and then the second filter paper. Air bubbles trapped between the gel and the membrane were removed by a glass rod. The “sandwich” was placed on the transfer device, the membrane was properly oriented to the anode side. A current of 3mA/cm<sup>2</sup> was applied for 1 hour.

## 3. Immunodetection

The membrane was placed in petri dish with 10mL of blocking buffer for 2 hours at 4°C and gentle agitation. Then, it was placed in a new petri dish with 10mL of Antibody solution, Anti-His (C-term)/AP Ab Novex® (*Invitrogen*) and was incubated at 4°C for overnight under gentle agitation.

## 4. Detection

The membrane was washed 3 times 5min with 10 mL of TBS-Tween under gentle agitation at 4°C. Then it was drained on a paper towel and placed in new Petri dish face protein above. Finally, it was completely covered with 500  $\mu$ L of substrate solution Novex® AP-Chromogen (*Invitrogen*). Colored bands appear in few minutes.

# IV. Enzyme activity assays and kinetics parameters determination

## A. Malachite green assay

Three solutions were freshly prepared on the day of reactions:

Solution A containing 1.1M sulfuric acid and 42mM Molybdate.

Solution B containing 0.45mM malachite Green and 0.1 % Poly-Vinyl-Alcohol.

Solution C containing 7.8% of sulfuric acid.

First, different concentration of phosphate solution was prepared; 0; 3.125; 6.25; 12.5; 25; 50 and 100  $\mu\text{M}$ . In parallel, the same concentration of UDP solution were also prepared. Then 20  $\mu\text{l}$  of each dilution of phosphate solution were deposit in 1.5 mL microtube with 20 $\mu\text{l}$  of Tris 1M pH8, in 200  $\mu\text{l}$  final volume. With the same way 20  $\mu\text{l}$  of each UDP solution were added to 20  $\mu\text{l}$  of Tris 1M pH8, and 2U of FastAP Thermosensitive Alkaline phosphatase (*Thermo Scientific™*). After 1 hour and 2 hours incubation at 37°C, 100 $\mu\text{l}$  of each condition were deposit in 1.5 mL microtube with 70  $\mu\text{l}$  of solution A and 30  $\mu\text{l}$  of solution B. Reaction mixture were mixed and left at room temperature for 2 min. After that, 200 $\mu\text{l}$  of solution C were added. The final reaction mixture was left for 20 min at room temperature followed by an absorbance lecture at 625 nm. This allowed the plotting of an experimental calibration curves of released phosphate by the phosphatase alkaline. This was compared to the calibration curves of phosphate solutions.

### B. HPLC detection

Chromatographic analysis was performed on the Agilent 1220 Infinity LC (Agilent®). HPLC separation was carried out with an Agilent C18 column (150 X 6,4mm; 3.5 $\mu\text{m}$ ). Separation was done at a flow rate of 0.8ml/min. The mobile phases employed for the separation of PNP sugars was composed of water + 0.1% TFA (solvent A) and acetonitrile + 0.1% TFA (solvent B). Gradient elution protocol was used: 0-4min: (10%B), 4-20min: (10-80%B), 20-24: (80%B), 24-25 (80 – 10%B). The injection volume was 20 $\mu\text{L}$ . UV absorbance was measured at 360nm. The acquisition and data analysis were performed with Agilent Open Lab software.

### C. Spectrophotometric assays

All reactions were performed in 200 $\mu\text{l}$  final volume buffer containing 50mM MOPS pH8, 20mM MgCl<sub>2</sub>, 50mM KCl, 3.5mM PEP (*Sigma Aldrich®*), 7.5 U pyruvate kinase (*Sigma Aldrich®*), 16.8 U lactate dehydrogenase (*Sigma Aldrich®*) and 1.1mM NADH (*Sigma Aldrich®*). All the solutions were freshly prepared on the day of use. Donors were used at different concentration from 10 $\mu\text{M}$  to 10mM, and acceptors were fixed at 1mM final concentration. After incubation of the reaction mixture for 30 second at 37°C, 0.2 – 10  $\mu\text{g}$  enzymes were added. Many negative controls were prepared, without enzymes, without NADH, without acceptors, without donors and without PK-LDH. The reactions were performed in 96-well microplate. They were monitored at 340 nm using a Multiskan™ GO (Thermo Scientific) microplate reader for up to 20 min with 10 s intervals. UDP formation rates were assumed to be equal to NADH consumption rates, and kinetic parameters were calculated by fitting saturation curves (obtained from the average of triplicate measurements) with standard the Michaelis–Menten equation using Prism 6 (GraphPad).

#### D. Galactofuranosyltransferase preparative reactions

A magnetically stirred 3 mL solution containing methyl  $\alpha$ -D-mannopyranoside (30 mM, 6 mg), the UDP-sugar (10 mM, 2 mg), 50 mM Tris pH 8, 20 mM MgCl<sub>2</sub> and 2 mg of the GalT to be tested was prepared. The reaction was incubated at 37 °C for 24 h.

#### E. Separation and TLC analysis

Mixture reaction were purified on Sephadex G-10 Gel (*Sigma Aldrich*<sup>®</sup>). Then, analytical thin layer chromatography of obtained fractions was performed on silica gel aluminum supported plates. The eluent was composed of ethyl acetate, methanol and water with the ratio of 7/2/1. Sugars were detected with orcinol solution (95% ethanol 100%, 5% H<sub>2</sub>SO<sub>4</sub> and orcinol 200 mg) after heating at 100 °C.

#### F. Acetylation of sugars and analysis

After reaction solvent evaporation, 1 mL of acetic anhydride and 1 mL of pyridine were added and the reaction was left at room temperature for 48 h. Then, the residue was concentrated through co-evaporation with toluene. The reaction mixture was then resuspended and the peracetylated sugar was isolated by extraction into CH<sub>2</sub>Cl<sub>2</sub>. Finally, high-resolution accurate mass measurements were performed in positive mode with an ESI source on a Q-TOF mass spectrometer (Bruker MaXis) with an accuracy tolerance of 2 ppm by the "Fédération de Recherche" ICOA/CBM (FR2708) analytical platform.

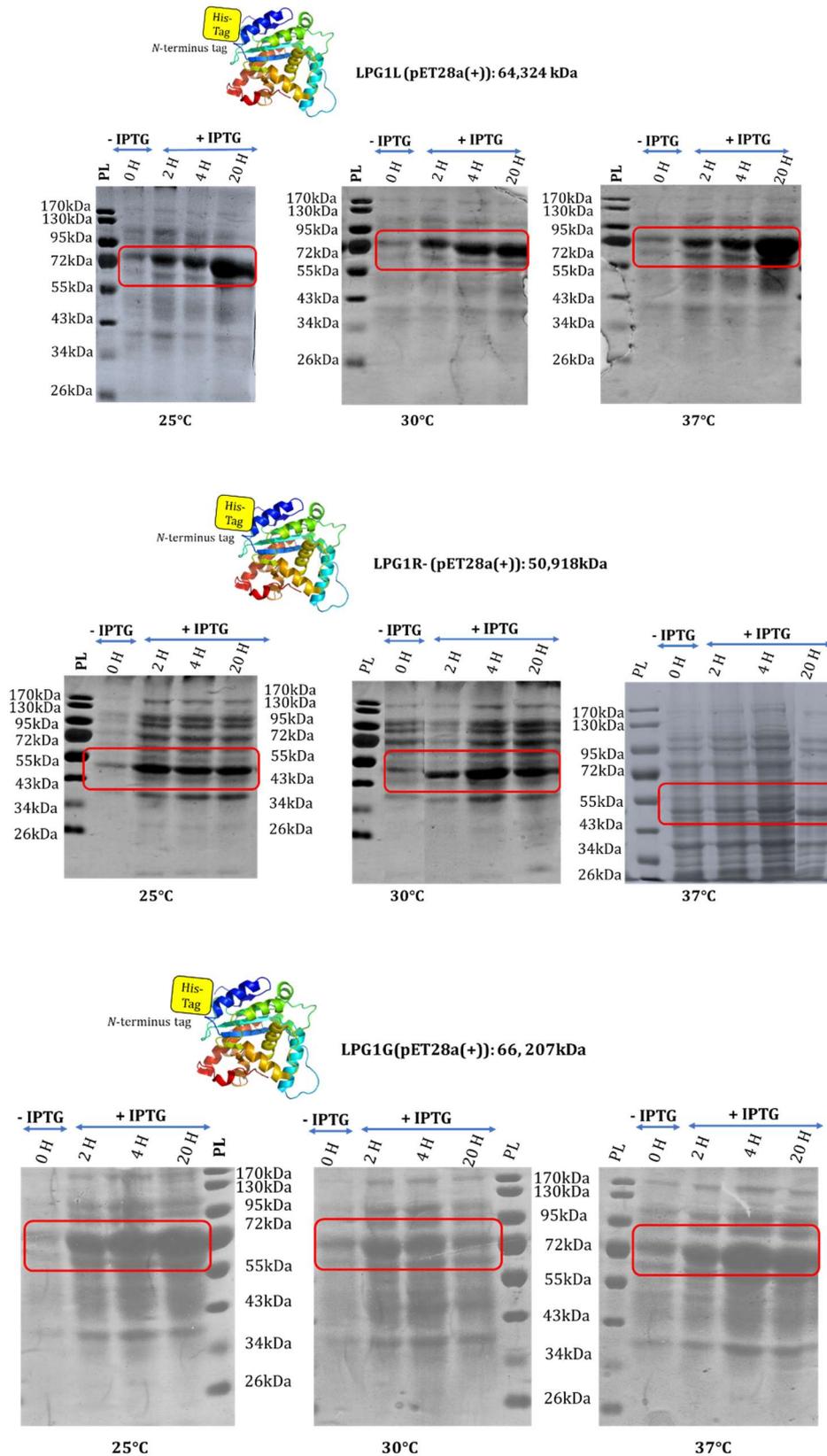
### V. Cristallogenesis of proteins

After purification steps, proteins were concentrated using protein concentrator 30,000 MWCO or 50,000 MWCO (*Thermo Scientific*<sup>™</sup>) according to the molecular weight of proteins. Final concentration varies from 1 to 20mg/ml. JBScreen Basic HTS HTS-1 and HTS-2 kits and JBScreen Classic kits HTS-1 and HTS-2 (*Jena Bioscience*) containing 96 reagent mixtures for screening a wide range of pH and various salts and precipitants were tested. Protein crystals were grown at 4°C by sitting-drop vapor diffusion in 2  $\mu$ l drop containing equal amount of pure proteins and screening buffer.

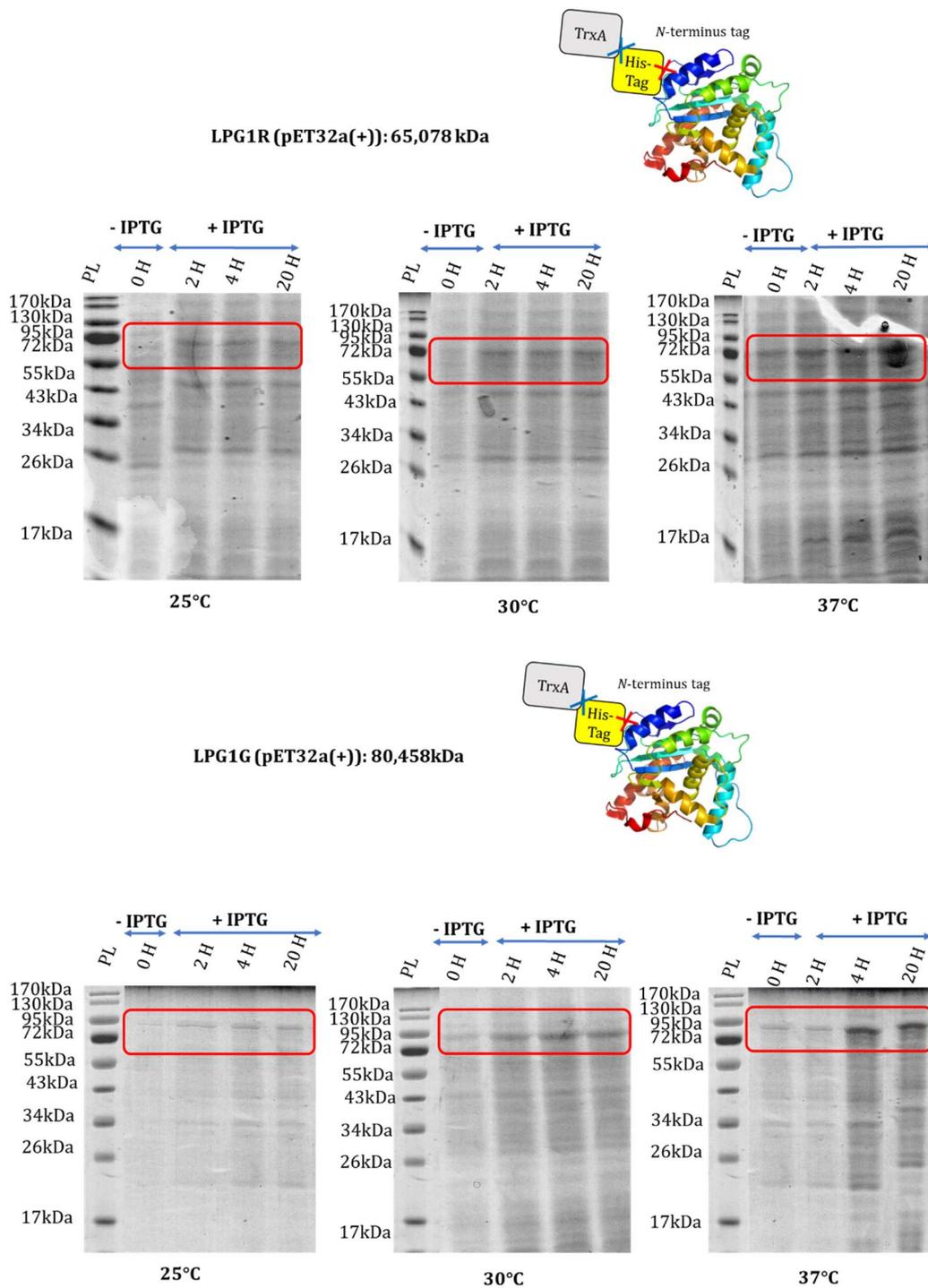
For LPG1(pMal), potential protein crystals were obtained with 30%PEG 8000 and 200mM Ammonium Sulfate. So new plate with three different concentrations 16mg/mL, 13.8mg/mL and 5mg/mL of proteins was prepared with 22 to 42% of PEG 8000 and 160 to 220 mM of Ammonium Sulfate.



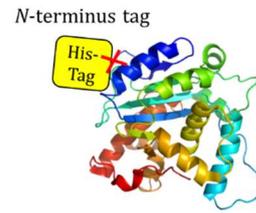
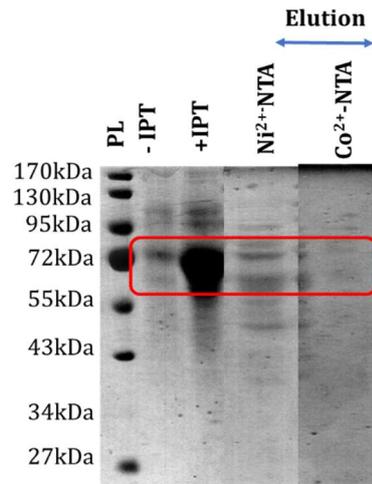
# *Appendix*



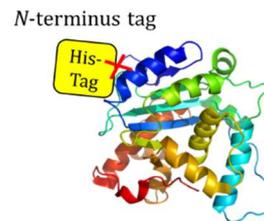
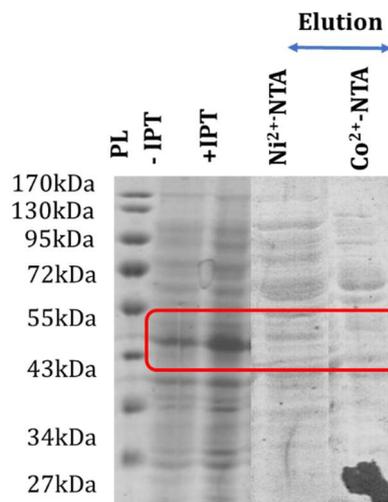
**Appendix 1:** SDS-PAGE (12%) analysis of the overexpression of LPG1L/LPG1R/LPG1G (pET-28a(+)), under different conditions of the induction; 2,4 and 20 hours of induction after the addition of IPTG in the medium at 25, 30 and 37°C. (PL: protein ladder).



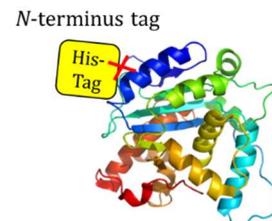
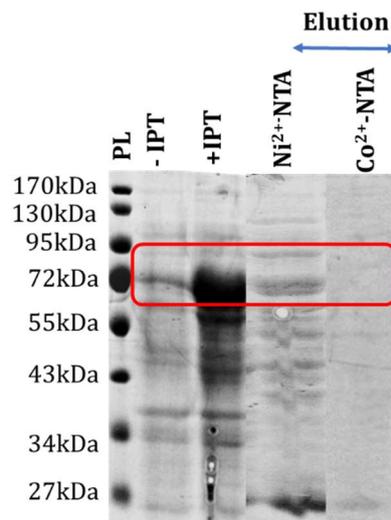
**Appendix 2:** SDS-PAGE (12%) analysis of the overexpression of LPG1L-1H (pET-32a(+)) and LPG1G (pET-32a(+)), under different conditions of the induction; 2,4 and 20 hours of induction after the addition of IPTG in the medium at 25, 30 and 37°C. (PL: protein ladder).



LPG1L-1H (pET28a(+)): 64,324 kDa

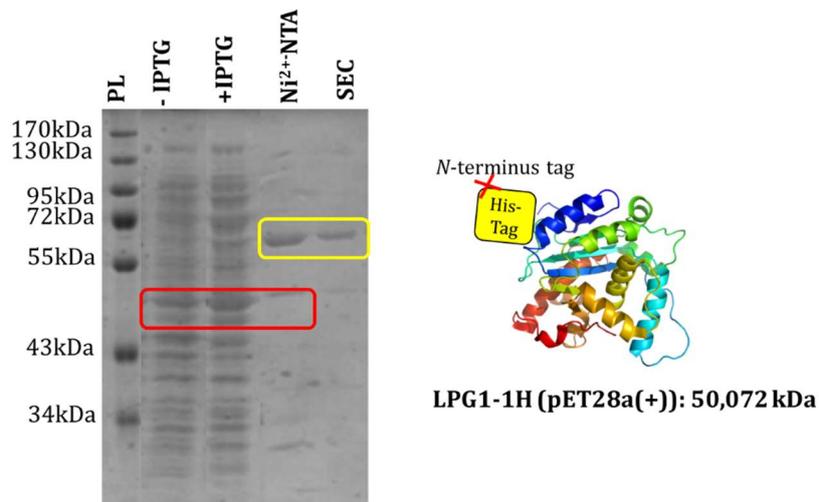
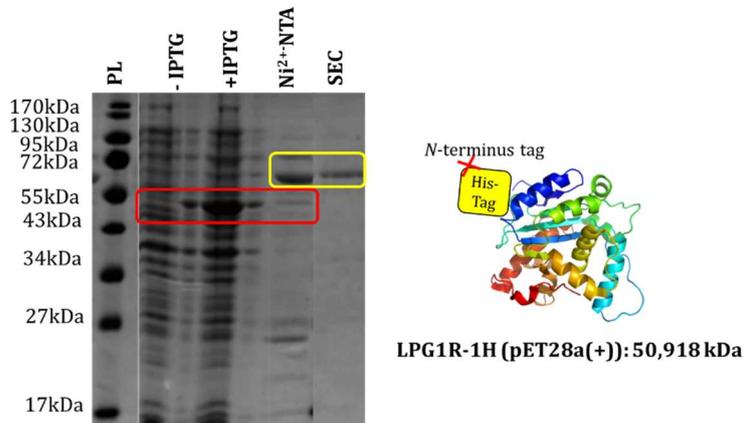
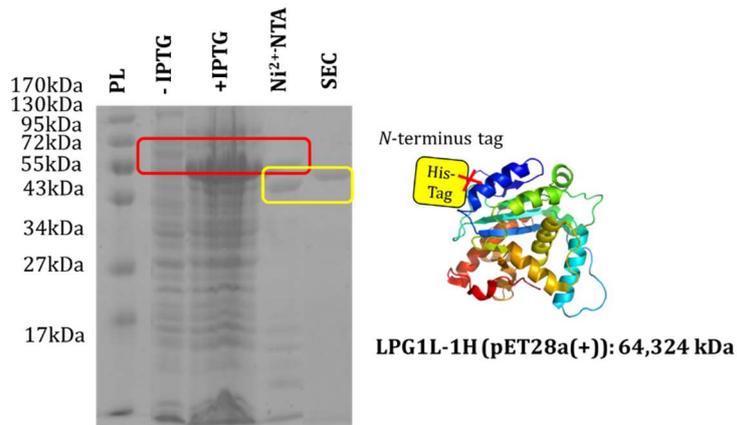


LPG1R (pET28a(+)): 50,918 kDa

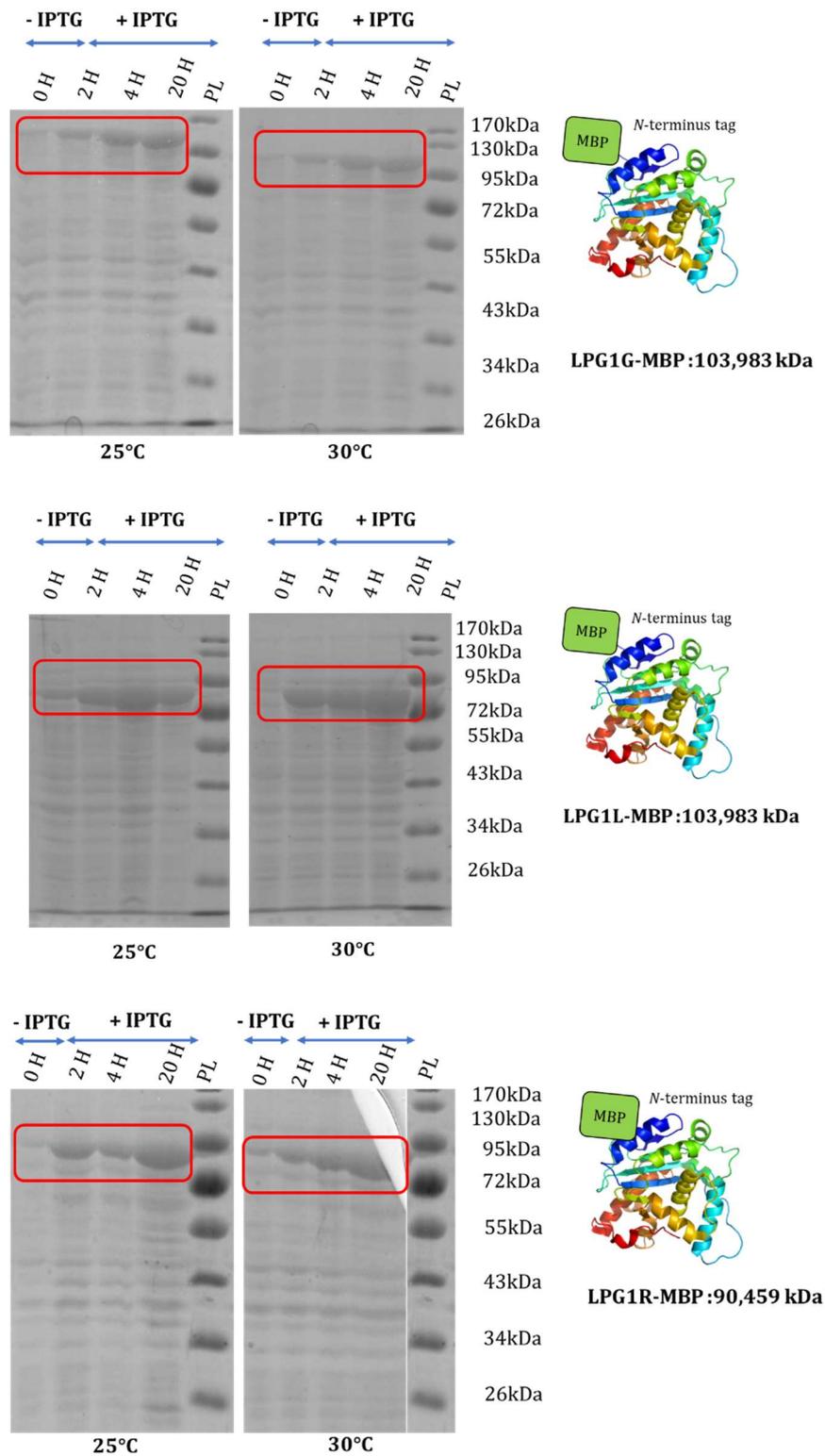


LPG1G (pET28a(+)): 66,207 kDa

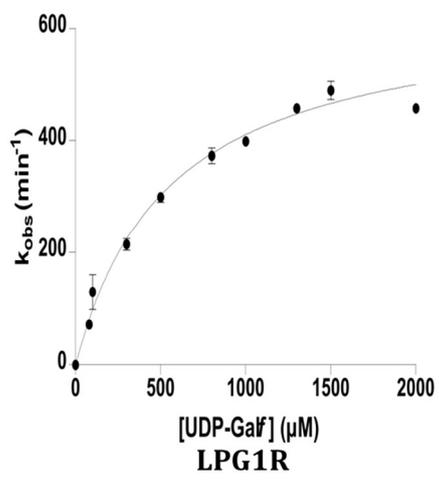
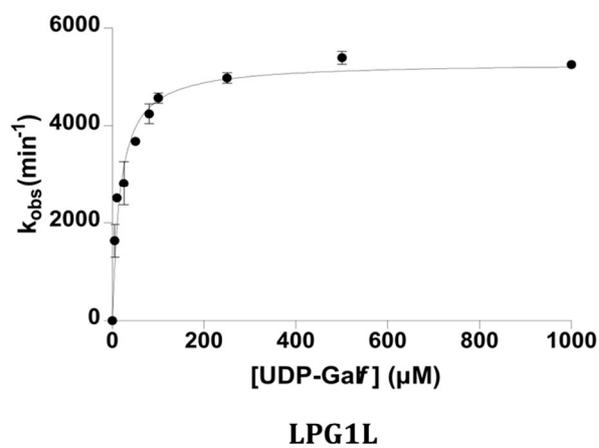
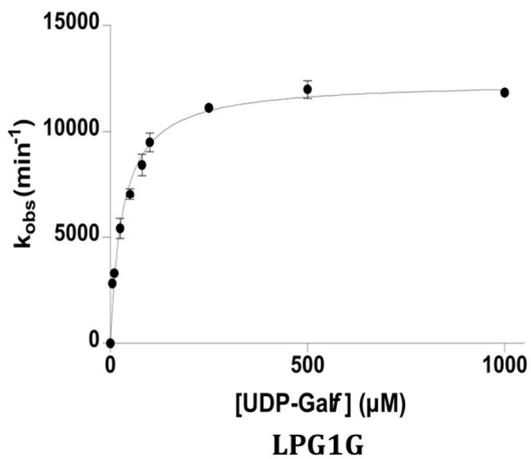
**Appendix 3:** SDS-PAGE (12%) analysis of the overexpression of LPG1L/LPG1R/LPG1G (pET-28a(+)), under different conditions of the induction; 2,4 and 20 hours of induction after the addition of IPTG in the medium at 25, 30 and 37°C. (PL: protein ladder).



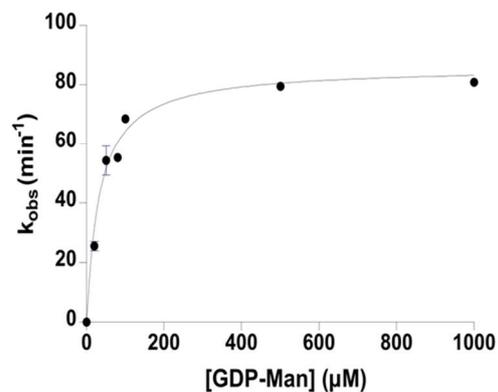
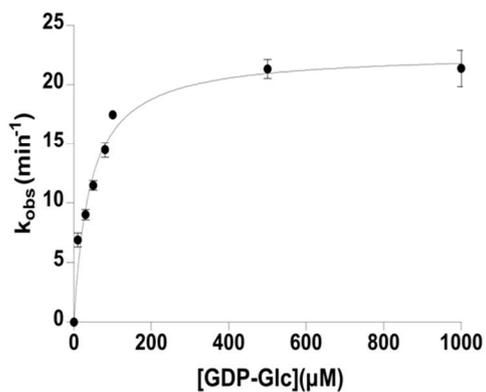
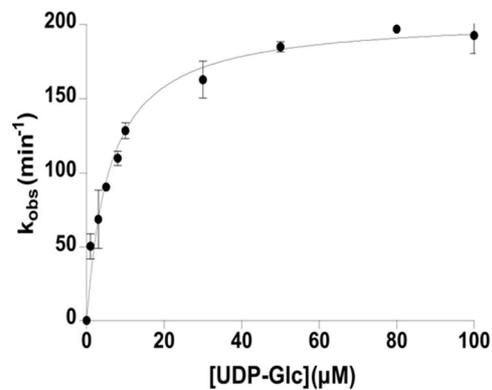
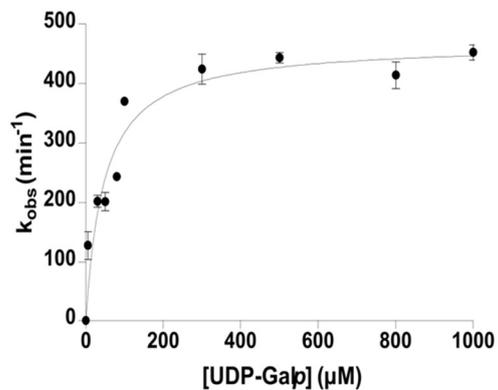
**Appendix 4:** SDS-PAGE (12%) analysis of overexpression of LPG1L, LPG1R and LPG1 (pET-28a(+)) proteins after the induction at 25°C for 2 hours, and its purification after lysis strategy 2 on Ni-NTA column followed by size exclusion chromatography (SEC). PL: Protein Ladder.



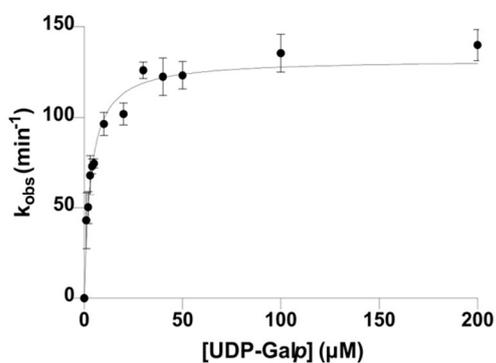
**Appendix 5:** SDS-PAGE (8%) analysis of the overexpression of LPG1L-MBP, LPG1G-MBP and LPG1R-MBP under different conditions of the induction; 2,4 and 20 hours of induction after the addition of IPTG in the medium at 25 and 30°C. (PL: protein ladder).



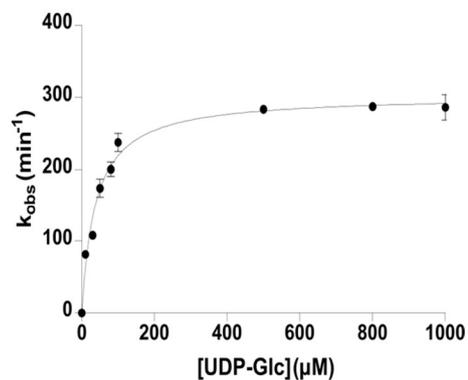
**Appendix 6:** Michaelis-Menten plot when using UDP-Galf as donor substrate for LPG1R, LPG1L and LPG1G.



**LPG1L**

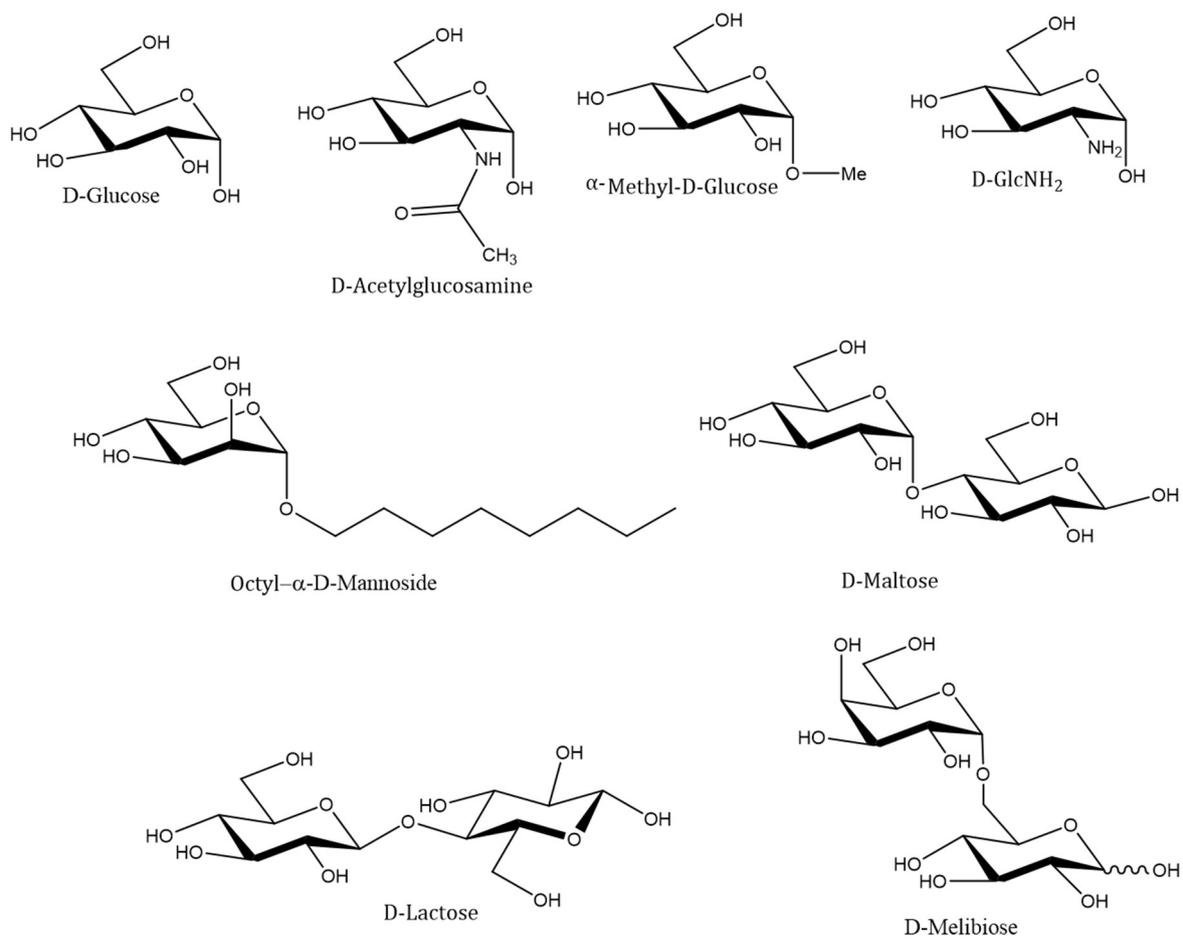


**LPG1G**

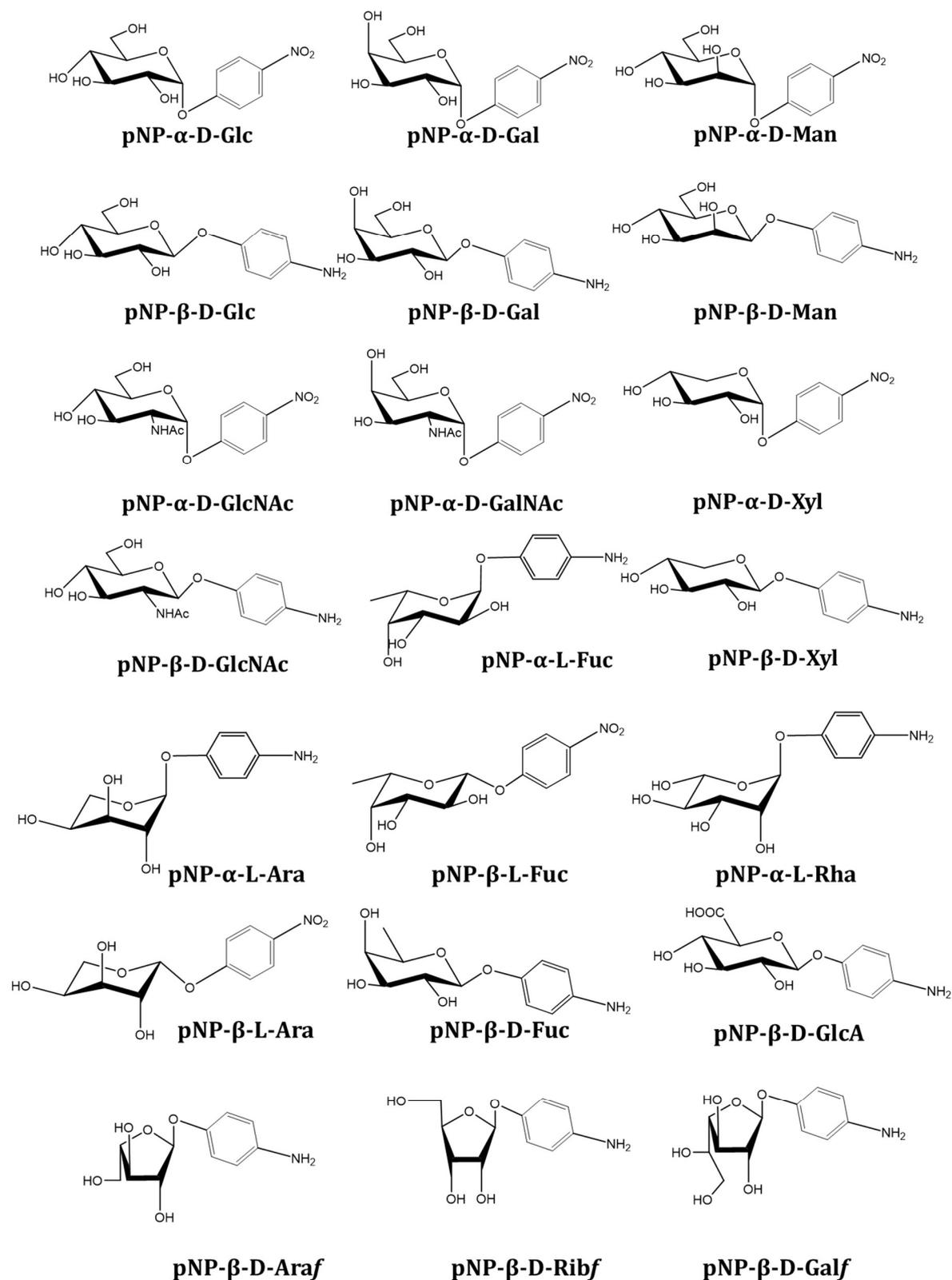


**LPG1R**

**Appendix 7:** Michaelis-Menten plot when using NDP-sugars as donor substrate for LPG1L, LPG1R and LPG1G.



**Appendix 8:** List of acceptors tested with coupled- spectrophotometric assay



Appendix 9: List of pNP-sugars tested as acceptors with LPG1x.



## ***The last chapter of thesis***

I would like to conclude my manuscript with this last chapter to tell you what these doctoral studies have brought me. Indeed, carrying out this thesis project was not only about making experiments and producing results, there was much more beyond this.

During this thesis, I was faced with lot of challenges, or should I say a lot of complications and deceptions moments. But, thanks to it I had the ability to investigate and consider alternative methods and techniques to reach my goals with never giving up. I have also acquired a novel and independent way of thinking and learned how to expand my curiosity to other fields. Indeed, I had the chance to participate to a workshop about crystallography through which I fall in love with this field and considered to pursue my postdoctoral studies in this area and gain new techniques and knowledge that will allow me to go from the genes to the protein structures.

In addition to my research, I have had the rewarding to teach practical and technical biochemistry and enzymology to 1<sup>st</sup> and 2<sup>nd</sup> year bachelor degree student during three years (64 hours/ year). My participation in such an interesting program has given me the chance to communicate with students and to share with them my love for science using different approaches. Indeed, I have learned how to interpret and popularize my explanations which is not an easy task. I also, had the opportunity to supervise different students during their research internship. This allowed to learn how to manage people, cooperate with them and trust int heir work. This experiment made me consider teaching in the future.

During this three year, I had the possibility to attend different congress and conferences to present my work at both national and international levels. This experience allowed me to put together a persuasive presentation and communicate about my project and most important thing gaining self-confidence. It allowed also to create an international professional network.

I also had the opportunity to patriciate in workshop about entrepreneurship through which I worked in group in order to create new company in only 8 hours and presented front of jury members. I have learned how to co-work with foreign persons, to be persuasive and be a decision-maker in a competitive situation. In addition to the challenging context, I enjoyed writing an innovative project and managing a team. This confirmed my desire to have my own research team in few years and leading project.

To conclude, with these three years of doctoral studies, I can say that I become a new person with new personal projects and new personal characters.

### **Leishmanian Galactofuranosyltransferases as promising versatile tools for therapeutic and chemoenzymatic approaches.**

Cells are heavily decorated by diverse glycoconjugates that are involved in important biological events such as cell-cell communication, growth of healthy or cancerous cells and pathogens infection process. Among these polysaccharidic structures, Gal $f$ -containing glycans have been the subject of increasing interest in the last decades. Indeed, the galactofuranose can be found in many pathogenic species, such as *Mycobacterium tuberculosis*, *Aspergillus* and *Leishmania*, but is absent in mammals. Therefore, these glycoconjugates are considered as interesting targets for therapeutic approaches.

Galactofuranosyltransferases (Gal $f$ Ts) catalyse the transfer of galactofuranose residues into glycoconjugates structures. However, Gal $f$ Ts are poorly described enzymes despite their crucial role in the virulence and the pathogenicity of numerous microorganisms. Up to date, only one mycobacterial Gal $f$ T has been fully characterized.

In this thesis, four putative Gal $f$ Ts of *Leishmania major*, the causing agent of leishmaniosis diseases, were characterized. They were first cloned, overexpressed in *E. coli* and purified. Then, their respective kinetic parameters were determined. In addition, since these Gal $f$ T are located in the Golgi apparatus of *Leishmania*, we assumed that their glycosylation could be an important element for their stability and activity. So, glycosylated Gal $f$ Ts were produced using, *Leishmania tarentolae*, and preliminary results of their enzymatic activity were obtained.

Still, leishmanian Gal $f$ Ts demonstrate promising results for the development of new chemoenzymatic strategies for Gal $f$ -containing glycoconjugates synthesis, as well as the design of new drugs against leishmaniasis.

Keyword : Galactofuranose, Galactofuranosyl Transferases, *Leishmania*

### **Leishmanian Galactofuranosyltransferases as promising versatile tools for therapeutic and chemoenzymatic approaches.**

Les cellules exposent à leurs surface glycoconjugués qui jouent important dans des événements biologiques importants tels que la communication entre cellules, la croissance de cellules saines ou cancéreuses et les processus d'infection d'agents pathogènes. Certaines structures polysaccharidiques qui contiennent le résidu Gal $f$  ont attiré beaucoup d'intérêt au cours des dernières décennies. En effet, le galactofuranose peut être exprimé chez de nombreuses espèces pathogènes, telles que *Mycobacterium tuberculosis*, *Aspergillus* et *Leishmania*, mais il est absent chez les mammifères. Par conséquent, ces glycoconjugués sont considérés comme des cibles intéressantes pour des approches thérapeutiques.

Les galactofuranosyltransférases (Gal $f$ T) catalysent le transfert des résidus de galactofuranose dans les structures des glycoconjugués. Cependant, ces Gal $f$ T sont des enzymes faiblement décrites, malgré leur rôle crucial dans la virulence ainsi que dans la pathogénicité de nombreux micro-organismes. Jusqu'à présent, seule la Gal $f$ T2 de *Mycobacterium tuberculosis* a été entièrement caractérisée.

Dans cette thèse, quatre Gal $f$ Ts de *Leishmania major*, l'agent responsable de la leishmaniose, ont été caractérisées. Elles ont été d'abord clonées, surexprimées dans *E. coli* et purifiées. Ensuite, leurs paramètres cinétiques respectifs ont été déterminés. De plus, puisque ces Gal $f$ T sont situées dans l'appareil de Golgi de *Leishmania*, nous avons supposé que leur glycosylation pourrait être un élément important pour leur stabilité et leur activité. Ainsi, des Gal $f$ T glycosylés ont été produites à l'aide de *Leishmania tarentolae* et les résultats préliminaires de leur activité enzymatiques ont été obtenus.

Les Gal $f$ T leishmaniennes démontrent des résultats prometteurs pour le développement de nouvelles stratégies chimio-enzymatiques pour la synthèse de glycoconjugués contenant du Gal $f$ , ainsi que pour la conception de nouveaux médicaments contre la leishmaniose.

Mots clés : Galactofuranose, Galactofuranosyl Transferases, *Leishmania*