Universitat Autònoma de Barcelona

DEVELOPMENT OF PICHIA PASTORIS FOR THE PRODUCTION OF HUMAN α-GALACTOSIDASE A AS A MODEL LYSOSOMAL ENZYME

Antonio Barreiro Vázquez

Universitat Autònoma de Barcelona

ESCOLA D'ENGINYERIA

DEPARTAMENT D'ENGINYERIA QUÍMICA, BIOLÒGICA I AMBIENTAL

DEVELOPMENT OF *PICHIA PASTORIS* FOR THE PRODUCTION OF HUMAN α-GALACTOSIDASE A AS A MODEL LYSOSOMAL ENZYME

PhD Thesis

Antonio Barreiro Vázquez

Barcelona, December 2015

DAVID RESINA RODRÍGUEZ, director executiu i científic de Bioingenium S.L., i PAU FERRER ALEGRE, professor agregat en el Departament d'Enginyeria Química, Biològica i Ambiental de la Universitat Autònoma de Barcelona,

CERTIFIQUEN:

Que el llicenciat en Biotecnologia i Bioquímica Antonio Barreiro Vázquez ha dut a terme sota la nostra direcció el treball amb el títol "**Development of *Pichia pastoris* for the production of human α-galactosidase A as a model lysosomal enzyme**" que es presenta en aquesta memòria, el qual constitueix la seva Tesi per optar al Grau de Doctor en Biotecnologia per la Universitat Autònoma de Barcelona.

I perquè es prengui coneixement i consti a efectes oportuns, presentem a l'Escola d'Enginyeria de la Universitat Autònoma de Barcelona l'esmentada Tesi, signant el present certificat a

Barcelona, 11 de desembre de 2015

Dr. David Resina Rodríguez
(Director)

Dr. Pau Ferrer Alegre
(Tutor)

Antonio Barreiro Vázquez
(Autor)

A g r a ï m e n t s

En primer lloc vull agraïr als meus directors, el David i el Pau, per donar-me la oportunitat de dur a terme aquest treball i per haver-me guiat pacientment durant aquest llarg camí. També vull fer partícep d'aquest agraïment al Ramon, que juntament amb el David, per convidar-me a formar part del projecte de Bioingenium i que han confiat tant en mí i en les meves iniciatives des del primer día.

També vull recordar als meus companys, amb els que he compartit el día a día durant aquests 5 anys: la Rosa, la Vero, la Yolanda i l'Albert. He aprés moltíssim de tots vosaltres durant tot aquest temps. Vosaltres feu de Bioingenium un bon lloc per compartir, fins i tot en les èpoques més complicades.

Vull agraïr també a tots els estudiants de pràctiques la seva inestimable ajuda. I m'agradaría mencionar i reconèixer especialment l'esforç d'aquells que han participat de forma directa en el treball d'aquesta tesi: la Gemma, el Christian, la María, el Juanjo, la Núria, la Laia i l'Ángel. Vosaltres també heu ajudat a que això sigui una realitat, i espero que hagueu après tant de mi com jo al vostre costat.

A la Cata de Maltas™: Arnau, Eloy, Albert, Guillem, Peri y Roger, con quienes he compartido la tortura de escribir la quétesis! durante tanto tiempo. A pesar de las semanas difíciles, siempre hay tiempo para reírnos (sobretodo de nosotros mismos), de unas cervezas o de una escapada a las montañas. También quiero acordarme de Álex y de los buenos ratos que hemos pasado, uno haciendo deporte y el otro intentándolo.

A mi família, que aun sin entender muy bien que es esto de las levaduras y las proteínas recombinantes, me han apoyado con esfuerzo, incondicionalmente y sin descanso no solo a lo largo de estos años de tesis, si no también en todo el largo el camino recorrido hasta hoy.

Y entre ellos, mi último agradecimiento, el más especial de todos, es para mi otra mitad. Porque sin tu ayuda y apoyo esto nunca habría sido posible. Gracias Xènia.

Most people say that it is the intellect which makes a great scientist.
They are wrong: it is character.

Albert Einstein

A mi familia

SUMMARY 1

RESUM 3

Nomenclature 6

Abbreviations 8

OBJECTIVES 10

Chapter 1 14

1. General introduction .....

1.1. Lysosomal Storage Diseases .....

1.1.1. Lysosomal enzymes and their role in lysosomal storage diseases .....

1.1.2. Biogenesis and sorting of lysosomal enzymes .....

1.1.3. Treatment of lysosomal storage diseases .....

1.1.4. Current scenario in the market of recombinant enzymes for ERTs .....

1.1.5. Production of recombinant lysosomal enzymes: current expression systems and upcoming technologies .....

1.2. α-Galactosidase A as the model enzyme .....

1.2.1. Fabry disease .....

1.2.2. Human α-Galactosidase A .....

1.2.3. Expression of human α-Galactosidase A in recombinant systems 31

1.3. *Pichia pastoris* as a heterologous protein expression system 32

1.3.1. A brief history 32

1.3.2. Advantages of *Pichia pastoris* for the production of recombinant proteins 33

1.3.3. The methylotrophic nature of *P. pastoris* 36

1.3.4. Strains and phenotypes 38

1.3.5. Tools and methods for protein expression in *Pichia pastoris* 40

1.3.5.1. Integrative expression vectors 40

1.3.5.2. Integration of expression vectors into the yeast genome 43

1.3.5.3. Multi-copy strains 44

1.3.5.4. Episomal expression vectors 45

1.3.5.5. Protein glycosylation in *P. pastoris* and strain glycoengineering 46

1.3.5.6. Cultivation techniques 50

1.4. *Pichia pastoris* for the production of biopharmaceuticals: current situation 52

Chapter 2 59

2. CLONING AND EXPRESSION OF HUMAN α-galactosidase A IN *PICHIA PASTORIS* 59

2.1. Introduction 59

2.2. Materials and methods 61

2.2.1. Strains and culture media .....

2.2.2. Cloning of α-galactosidase A into a novel *Pichia pastoris* shuttle vector: pBIZα .....

2.2.3. Generation of rhGLA multi-copy vectors .....

2.2.4. Generation of a *P. pastoris* expression vector containing HAC1 for inducible protein expression under AOX promoter .....

2.2.5. Transformation of vectors into *P. pastoris* and clone isolation .....

2.2.6. Screening of *Pichia pastoris* isolated clones: expression assays 71

2.2.7. Enzymatic activity assay .....

2.2.8. Generation of episomal vector encoding for Cre recombinase and resistant-phenotype reversion procedure. .....

2.3. Results and discussion .....

2.3.1. Comparison of expression levels of alpha-galactosidase A between Mut^{+} and Mut^{S} strains.....

2.3.2. Effect of the overexpression of helper proteins on the productivity of α-galactosidase A.....

2.3.2.1. Overexpression of HAC1p (methanol-induced) in Mut^{+} and Mut^{S} strains .....

.....

2.3.2.2. Overexpression of HAC1 and PDI in a Mut^{+} α-galactosidase A producer clone .....

2.3.3. Improvement of α-galactosidase A productivity by increase of gene dosage 81

2.3.3.1. Evaluation of multi-copy strains isolated from agar plates of high antibiotic concentration 82

2.3.3.2. Evaluation of directed multi-copy strains generated by integration of multi-copy vectors 84

2.3.4. Summary of α-galactosidase A productivity improvements 86

2.3.5. Expression of α-galactosidase A variants 89

2.3.6. Generation of an $rhGLA_{C\text{-}HIS}$ producing strain as a base for the co-expression of glycosylation-related enzyme complexes. 92

2.3.6.1. Selection of $rhGLA_{C\text{-}HIS}$ producing clones: constitutive expression under GAP promoter or methanol-inducible co-expressed with HAC1p 92

2.3.6.2. Reversion of Zeocin resistant phenotype by transient expression of Cre recombinase 94

2.4. Conclusions 98

Chapter 3 103

3. Production of recombinant HUMAN α-galactosidase A in bioreactor 103

3.1. Introduction 103

3.2. Materials and methods 107

3.2.1. Strains 107

3.2.2. Preparation of pre-inoculums: media culture and cell growth 107

3.2.3. Bioreactor media formulation and additives composition 10

3.2.4. Bioreactor equipment 10

3.2.5. Methanol-limited fed-batch cultivation strategy 1

3.2.6. O_2-limited fed-batch cultivation strategy 1

3.2.7. Biomass determination 1

3.2.8. α-galactosidase A activity assay 1

3.3. Results and discussion 1

3.3.1. Production of GLA by methanol-limited induction strategy 113

3.3.2. Production of GLA by oxygen-limited induction strategy 1

3.3.3. Comparison between MetOH-limited and O_2-limited strategies 1

3.3.4. Production of GLA co-expressed with HAC1 under O_2-limited conditions 1

3.4. Conclusions 1

Chapter 4 134

4. Purification and characterization of HUMAN recombinant α-galactosidase A 1

4.1. Introduction 1

4.2. Materials and methods 1

4.2.1. Sample recovery from bioreactor and conditioning....136

4.2.2. Chromatography equipment....136

4.2.3. Purification of un-tagged rhGLA by chromatography....138

4.2.3.1. Step 1 – Anion exchange chromatography (AIEX)....138

4.2.3.2. Step 2 – Hydrophobic chromatography (HIC)....139

4.2.3.3. Step 3 – Gel filtration (GF) or Size exclusion chromatography (SEC)....140

4.2.4. Purification of rhGLA variants by galactose-affinity chromatography....140

4.2.5. Purification of $rhGLA_{C\text{-}HIS}$ by immobilized-metal affinity chromatography (IMAC)....141

4.2.6. Analytic procedures for sample characterization....142

4.2.6.1. SDS-PAGE and Western-Blot....142

4.2.6.2. Protein quantification....143

4.2.6.3. Protein identification by peptide mass fingerprinting and molecular weight determination by mass spectrometry....144

4.2.6.4. Protein size determination by dynamic light scattering (DLS) 145

4.2.6.5. Glycan analysis....145

4.3. Results and discussion....146

4.3.1. Early downstream....146

4.3.2. 3-step chromatographic process for the purification of rhGLA 148

4.3.2.1. Capture step: anion exchange chromatography 14

4.3.2.2. Hydrophobic chromatography..... 1

3.1.1. Size-exclusion chromatography 1

4.3.2.3. Overall process evaluation and discussion..... 1

4.3.3. Purification of rhGLA variants (rhGLA, rhGLAΔKDLL, $rhGLA_{C\text{-}HIS}$, $rhGLA_{C\text{-}RGD}$) by galactose-affinity chromatography 1

4.3.3.1. Set-up of galactose-affinity chromatography with rhGLA 171

4.3.3.2. Purification of rhGLA variants by galactose-affinity chromatography 1

4.3.4. Purification of 6xHis-tagged rhGLA ($rhGLA_{C\text{-}HIS}$) by nickel affinity chromatography..... 1

4.3.5. Protein characterization 1

4.3.5.1. Protein identification by peptide mass fingerprinting..... 1

4.3.5.2. Molecular weight by MALDI-TOF 1

4.3.5.3. Protein size by dynamic light scattering (DLS)..... 1

4.3.5.4. Glycosylation profile 1

4.4. Conclusions 1

Chapter 5..... 193

5. Development of *In vitro* assays for the determination of HUMAN RECOMBINANT α-galactosidase A UPTAKE and intracellular activity 193

5.1. Introduction 193

5.2. Materials and methods 195

5.2.1. Proliferation, subculture and preparation of cell lines for *in vitro* assays 195

5.2.2. Quantification of internalized enzyme by activity measurement 197

5.2.3. Fluorometric quantification of internalized enzyme by cytometry 198

5.2.4. Evaluation of internalized enzyme by fluorescence microscopy 198

5.3. Results and discussion 199

5.3.1. Evaluation of 2 cell lines (human Fabry fibroblasts vs. HUVEC) and different techniques for the evaluation of the uptake capacity of recombinant human α-galactosidase A 199

5.3.1.1. Quantification by extraction and activity measure of internalized enzyme 200

5.3.1.2. Fluorometric quantification of internalized enzyme by flow cytometry 203

5.3.1.3. Evaluation of internalized enzyme by fluorescence microscopy 207

5.3.1.4. Overview of cell types and methods evaluated 208

5.3.2. Cellular uptake of rhGLA produced in *Pichia pastoris* versus the therapeutic product Agalsidase alfa 2

5.3.2.1. Quantification by extraction and activity measure of internalized enzyme 2

5.3.2.2. Fluorometric quantification of internalized enzyme by flow cytometry 2

5.3.2.3. Evaluation of internalized enzyme by fluorescence microscopy 2

5.4. Conclusions 2

GENERAL CONCLUSIONS 218

REFERENCES 220

S U M M A R Y

Human α-galactosidase A (GLA) belongs to the family of lysosomal hydrolases involved in lysosomal storage diseases (LSDs). Enzyme replacement therapies (ERTs), based on the administration of the recombinant missing enzyme, are an efficient measure to slow the advance of some LSDs. However, the cost of currently available ERTs is extraordinarily high. Therefore, the use of alternative production systems is proposed as a solution. *P. pastoris* is a promising system for the production of biopharmaceuticals, even though its glycosylation pattern needs to be engineered for such purpose. GLA, involved in Fabry disease, is the enzyme selected by Bioingenium to evaluate the feasibility of the yeast to produce human lysosomal hydrolases and, if so, to use it as the model protein in strain glycoengineering.

First, the expression of recombinant human GLA (rhGLA) in *P. pastoris* was evaluated in different strains. Once selected the most suitable one, the expression levels were significantly improved by overexpression of HAC1p, a transcriptional activator. Then, the specific productivity of rhGLA was optimized by cloning multiple copies of the coding gene. Overall, the productivity was improved about 7-fold compared to the initial. Additionally, variants of the enzyme containing a tag, a cell internalization motif or lacking a C-terminal sequence were successfully expressed. Lastly, a markerless clone expressing rhGLA was generated for its use as a base strain for glycan modification.

A scalable bioreactor process based on O_2-limiting conditions has been established for the production of high quantities of the enzyme. Moreover, purification methods have been developed for the isolation rhGLA. Specifically, a 3-step chromatography process is described for the purification of significant amounts of untagged enzyme, while affinity-chromatography procedures were defined for the fast isolation of small amounts of rhGLA variants aimed for product characterization.

Finally, a set of analytical methods have been implemented for the characterization of rhGLA, including glycan analysis and *in vitro* assays. By using fibroblasts from a Fabry patient, the uptake efficiency of the recombinant enzyme could be easily assessed. As expected, the rhGLA having yeast-like glycans showed poor internalization levels.

Overall, *P. pastoris* has been shown as a suitable production system for rhGLA. The methods developed in this work are of great value for future strain

glycoengineering and production scale-up, which will ultimately qualify the yeast for the production of different lysosomal enzymes for ERTs.

R E S U M

L'α-galactosidasa A humana (GLA) pertany a la família d'hidrolases lisosomals involucrades en malalties d'emmagatzemat enzimàtic (MEEs). Les teràpies de reemplaçament enzimàtic (TREs), basades en l'administració de l'enzim disfuncional, són una mesura eficient per alentir l'avanç d'algunes MEEs. Tanmateix, el cost de les TREs disponibles actualment es extremadament elevat. Per tant, es proposa com a solució l'ús de sistemes de producció alternatius. *P. pastoris* és un sistema prometedor per la producció de productes terapèutic biològics, tot i que el seu patró de glicosilació ha de ser adaptat per a aquest ús. GLA, involucrat en la malaltia de Fabry, és l'enzim seleccionat per Bioingenium per avaluar la viabilitat del llevat per produir hidrolases lisosomals, i en tal cas, per al seu ús com a proteïna model per a la modificació del patró de glicosilació *P. pastoris*.

En primer lloc, es va avaluar l'expressió GLA recombinant (rhGLA) en diferents soques de *P. pastoris*. Un cop seleccionada la més adient, els nivells d'expressió és van millorar mitjançant la sobreexpressió de HAC1p, un activador transcripcional. A continuació, es va optimitzar la productivitat específica de rhGLA mitjançant el clonatge de múltiples còpies del gen codificant. En conjunt, es va aconseguir millorar 7 vegades la productivitat comparat amb els valors inicials. Addicionalment, es van expressar amb èxit variants de l'enzim amb *tag*, amb un domini d'internalització cel·lular, o mancat d'una seqüència C-terminal. En darrer lloc, es va obtenir un clon lliure de marcadors de resistència per al seu ús com a soca base per la modificació del patró de glicosilació.

Es va establir un procés escalable de producció en bioreactor basat en condicions d'O_2 limitant, permetent la producció d'elevades quantitats d'enzim. A més, es van desenvolupar mètodes cromatogràfics per a l'aïllament de rhGLA. En concret, s'ha descrit un procés basat en 3 passos cromatogràfics per la purificació de quantitats significants d'enzim lliure de *tag*, així com procediments de cromatografia d'afinitat per la purificació de petites quantitats de variants de rhGLA destinades a ser caracteritzades.

Finalment, s'ha implementat un conjunt de tècniques per la caracterització de l'enzim, incloent anàlisis de glicans i assajos *in vitro*. Mitjançant l'ús de fibroblasts provinents d'un pacient de Fabry, s'ha pogut valorar amb simplicitat la eficiència d'internalització de l'enzim recombinant. Tal i com s'esperava, la rhGLA amb

estructures de glicosilació pròpies de llevat va mostrar nivells d'internalització pobres.

En general, s'ha demostrat que *P. pastoris* és un sistema d'expressió adequat per a la producció de rhGLA. Els mètodes desenvolupats en aquest treball són de gran valor per futures investigacions relacionades amb la humanització dels patrons de glicosilació de *P. pastoris*. A més, els resultats són d'utilitat per l'escalat dels processos de producció, que habilitaran el llevat per la producció de diferents enzims lisosomals destinats a l'ús en TREs.

AU	Activity units (nmol MU h^{-1} or µmol MU h^{-1})
dO	Dissolved oxygen (% saturation)
µ	Specific growth rate (h^{-1})
mAU	miliAbsorbance units (Abs 280 nm)
q_S	Specific substrate consumption rate (mg substrate g X^{-1} h^{-1})
T	Temperature (ºC)
t	Time (h)
V	Volume (L)
X	Biomass concentration (g DCW L^{-1})
$Y_{P/S}$	Product/Substrate Yield

CHO	Chinese Hamster Ovary
CiP	Cleaning-in-place
CV	Column volume
DCW	Dry cell weight
DF	Depth filtration
DLS	Dynamic light scattering
DoE	Design of experiments
ER	Endoplasmic reticulum
ERT	Enzyme replacement therapy
FDA	Food and Drug Administration
FT	Flow through
FTO	Freedom-to-operate
GF	Gel filtration
GILT	Glycosylation-independent lysosomal targeting
GLA	Human α-galactosidase A
GOI	Gene of interest
GRAS	Generally recognized as safe
HUVEC	Human Umbilical Vein Endothelial Cells
IMAC	Immobilized-metal affinity chromatography
LSD	Lysosomal storage disease
MetOH	Methanol
MF	Microfiltration
MLFB	Methanol-limited fed-batch
Mut^S	Methanol utilization slow

OLFB	Oxygen-limited fed-batch
OUR	Oxygen uptake rate
PARS	*P. pastoris* autonomous replicating sequence
PBS	Phosphate buffered solution
PDI	Protein disulfide isomerase
PTM	Post-translational modification
RCT	Research Corporation Technologies
rhGLA	Recombinant human α-galactosidase A
ROS	Reactive oxygen species
SCP	Single-cell protein
SEC	Size exclusion chromatography
SN	Supernatant
TFF	Tangential flow filtration
TGN	Trans-Golgi network
TLFB	Temperature-limited fed-batch
UF	Ultrafiltration
UPR	Unfolded protein response
VLP	Virus-like particles
WO	Wash-out

OBJECTIVES

The main objective of this thesis is to optimize the expression of human α-galactosidase A in the methylotrophic yeast *Pichia pastoris*, and develop a set of tools and procedures to enable its use as a model lysosomal enzyme in future strain glycoengineering. The pursued outcome would be useful for the development of the expression system to produce therapeutic recombinant lysosomal hydrolases, which could be used for the treatment of lysosomal storage diseases.

The following milestones can be derived from the main objective:

- Maximize the productivity of the expression system.
- Develop a scalable, inexpensive and robust bioreactor process to produce high amounts of product.
- Define a scalable purification procedure to obtain highly purified enzyme with high recovery yields. Additionally, simple and fast purification methods should be developed easily obtain purified enzyme for analysis.
- Implement analytical methods for the characterization of the enzyme, with special focus on glycan determination and assessment of cellular uptake efficacy by means of *in vitro* assays.

In the end, the methodologies should allow to stablish a platform for the screening of strain modifications regarding the glycosylation profile, and their effect over the glycosylation of human α-galactosidase A defined as the model lysosomal protein. The glycan pattern of lysosomal enzymes is key not only for *in vivo* stability, but to achieve the required cellular uptake and the desired therapeutic effect.

The accomplishment of these objectives would represent a step forward in the research related with the humanization of *P. pastoris*, a promising system for the attainment of more affordable, efficient and secure production processes for enzyme replacement therapies.

Chapter 1

General introduction

Chapter 1

GENERAL INTRODUCTION

1.1. Lysosomal Storage Diseases

1.1.1. Lysosomal enzymes and their role in lysosomal storage diseases

Lysosomal storage diseases (LSDs) are inherited metabolic disorders caused by the absence or dysfunction of lysosomal-related proteins, leading to the accumulation of intermediate substrates into the lysosomes. Even though they are individually classified as rare diseases due to their low prevalence (fig. 1.1), they are estimated to affect cumulatively 1 in 8000 live births (Fuller et al., 2006).

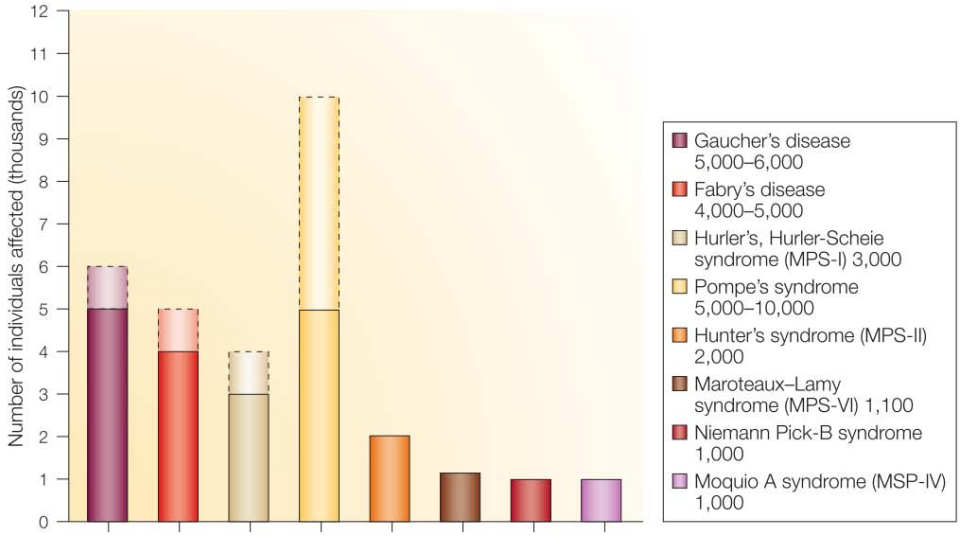


Fig. 1. 1. Global prevalence of the main lysosomal storage diseases. (Source: Werber, 2004).

Depending on the substrate accumulated and the organs and systems affected, different clinical manifestations arise that characterize each LSD. Neurological,

skeletal, visceral, haematological and/or ocular affectations, between others, are typically experienced by patients (Parenti et al., 2015). Genetically, LSDs originate from inheritable mutations in genes involved in the lysosomal function, encoding for different types of proteins: membrane proteins, activator proteins, transporters, non-lysosomal proteins, and especially acid hydrolases. Soluble acid hydrolases, or lysosomal enzymes, constitute a family of more than 50 members, each one responsible for the degradation of a specific substrate (Filocamo and Morrone, 2011). Due to the heterogeneity in the nature of the molecules to be processed (mainly glycosaminoglycans, sphingolipids and glycoproteins) and the multiple catalytic steps required, a wide range of enzymatic activities are represented in this group, such as glycosidases, lipases, peptidases, sulfatases, phosphatases and nucleases (Saftig, 2006). Since each hydrolase is responsible for a specific step within the degradation cascade, the lack or deficiency of a single one causes the lysosomal accumulation of undegraded substrates in several organs and tissues, leading to the clinical manifestations of the disease.

1.1.2. Biogenesis and sorting of lysosomal enzymes

Expression of lysosomal enzymes is directed to the secretory pathway in order to reach their final destination: the endosomes that will later mature into lysosomes by progressive luminal acidification. Thus, their respective mRNAs encode signal peptides to direct the expression to the rough endoplasmic reticulum (ER), where protein precursors are co-translationally glycosylated. As any other secreted glycoprotein, lysosomal hydrolases follow the common glycosylation process that takes place into the ER, ending in high-mannose Man8-NAcGlc2 structures (fig. 1.3, part b). Once the primary glycosylation is completed, proteins are loaded into COPII vesicles and transferred to the *cis*-Golgi compartment. From this point, glycosylation pathways differ between those proteins aimed to be secreted and lysosomal enzymes that need to be directed to the endosomes. While the first

undergo a multi-step process involving different glycosidases and glycosyltransferases to yield complex oligosaccharides ending in sialic acid residues, lysosomal proteins follow the mannose-6-phosphate (M6P) glycosylation pathway. Firstly, these proteins are specifically recognized in the *cis*-Golgi by the N-acetylglucosaminyl-phosphotransferase that catalyzes the addition of UPD-GlcNAc residues to the Man8-NAcGlc2 core structure. In a second step, taking place in the *trans*-Golgi and the *trans*-Golgi network (TGN), phosphate residues are uncovered by a phosphodiesterase, generating the mature M6P structure (Coutinho et al., 2012) (fig. 1.2).

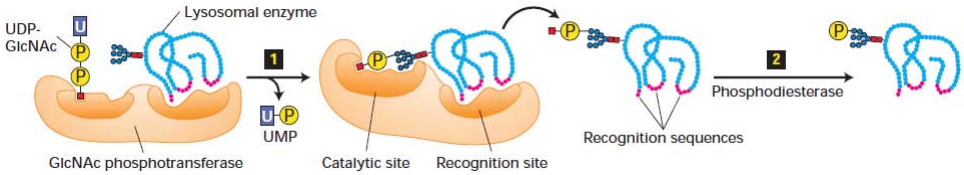


Fig. 1. 2. Mannose-6-phosphate glycosylation pathway.
(From: Molecular Cell Biology 7ed., by Harvey Lodish, et al., Copyright 2013, by W.H. Freeman and Company. Used by Permission of the publisher).

Indeed, M6P glycosylation pattern is the key signal through which lysosomal enzymes are sorted from other proteins in the secretory pathway. Phosphorylated glycans are recognized by M6P receptors located in budding vesicles from the TGN and in the plasma membrane (Kor nfeld, 1990). These vesicles fuse each other, with endosomes or directly to lysosomes, where the acidic environment releases the cargo proteins attached to the receptors (which will be recycled into vesicles directed back to the TGN to bind newly synthetized enzymes) (fig. 1.3, part a). Since the catchment of phosphorylated proteins is not fully efficient, some of them are released through the constitutive secretory pathway. Here come into play M6P receptors from the cell surface (specifically cation-independent M6P receptors, CI-M6PRs), which are responsible for the uptake and recovery of lysosomal enzymes accidentally secreted, also from neighbor cells (Dahms et al., 2008).

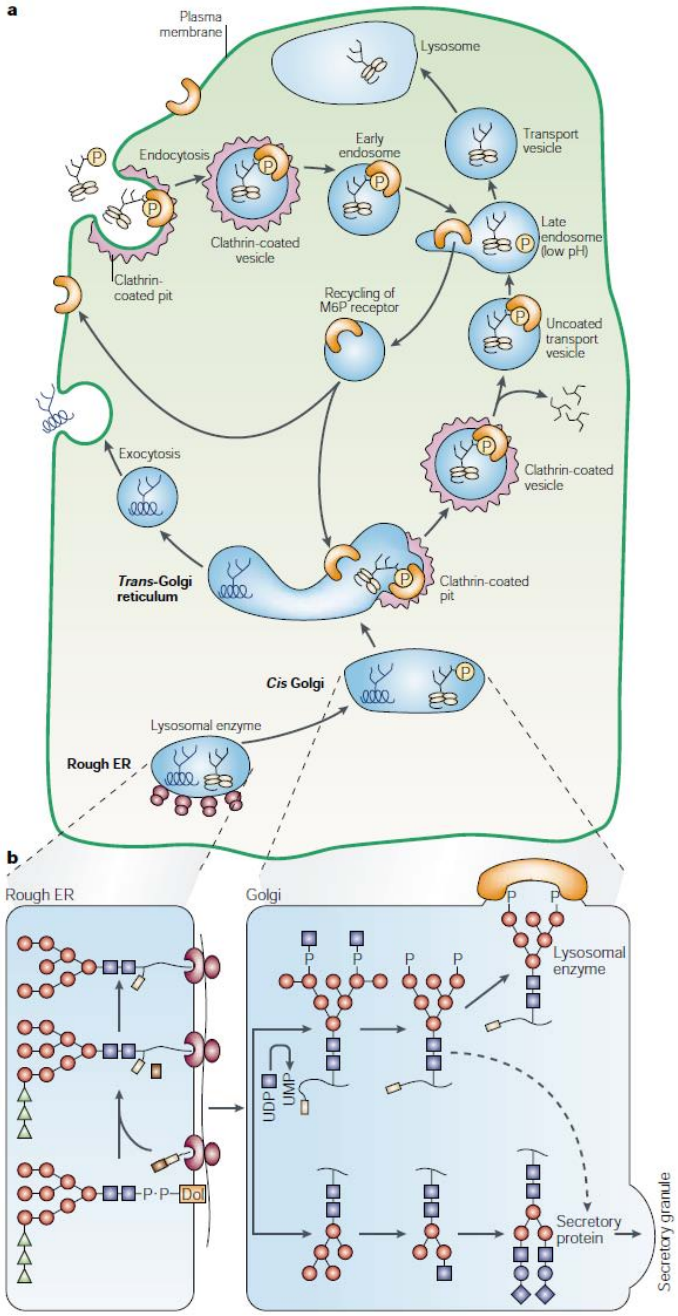


Fig. 1. 3. Sorting of lysosomal enzymes through the secretory pathway by the M6P receptor system. (Source: Desnick and Schuchman, 2002).

Once delivered to the endosomes (ph ~6.5), the subcellular vesicles undergo a progressive acidification responsible for their maturation into lysosomes (pH 4.5-5). This process, generated by active pumping of protons into the lumen by the V-type ATPase (Mindell, 2012), leads to the catalytic activation of the acidic hydrolases. Once activated, the enzymes will be ready to degrade their specific substrates, coming from autophagy of cell debris, phagocytosis of pathogens and endocytosis of extracellular material (Coutinho and Alves, 2015) (fig. 1.4).

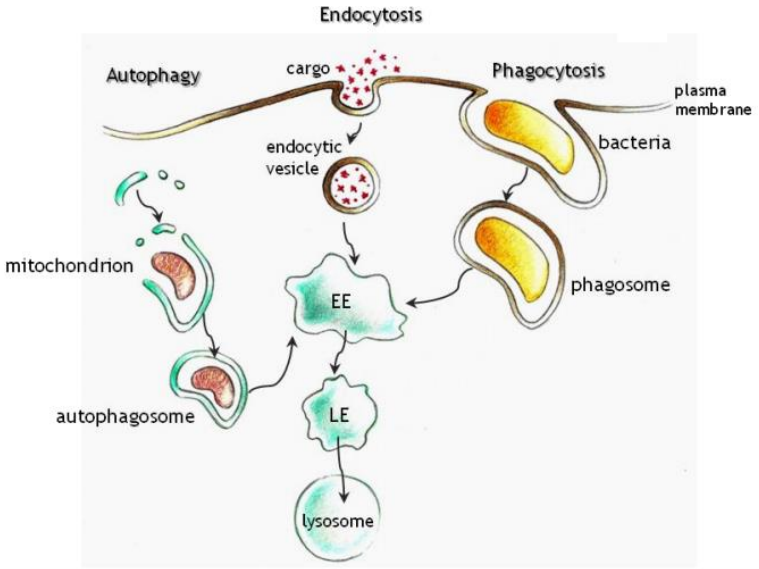


Fig. 1. 4. Pathways leading to lysosome biogenesis. (Source: Coutinho and Alves, 2015).

1.1.3. Treatment of lysosomal storage diseases

For a significant amount of LSDs, the only available treatment is still based on palliative therapy focused on the relieving of specific symptoms caused by the disorder. Typical therapies include a multidisciplinary approach based on the management of respiratory, neurological, cardiac, vascular, orthopedic, ocular and hematological defects, between others, either by continuous treatment or surgical interventions (Parenti et al., 2015).

For a selection of LSDs, the scenario began to change two decades ago with the emergence of the first enzyme replacement therapies (ERTs). Although palliative treatments remain to be essential in patient care, ERTs proved to be efficient in the reversion of critical manifestations and pathophysiology improvement of the target disease. These treatments are based on the correction of the missing or defective activity by providing periodically recombinant enzyme. Up to date, approved therapies are administered intravenously (Grubb et al., 2010), although new routes of administration are being evaluated, e.g. oral (www.protalix.com). Glucocerebrosidase (Imiglucerase) for Gaucher disease, the most common lysosomal storage disorder, was approved in 1991 as the first ERT. Although glucocerebrosidase was initially obtained from placental tissue, the recombinant enzyme has been available since 1994 (Werber, 2004). From that moment, several ERTs based on the infusion of recombinant enzyme have been progressively approved by the regulatory agencies, helping to revert significantly the signs of patients suffering different LSDs (Table 1.1). The group includes, apart from glucocerebrosidase, α-galactosidase A for Fabry disease, a-L-iduronidase for Mucopolysaccharidosis (MPS) type I, iduronate-2-sulfatase for MPS type II, N-acetylgalactosamine-6-sulfatase for Morquio syndrome, Arylsulfatase B for Maroteaux-Lamy syndrome, Sebelipase alfa for Wolman disease, and α-glucosidase for Pompe disease.

Table 1. 1. Approved treatments for LSDs. (Adapted from Parenti et al., 2015).

Disease	Approved treatments	Company	Production system	Type of treatment (route of administration)
Pompe disease	Alglucosidase alfa	Genzyme (Sanofi)	CHO	ERT (IV)
Fabry disease	Agalsidase alfa	Shire	Human fibroblast cell line (HT1080)	ERT (IV)
	Agalsidase beta	Genzyme (Sanofi)	CHO	
Gaucher disease type I	Imiglucerase	Genzyme (Sanofi)	CHO	ERT (IV)
	Velaglucerase alfa	Shire	Human fibroblast cell line (HT1080)	
	Taliglucerase alfa	Protalix Biotherapeutics	Carrot root cells	
	Miglustat	Actelion Pharmaceuticals	Synthetic small molecule	SRT (O)
	Eliglustat	Genzyme (Sanofi)	Synthetic small molecule	
MPS I (Hurler, Hurler-Scheie and Scheie syndromes)	Laronidase	BioMarin Pharmaceuticals	CHO	ERT (IV)
MPS II (Hunter syndrome)	Idursulfase	Shire	Human fibroblast cell line (HT1080)	ERT (IV)
MPS IVA (Morquio A syndrome)	Elosulfase alfa	BioMarin Pharmaceuticals	CHO	ERT (IV)
MPS VI (Maroteaux-Lamy syndrome)	Galsulfase	BioMarin Pharmaceuticals	CHO	ERT (IV)
Niemann-Pick disease type C	Miglustat	Actelion Pharmaceuticals	Synthetic small molecule	SRT (O)
LAL (Wolman disease)	Sebelipase alfa	Synageva BioPharma (Alexion)	Egg white from transgenic *Gallus*	ERT (IV)

ERT: enzyme replacement therapy. SRT: substrate reduction therapy.
IV: intravenous. O: oral.

Alternative approaches, which can be used alone or in combination with available ERTs, are being assessed for the treatment of LSDs. In those cases where the disease is caused by mutations affecting the stability of the enzyme, chaperone-based therapies are a promising option. These are based on the enzyme stabilization by interaction with chemical entities in the ER, thus preventing the potential degradation of the enzyme. The chaperones are bound to the protein through the secretory pathway until they reach the lysosome, where the acidic pH contributes to the dissociation and the enzyme activation. Some chaperone-based therapies are already facing late stage clinical trials (Warnock et al., 2015). With treatments already approved, other therapies focused on the reduction of available substrate (substrate reduction therapies, SRTs) have shown to be an efficient alternative to treat Gaucher disease type I and Niemann-Pick disease type C. Miglustat, an iminosugar that reversibly inhibits glucosylceramide synthase and reduces the biosynthesis of macromolecular substrates that accumulate pathologically in glycosphingolipidoses (Shemesh et al., 2015), is used for the treatment of mild Gaucher disease type I patients for whom the ERT is not an option (Cox et al., 2003) and for patients with Niemann-Pick type C disease. Eliglustat, another glucosylceramide synthase inhibitor, is used to treat Gaucher disease patients and can be prescribed as a combination regimen with available ERTs (Marshall et al., 2010).

All the treatment strategies discussed above are either palliative or corrective, and none of them is aimed to cure the underlying disease. Gene therapy is, on the other hand, a promising future alternative that would heal patients by delivering a wild-type copy of the defective gene. The monogenic defects causing the disease make LSDs patients ideal candidates for such therapies. Moreover, the severity of some untreatable LSDs (some of which are characterized by a pediatric onset), together with recent developments in the security increase for gene therapies, have enabled a permissive position for conducting experimental clinical trials using retroviral, lentiviral or adeno-associated viral vectors (Parenti et al., 2015).

1.1.4. Current scenario in the market of recombinant enzymes for ERTs

LSDs patients and health agencies deal with different obstacles to respectively receive and provide effective treatments. In first place, a great amount of LSDs remain without any approved corrective therapy to date, and palliative measures are the only option available. Great efforts are dedicated to promote the development and approval of new treatments for orphan diseases, but the extreme low prevalence of some of them can compromise the economic viability of research initiatives. However, orphan drug designations receive special advantages that makes them an attractive spot for new developments: faster regulatory pathways, longer market exclusivities, lower marketing costs and premium pricing (Hall and Carlson, 2014). Indeed, the cost of LSDs is usually above €200,000 per patient per year, ranging from €30,000 for infants to over €400,000 for adults because dosage depends on patients' weight (Tambuyzer, 2010). Pharmaceutical companies already supplying ERTs claim that additional factors affect pricing apart from the manufacturing process for small patient populations, including the development of new products, free-drug programs and, specially, the health benefits for patients and the consequent value added to health care systems. However, the cost of treatments suggest an evident correlation with the prevalence of the disease: the rarer it is, the more expensive the treatment (Jahnke, 2014). Moreover, for those diseases having more than one eligible product (Gaucher and Fabry), costs of ERTs are palpably lower (fig. 1.5).

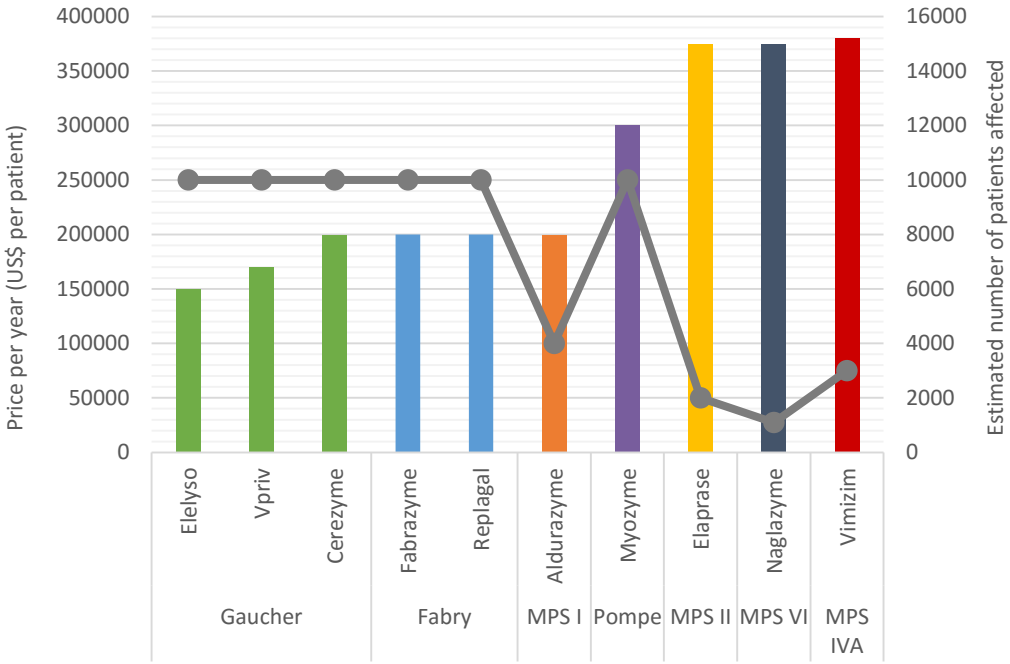


Fig. 1. 5. Correlation between global estimated prevalence of LSDs and the cost of their respective ERTs. Colored bars: cost of ERTs (in US$) per patient per year. Grey line: estimated number of patients affected worldwide. (Adapted from Jahnke, 2014).

Definitely, these treatments provide significant earnings to their respective marketing companies, proving that the market of orphan-drugs (and specifically those dedicated for the treatment of LSDs) can lead to profitable business models. In fact, this situation could explain why the treatment of some diseases have experienced the irruption of alternative ERT suppliers (e.g., Velaglucerase alfa in 2010 and Taliglucerase alfa in 2012 after Cerezyme, which was first approved in 1994 for the treatment of Gaucher disease). The attractiveness of biopharmaceuticals for LSDs is even enhanced by predictions of an increasing orphan-drug market (Hadjivasiliou, 2014). Thanks to the improvement of diagnostic procedures and methods, together with the greater awareness by physicians regarding rare diseases, an increase in the number of patients previously undiagnosed or misdiagnosed that will require corrective treatments is expected.

This is directly linked to the forecasted rise in sales of currently approved ERTs (fig. 1.6).

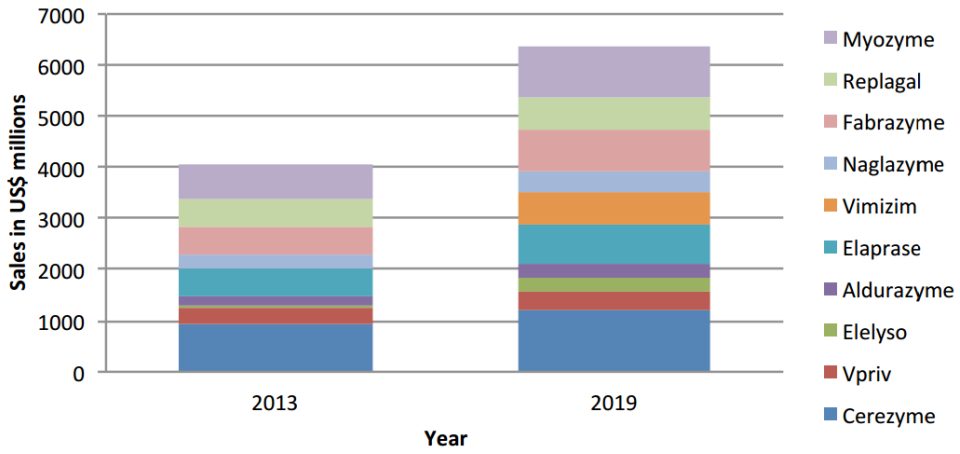


Fig. 1. 6. Market growth forecast for the available ERT products. (Source: Jahnke, 2014).

1.1.5. Production of recombinant lysosomal enzymes: current expression systems and upcoming technologies

Mammalian cell lines are currently the workhorse for the production of recombinant lysosomal hydrolases. Chinese Hamster Ovary (CHO) cells and HT1080 (fibroblasts derived from fibrosarcoma) are the expression systems of choice for the manufacturing of actual marketed products, the first being used mainly by Genzyme and BioMarin and the later by Shire. The main advantage of mammalian systems is their ability to perform human-like glycosylation patterns, which turns to be a critical feature for lysosomal enzymes because, as mentioned previously, M6P glycosylations are the key signals for protein uptake by cells towards the endosomal pathway. An exception to the rule is glucocerebrosidase, which requires exposed mannose residues to be targeted to the macrophages from the reticuloendothelial system (LeBowitz et al., 2014). The first recombinant glucocerebrosidase, Imiglucerase (developed by Genzyme and produced in CHO

cells), requires glycan trimming by exoglycosidases in the downstream process (Walsh and Jefferis, 2006). On the other hand, Velaglucerase (produced in HT1080, by Shire) contains high-mannose glycans thanks to the use of kifunensin in culture, a mannosidase I inhibitor that prevents the formation of complex glycans. Interestingly, a third version of glucocerebrosidase produced by Protalix in genetically modified carrot cells has been recently approved (known as Taliglucerase alfa), containing a Man_3 exposed structure (Tekoah et al., 2013). In principle, the rest of lysosomal enzymes require the mentioned M6P glycosylation (Oh, 2015).

Production in mammalian cell lines implies significant drawbacks. Cultures require complex media that contributes significantly to high manufacturing costs, affecting consequently the price of treatments. Moreover, these type of cultures are susceptible to viral infections that can seriously compromise the continuous supply needed by patients. In 2009, a contamination by Vesivirus 2117 in one of Genzyme's plants forced to halt the manufacturing process of Agalsidase beta (Fabrazyme®), used to treat for Fabry patients, and Imiglucerase (Cerezyme®), for Gaucher patients (Bethencourt, 2009). The situation caused extended shortages in the US that affected patients with ongoing treatments. The situation lasted until 2012, when the production was reallocated in a new manufacturing plant.

The development of alternative expression systems for the production of recombinant lysosomal enzymes could help to lower the cost of treatments and increase manufacturing safety, therefore improving the availability of corrective therapeutic options for patients and contributing to minimize the probability of product shortages. In addition, the recent expiration of patents and market exclusivities for some products (e.g. α-galactosidase A products, namely Agalsidase alfa from Shire and Agalsidase beta from Genzyme), opens the door to the emergence of biosimilars as long as they reach at least equal therapeutic effects. In some cases, the availability of novel expression systems could enable the production of improved therapeutic entities (also known as "biobetters" or

"biosuperiors") showing superior performance, either by increasing the potency, uptake capacity, stability or blood half-life time of existing enzymes, between others (Courtois et al., 2015). Finally, the emergence of such alternative systems could also help to develop new recombinant enzymes to treat LSDs for which no corrective therapy is still available, such as α-manosidosis, metachromatic leukodystrophy, Sanfilippo A and B syndromes (MPSIIIA and B), Sly syndrome, Neuronal ceroid lipofuscinosis or Niemann-Pick disease type B, among many others. The production of Taliglucerase alfa (glucocerebrosidase for the treatment of Gaucher disease) is a great example of a new alternative system for the expression of an existing enzyme: as mentioned previously, it is produced in genetically modified carrot cells (ProCellEx® system, developed by Protalix) (Tekoah at al., 2013, Shemesh et al., 2013). Another example is provided by a singular novel expression system based on enzyme purification from egg whites laid by transgenic hens, which has been developed for the manufacturing of Sebelipase alfa (acid lipase) (Shirley, 2015). The resulting product has been recently authorized (August 2015) for the treatment of Wolman disease, becoming its first corrective therapy available.

In this context, yeast expression systems are of great interest for the production of biopharmaceuticals. Thanks to their associated advantages (which will be discussed in detail later in this chapter), great productivity levels of fully functional proteins can be expected while reducing production costs. Indeed, yeast-based systems are already being assessed for the production of lysosomal enzymes for ERTs, and significant advances have been made in *Yarrowia lipolytica*. In order to overcome the lack of endogenous M6P glycan structure, *Y. lipolytica* was firstly engineered to yield mannose-P_i-6-mannose structures. Later in the downstream process, a novel bacterial glycosidase was used to uncap the P_i by removing the coupled mannose residues, and the glycan structures were finally refined thanks to the over-expression of an α-manosidase, thereby obtaining a functional enzyme *in vivo* (Tiels et al., 2012). The technology is being actually exploited by Oxyrane (Belgium).

Finally, the use of new targeting strategies is being evaluated to bypass the dependence on M6P glycans and achieve an efficient cellular uptake. Known as Glycosylation-independent lysosomal targeting (GILT), this strategy was successful in targeting β-glucuronidase (LeBowitz et al., 2004) and α-glucosidase (Maga et al., 2013) to the lysosomes: by taking advantage of the CI-M6PR ability to recognize IGF-II, the enzymes were fused to a portion of this protein hormone. The modification enabled their efficient delivery in relevant disease mice models (MPS VII for β-glucuronidase and Pompe disease for α-glucosidase).

Although ERTs are usually defrayed by insurers or governments, the extreme high pricing of these treatments potentially compromise the accessibility to all diagnosed patients. Hence, the use of alternative expression systems allowing better productivities at lower production costs, and with increased security, would be an interesting advance not only for patients, but for social security systems and private insurers too. Along this line, Bioingenium SL is committed to the development of a yeast-based expression system engineered for the production of a wide range of recombinant lysosomal enzymes.

1.2. α-Galactosidase A as the model enzyme

1.2.1. Fabry disease

As mentioned previously, Fabry disease is a human pathology included within the group of LSDs. The disease, initially described as “angiokeratoma corporis diffusum” by dermatologists Johannes Fabry and William Anderson (Schiffmann, 2009), is caused by deficiencies in the enzyme α-galactosidase A. The partial or complete loss of its activity leads to the progressive accumulation of globotriaosylceramide and glycosphingolipids within lysosomes (Germain, 2010). Fabry disease is a pan-ethnic hereditary X-linked disorder mainly affecting males, although females can present mild variants of the disease presumably due to random X-chromosome inactivation (Deegan et al., 2006). The classic form is due

to α-galactosidase A deficiencies leading to less than 1% of the normal activity, and clinical manifestations start to appear in the childhood or adolescence. However, diagnosis is usually delayed due to the unspecificity of initial symptoms and signs. Patients suffer gradual deterioration of renal, cerebrovascular and/or cardiac functions, the latter being the main cause of death. In the US, life expectancy has been estimated to be 58.2 years in males and 75.4 for females (4.6 years less than the general female population) (Mehta and Hughes, 2013). Even though the exact incidence of the disease is not clear, it has been globally estimated to be between 1: 40,000 and 1:117,000. Nevertheless, studies in Italy and Taiwan suggest that incidence could be as high as 1:3,100 and 1:1,600, respectively, although only a small fraction of affected neonates would carry a mutation leading to the classic phenotype (1 of 11 in the Italian study) (Zarate and Hopkin, 2008). Fabry disease is the second most prevalent LSD, which is one of the reasons why it was chosen for the present work.

Fabry disease is one of the few LSDs with an available ERT and, in fact, two products are available on the market. Fabrazyme® (agalsidase beta, manufactured by Genzyme), is expressed in a DUKX B11 CHO cell line (Ioannou et al., 2001), while Replagal® (agalsidase alfa, manufactured by Shire) is produced in a fibrosarcoma HT1080 cell line. Unlike in Europe, where both products are used, in the US Fabrazyme is the only available thanks to market exclusivity issues (Morel and Clarke, 2009). While they share identical amino acid sequences and are considered equivalent therapeutic entities, minor differences have been described regarding their glycosylation profiles (Lee et al., 2003) that could explain why a higher percentage of patients develop antibodies against Agalsidase beta than Agalsidase alfa (Vedder et al., 2008). As other ERTs, these treatments are extraordinarily costly. The average per patient per year is estimated between €115,000 and €155,000. As an example of the expenditure for social security systems, the annual expense to treat Fabry patients diagnosed in Catalonia is estimated within the range of €1,700,000 – €2,330,000 (Generalitat de CAtalunya,

Departament de Salut. Informe CAMUH. "Agalsidasa alfa (Replagal®) i agalsidasa beta (Fabrazyme®) per al tractament de pacients amb malaltia de Fabry", 2012). The combined global sales revenue of both products is forecast to reach $1,530 million in 2019 (Jahnke, 2014).

1.2.2. Human α-Galactosidase A

The human lysosomal enzyme α-galactosidase A (GLA, EC 3.2.1.22) is encoded by the GLA gene, located on the Xq22.1. The gene encodes for a 429 amino acid precursor which, upon translation and translocation into the ER lumen, is processed into a 398 amino acid glycoprotein by cleavage of its signal peptide. Structurally, the enzyme is an homodimer in which each monomer contains a $(\beta/\alpha)_8$ domain (harboring the active site) and a C-terminal domain with eight antiparallel β strands on two sheets in a β sandwich (Garman and Garboczi, 2004) (fig. 1.7).

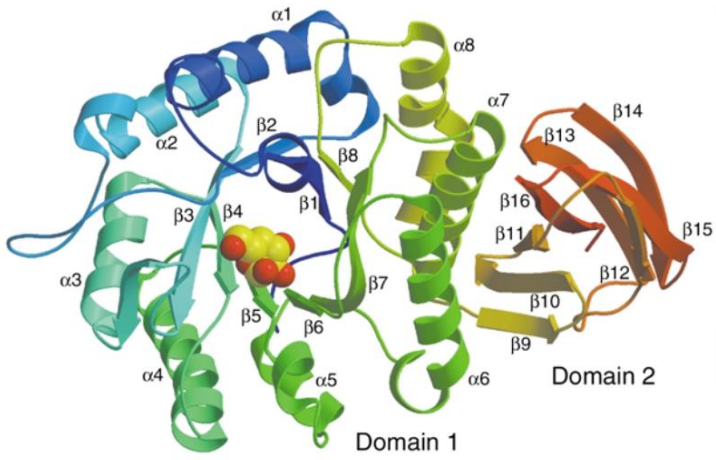


Fig. 1. 7. Structure of the α-galactosidase A monomer. The sequence is colored from N-terminus (blue) to C terminus (red). In yellow and red CPK atoms is shown the galactose ligand. (Source: Garman and Garboczi, 2004).

Each monomer contains the three N-linked carbohydrate sites (at N139, N192, and N215) and five disulfide bonds (C52–C94, C56–C63, C142–C172, C202– C223, and C378–C382). Interestingly, both endogenous and recombinant GLA show

significant heterogeneity in the attached carbohydrates (Ioannou et al., 2001; Lee et al., 2003). Moreover, the type of glycosylation is heterogeneous within the same molecule: while N139 contains a complex sialylated structure, N192 and N215 are glycosylated with the M6P pattern, responsible for the lysosomal targeting (Garman and Garboczi, 2004). Elimination of any two glycosylation sites lead to an unstable enzyme and premature degradation, while N215 was shown to be essential for protein activity, solubility and stability (Kang and Stevens, 2009). The homodimer has a molecular weight of ~100 kDa (~50 kDa per monomer) (Morel and Clarke, 2009).

α-galactosidase A is a hydrolase that catalyzes the removal of terminal α-galactose residues from polysaccharides, glycolipids, and glycopeptides during the catabolism of macromolecules (Guce et al., 2009) (fig. 1.8). The optimum pH for enzyme activity is 4.2, although it is stable in the range 4.0 - 7.0 (Ishii et al., 2000).

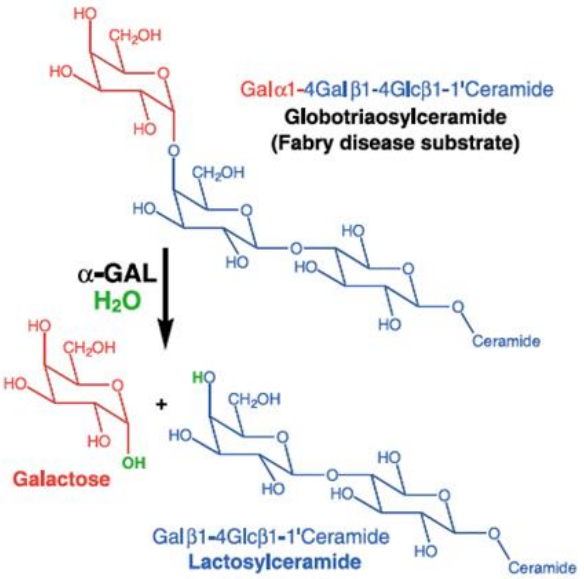


Fig. 1. 8. α-galactosidase A biological reaction. The lack of activity leads to globotriaosylceramide (GB_3) accumulation and Fabry disease. (Source: Garman and Garboczi, 2004).

Over 600 gene mutations of the protein have been described to date (Pisani et al., 2015). Patients with mutations leading to the total absence of α-galactosidase A suffer the classic form of Fabry disease, while those having missense mutations causing residual activity (2% to 25%) would suffer from milder forms of the metabolic disorder (Schiffmann, 2009).

1.2.3. Expression of human α-Galactosidase A in recombinant systems

α-galactosidase A has been expressed with different success in the main recombinant systems. Mammalian systems, including CHO and HT1080 cell lines as mentioned before (Lee et al., 2003) and human embryo kydney (HEK293) cell line (Corchero et al., 2011), are considered for the moment the most efficient for the expression of bioactive human lysosomal enzymes mainly because of their ability to perform naturally human glycosylation structures (including M6P). The expression was attempted in other well-known recombinant expression systems. While active protein could be purified from Sf9-baculovirus cultures (Yasuda et al., 2004; Chen et al. 2000a), *Escherichia coli* was unable to express active enzyme (Hantzopoulos and Calhoun, 1987; Unzueta et al., 2015). Arguably, the inability of common bacterial systems to carry out the essential glycosylations of GLA would be the main reason for the absence of catalytic activity.

Interestingly, more exotic systems based in plant cells were able to efficiently express active enzyme. Protalix Biotherapeutics used BY2 Tobacco cell cultures (Kizhner et al., 2015), while the moss *Physcomitrella patens* was successful too (Reski et al., 2015). Both systems have been engineered to perform human-like glycosylation patterns.

Finally, yeast systems have also been explored for the expression of GLA. *Saccharomyces cerevisiae* (Chiba et al., 2002; Sakuraba et al., 2006) and *Ogataea minuta* (Tsukimura et al., 2012) produced successfully the enzyme, and significant advances have been made in the humanization of strains (Chiba and Akeboshi, 2009). Moreover, the ability of *Pichia pastoris* to express the enzyme was

previously proven (Chen et al. 2000b). This represents one of key reasons for choosing α-galactosidase as the model enzyme for the developments presented in this work, whose main objectives are to tune up *P. pastoris* and optimize its expression capacity, develop scalable production and purification processes, and set-up analytical methods for the recombinant enzyme. Once achieved, GLA would be used as a model lysosomal enzyme for the yeast humanization.

1.3. *Pichia pastoris* as a heterologous protein expression system

1.3.1. A brief history

Pichia pastoris was firstly isolated in 1919 from exudates of a French chestnut tree by A. Guilliermond, who initially classified the new species as *Zygosaccharomyces pastori* and later as *Saccharomyces pastori.* Later in the 1940s, Herman Phaff obtained yeast isolates from an oak tree in the Yosemite region in California, and established the species *Pichia pastoris*. As the former *Saccharomyces pastori* was indistinguishable, it was merged into the newly described species (Cregg et al., 2010).

In 1968, the methylotrophic behavior of some yeast species (including *P. pastoris*) was discovered, attracting the attention of Phillips Petroleum Co. Aiming to use the yeast for the production of single-cell protein (SCP) for animal feed, the oil company started a development program based on the use of their excess methane (chemically oxidized to methanol) as a carbon source for cell growth. Between the different options explored, *P. pastoris* was selected for commercial development, and fermentation processes were successfully up-scaled to >100.000 L at >130 g/L (dry cell weight, DCW) of cell concentration (Macauley-Patrick et al., 2005). Unfortunately, the energy crisis in the 1970s increased notably the cost of natural gas and the program became economically unfeasible (Cereghino and Cregg, 2000). The newly available molecular biology techniques allowed Phillips to reformulate their strategy, this time focused on the production of high-value yeast biomass with enhanced amino acid content or containing recombinant growth

hormones, and a new research program to develop *P. pastoris* as a protein expression system was started in collaboration with the Salk Institute Biotechnology Industrial Associates Inc. (SIBIA). As a result, expression vectors and strains were ready in the late 1980s (Cregg et al., 1989). In 1994, Phillips transferred the associated patents to Research Corporation Technologies (RCT) and licensed Invitrogen Corp. to commercialize the expression system, making it available to researchers worldwide (Julien, 2006).

A key breakthrough occurred in 2000, when Tillman Gerngross and Charles Hutchinson co-founded Glycofi Inc., the company that developed the first *P. pastoris* strains with a fully humanized glycosylation pattern. Because of the great potential of the technology for the production of monoclonal antibodies, Glycofi was acquired by Merck & Co for $400 million in 2006.

In the late 1990s and early 2000s, newly available genome sequencing techniques revealed that the two original *P. pastoris* strains isolated independently in France and US are, in fact, different species (Kurtzman, 2005). Officially renamed as *Komagataella pastoris* (A. Guilliermond strain, NRRLY-1603) and *Kogamataella phaffii* (H. Phaff strain, NRRLY-11430), the later resulted to be surprisingly the biotechnological strain widely used for the expression of recombinant proteins (Kurtzman, 2009). Since then, *Pichia pastoris* remains as a colloquial designation for the expression system. Finally, in 2013 original patents of the *P. pastoris* system expired, allowing the commercialization of new vectors and strains by third parties and the use of freedom-to-operate (FTO) in-house systems.

1.3.2. Advantages of *Pichia pastoris* for the production of recombinant proteins

Pichia pastoris is one of the most extended eukaryotic systems for the production of recombinant proteins. The nature of yeasts allows them to be cultivated at high

cell densities in low-cost media (a typical feature of bacterial expression systems), while performing the main post-translational modifications required for the proper folding and/or functionalization of a large number of proteins. Moreover, it is able to secrete the product of interest, which is of great interest to simplify downstream processes. *Pichia* stands out above other yeast-based expression systems thanks to the easy accessibility to vectors, strains and methods related. The main advantages of the system for the production of heterologous proteins are summarized below (table 1.2).

Table 1. 2. Key features and benefits of *P. pastoris* expression system.

Feature	Benefit
Fast growth, achieving very high densities (>130 g/L of DCW)	High volumetric productivities, lower production costs, better scale economy
Generally, high product titers and high secretion performance	
Can grow on minimal, chemically defined media	Low batch-to batch variability, better process control, low cost of goods
Efficient and robust growth in different culture conditions. Can grow on wide ranges of pH and temperature, is able to assimilate several carbon sources and tolerates medium additives	Process parameters can be adapted for the convenience of the protein of interest (to maximize productivity and preserve bioactivity and stability)
Performs the main post-translational modifications, including N- and O-linked glycosylation, disulfide bonds or phosphorylation.	Suitable for the expression of proteins in which PTM's are essential for its final purpose
High success rate	High probability of expressing a functional recombinant protein
Well-established molecular biology tools and manipulation methods	Good technology accessibility, fast learning-curve

Stable integration of expression vectors into the yeast genome	High clone stability. Once isolated, there is no need to use antibiotics to maintain the expression profile.
Scalable production in bioreactor	Suitable for protein production at industrial scale
Low levels of endogenous protein secretion	Downstream processes simplified

Table 1. 2. Key features and benefits of P. pastoris expression system (continued).

Feature	Benefit
Complex mammalian-like glycosylation strains available	Potentially suitable for the production of antibodies, antigen vaccines, cytokines or growth factors
Since recently, accepted by regulatory agencies for the production of therapeutic proteins: endotoxin- and virus-free production system	Established framework for regulatory approval of product candidates

P. pastoris has been widely used for the expression of many different kinds of proteins, including industrial enzymes such as xylanases, cellulases, proteases, lipases, amylases, phosphatases, pectinases, phytases and laccases (Fickers, 2014), antibodies (full-length and fragments), hormones, cytokines, toxins, virus-like particles (VLPs), membrane proteins, peptides, fluorescent proteins, gelatins and collagens, among many others (reviewed in Fickers, 2014, Macauley-Patrick et al., 2005, Çalik et al., 2015; and Cereghino et al., 2002). It has shown to express successfully proteins from virtually any source possible: from bacteria, fungi, invertebrates, plants, viruses, non-human vertebrates, and human origin.

Remarkably, more than 1000 heterologous protein have already been expressed in *P. pastoris* (Cregg et al., 2009).

Moreover, *P. pastoris* is currently accepted by the Food and Drug Administration (FDA) as a GRAS (generally recognized as safe) production system for proteins used in animal feeding (Ciofalo et al., 2006), which extends its commercial interest to the field of food industry.

1.3.3. The methylotrophic nature of *P. pastoris*

P. pastoris belongs to the select group of known methylotrophic yeasts, which includes other species used for the expression of heterologous proteins such as *Hansenula polymorpha* (*Pichia angusta*), *Pichia methanolica*, *Candida boidinii* and *Ogataea minuta* (Mattanovich et al., 2012). This means that it is able to use methanol, a reduced C1-compound, as the sole source of carbon and energy. Methylotrophic yeasts have developed highly specialized metabolic pathways that are partly compartmentalized in peroxisomes, where specific enzymes are accumulated (van der Klei et al., 2006). The methanol assimilation pathway consist on several steps that transform methanol into dihydroxyacetone phosphate and glyceraldehyde-3-phosphate, intermediates of the glycolysis. The first reaction, responsible for the methanol oxidation to formaldehyde, is a key step in the pathway catalyzed by the alcohol oxidase (AOX) (fig. 1.9).

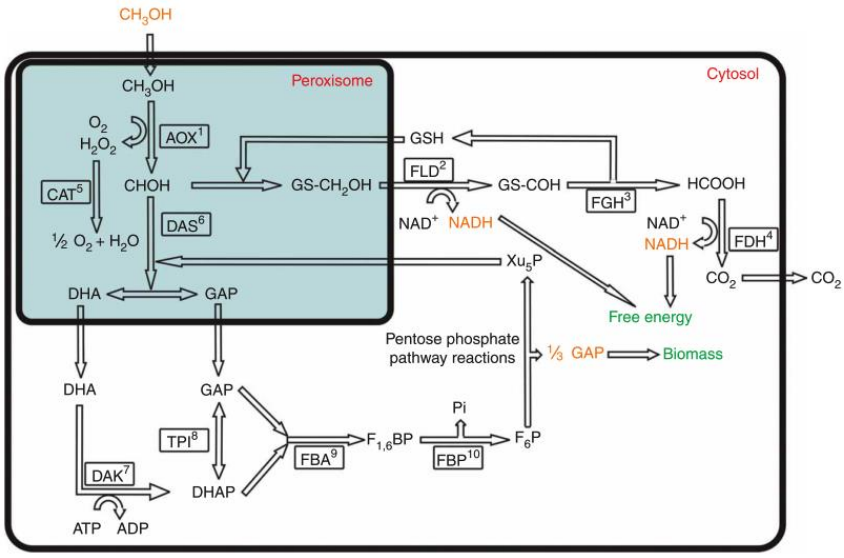


Fig. 1. 9. *Pichia pastoris* methanol utilization pathway. AOX1: alcohol oxidase, FLD2: formaldehyde dehydrogenase, FGH3: S-formylglutathione hydrolase, FDH4: formate dehydrogenase, CAT5: catalase, DAS6: dihydroxyacetone synthase, DAK7: dihydroxyacetone kinase, TPI8: triosephosphate isomerase, FBA9: fructose-1,6-biphosphate aldolase, FBP10: fructose-1,6-biphosphatase, DHA: dihydroxyacetone, GAP: glyceraldehyde-3-phosphate, DHAP, dihydroxyacetone phosphate, $F_{1,6}BP$: fructose-1,6-biphosphate, F_6P: fructose-6-phosphate, P_i: phosphate, Xu_5P: xylulose-5-phosphate, GSH: glutathione. (Source: De Schutter et al., 2009).

Alcohol oxidase has a low affinity for O_2, which is essential for methanol oxidation. For this reason, large amounts of enzyme are necessary to efficiently metabolize methanol as a substrate. This is achieved in part thanks to the presence of two alcohol-oxidase coding genes in the yeast genome: AOX1 and AOX2. However, and due to differences on their promoter strength, AOX1 is responsible for the 80-90% of the total AOX activity (Cregg *et al*, 1989). Therefore, AOX1 promoter is to a great extent the responsible of most of the AOX expressed, which can account for the 30% of total intracellular protein when the yeast grows on methanol. Interestingly, the expression of AOX is strongly repressed in the presence of preferred, energetically more efficient carbon sources such as glucose or glycerol. If methanol is the only carbon source available, then the alcohol oxidase expression is de-repressed and the enzyme is produced in high quantities (Gasser et al., 2013). The potential of the strong, tightly regulable AOX1 promoter for the expression of recombinant proteins did not go unnoticed (Cregg et al.,

1989), and soon became one of the key elements of common *P. pastoris* expression vectors.

The growth of *P. pastoris* in methanol as the sole carbon source has important bioprocess implications. Its consumption leads to high levels of heat generation, in part because of its higher enthalpy of chemical combustion compared to other carbon sources. This effect can be countered basically by an efficient reactor design (with efficient heat transfer capacity), adapting the feeding strategies and/or using mixed carbon sources as substrates (Meyer et al., 2008). For example, the combination of methanol and sorbitol during the induction phase can provide not only operational advantages, but improved product titers (Arnau et al. 2010, Jungo et al., 2007). Noticeable, sorbitol is one of the few non-repressible carbon sources assimilable by *P. pastoris* (Inan and Meagher, 2001).

1.3.4. Strains and phenotypes

As mentioned previously, the most extended *P. pastoris* strains for the production of recombinant proteins derive originally from the yeast isolate obtained by H. Phaff in California, US (Gonçalves et al., 2013). This wild-type parental strain, recently renamed as *Komagataella phaffi* (NRRLY-11430, also classified as CBS7435), was the source to generate GS115 (auxotrophic, *his4* phenotype). GS115 gave rise to X-33 (by reversion of the auxotrophy, arguably phenotypically equivalent to the wild-type) and KM71 (*his4 aox1*), in which AOX1 gene was disrupted. By disruption of the second AOX gene (AOX2), MC100-3 (*his4 aox1 aox2*) was generated. KM71 auxotrophic phenotype can easily be reverted by transforming any vector encoding for the HIS4 gene (used as a selection marker), obtaining KM71H (*aox1*). Returning to GS115, this strain was also used to generate protease deficient strains: SMD1163 (*his4 pep4 prb1*), SMD1165 (*his4 prb1*) and

SMD1168 (*his4 pep4*) (fig. 1.10). Thanks to the disruption of prb1 and pep4 (encoding for proteinase B and proteinase A, respectively), these strains represent an alternative for the expression of protease-sensitive proteins. However, they are usually recommended as the last resort because of their deficient growth behavior and lower transformation efficiencies (Lin-Cereghino and Lin-Cereghino, 2007).

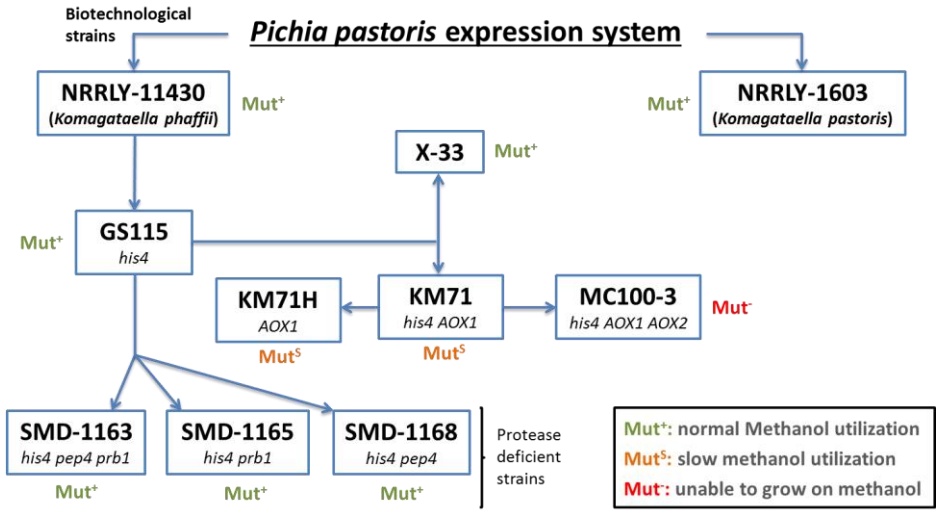


Fig. 1. 10. Genealogical scheme of *P. pastoris* strains.

P. pastoris strains are usually classified according to their methanol utilization profile. As this feature depends on the integrity of the two AOX genes of the yeast (AOX1 and AOX2), KM71 and KM71H have a Mut^S (methanol utilization slow) phenotype, while MC100-3 is a Mut $^-$ strain unable to use methanol as a carbon source. The rest of the mentioned strains are able to metabolize methanol normally.

GS115, X-33 and KM71 have been the best-seller biotechnological strains commercialized by Invitrogen Corp. for the last two decades, and the producers of the vast majority of reported heterologous proteins produced in the expression system. Nevertheless, additional strains have been generated (in some cases starting from the previous ones) and successfully used for protein expression (Felber et al., 2014). Furthermore, the genome sequences of the most extended strains were recently published (Küberl et al., 2011; Mattanovich et al., 2009; De

Schutter et al., 2009), a significant breakthrough in the system knowledge useful for rational gene editing and strain improvement.

1.3.5. Tools and methods for protein expression in *Pichia pastoris*

1.3.5.1. Integrative expression vectors

As in many other eukaryotic expression systems, the genes coding for the protein of interest are usually cloned into vectors containing a bacterial selection marker and replication origin. By doing this, the DNA sequences responsible for the gene transcription are easily manipulated and propagated in bacteria. For this reason, these vectors are usually referred as "shuttle vectors" due to its capacity to be functional in at least 2 different hosts (typically, in bacterial and eukaryotic systems). In the specific case of *P. pastoris*, a previous step of plasmid propagation in bacteria (*E. coli*) is essential to achieve the high DNA amounts required for the efficient generation of transformants (Jamshad and Darby, 2012).

Below are described in deeper detail the key components of *P. pastoris* shuttle vectors (fig. 1.11).

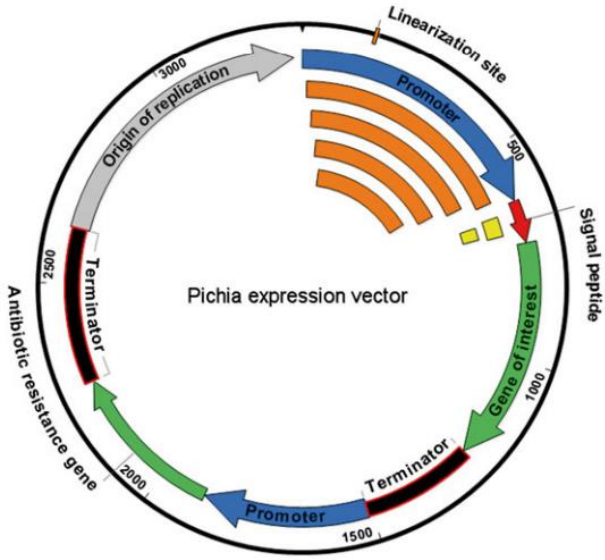


Fig. 1. 11. Schematic representation of a *P. pastoris* shuttle expression vector. (Source: Felber et al., 2014)

Promoters

AOX1 and GAP (methanol-inducible and constitutive, respectively) are the most extended promoters for protein expression. AOX1 is usually the first choice, as it allows a tight regulation simply depending on the carbon source available: while the promoter is strongly repressed by glycerol and specially glucose, it is strongly activated by methanol (Inan and Meagher, 2001). Its strength usually leads to high productivities, although promoter performance seems to depend also on the protein expressed. Recently, promoter libraries based on AOX1 (Hartner et al., 2008) and GAP (Qin et al., 2011) have been described, with some variants that led to significant increases in product titers compared with the original promoters. Moreover, the differential performance of the sequences generated were proposed as a useful tool to modulate the co-expression of accessory proteins (Ahmad et al., 2014). Other promoters have been shown to be efficient alternatives to express heterologous proteins in *Pichia* (Stadlmayr et al., 2010, Vogl and Glieder, 2013;

Çalik et al., 2015). Some of these avoid the use of methanol, a highly flammable and hazardous substance, as inducer in *P. pastoris* cultures.

Secretion signals

When the extracellular expression of the protein of interest is intended, the use of a secretion signal is required. In such case, vectors can already contain the appropriate coding sequence, and the gene of interest is cloned in frame accordingly. The α-mating factor from *S. cerevisiae* is normally the preferred choice, but several alternative sequences have been described for the successful secretion of heterologous proteins (Ahmad et al., 2014). For example, variants of the mentioned α-mating factor obtained by site-directed mutagenesis yielded higher expression titers than the original sequence (Lin-Cereghino et al., 2013). Again, the performance of different secretion signals is unpredictable because it depends on factors such as the mature protein N-terminal sequence (Yang et al., 2013) and structure (Cereghino et al., 2002). Therefore, whenever an expression optimization is needed, a screening of signal sequences is recommended. Furthermore, when expressing naturally secreted proteins the use of the native signal peptide cannot be completely discarded, as it could provide better performance than yeast-specific signals (Vadhana et al., 2013).

Selectable markers

Typical selectable markers can be classified between antibiotic resistances and auxotrophic markers. Usual antibiotics used for selection of *Pichia* transformants are Zeocin, G418, Hygromycin, Blasticidin and Nourseothricin, while *HIS4*, *ARG4*, *URA3* and *ADE2* are the typical auxotrophic markers (Cregg, 2007). Interestingly, the screening of transformants obtained by reconstitution of the

ADE2 gene allows the positive selection thanks to a change in the color of colonies (from red to white). Even though the use of auxotrophic markers avoid the generation of antibiotic-resistant strains, the use of auxotrophic strains is then strictly necessary for clone selection.

Other vector features

High amounts of plasmid are required for transformation of *Pichia*. For this reason, *E. coli* is used as the plasmid reservoir and producer, making necessary the presence of a bacterial replication origin (usually high-copy), as well as a bacterial selectable marker (typically, ampicillin or kanamycin). Even though Zeocin can be used both for bacterial and eukaryotic selection, the use of a specific bacterial marker is usually preferred.

1.3.5.2. Integration of expression vectors into the yeast genome

Thanks to the characteristic integration of expression vectors into the yeast chromosome, the clones isolated have a high stability and the expression performance is maintained over a large number of generations. Moreover, the presence of a stably integrated expression vector removes the need for antibiotics once the clone is properly isolated, which is of great interest at the industrial production level. To achieve an efficient integration event, the expression vector is usually linearized by restriction inside the promoter sequence (e.g. AOX1when the protein expression is methanol-induced). Then, the plasmid is introduced into the cells by electroporation, and the terminal regions of the linear DNA recombine specifically with the endogenous AOX1 promoter by homologous recombination. This process allow the insertion of the whole plasmid into the chromosome, and regenerates not only the truncated vector promoter but also the endogenous one

(responsible for the expression of AOX) (fig. 1.12). The process is analogous for any other promoter used.

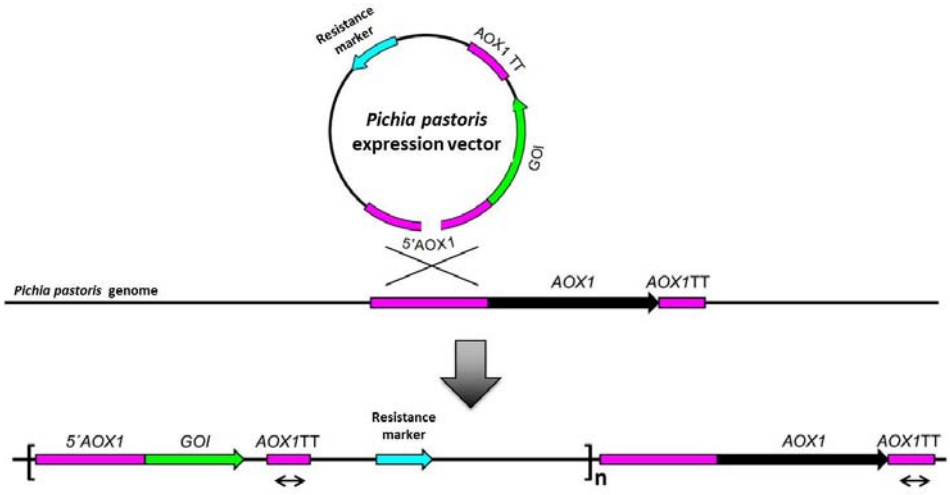


Fig. 1. 12. Depiction of the process leading to the vector recombination into the yeast chromosome (adapted from Nordén et al. 2011).

1.3.5.3. Multi-copy strains

The productivity of a given clone can be directly related to the number of copies of the gene of interest integrated into the yeast genome (Vassileva et al., 2001; Marx et al., 2009). Increasing the copy number can be achieved either by selecting clones where multiple integration events occurred (usually by means of selection in higher antibiotic concentrations, Sunga et al., 2008), or by transformation of multi-copy vectors (Vassileva et al., 2001; Zhu et al., 2009). However, several studies suggest that above a maximum amount of copies a detrimental effect can be experienced in the productivity, especially for secreted proteins (Hohenblum et al., 2004; Marx et al., 2009; Zhu et al., 2009; Liu et al. 2014). Apparently, an excess of gene copies can overburden the host cell physiology by overloading the secretory capacity (Gasser et al., 2013). Moreover, it has been shown that it can potentially lead to

strain instability issues that can seriously affect productivity in bioreactor cultures, presumably due to loop out recombination events between similar sequences closely localized into the genome (Aw and Polizzi, 2013). Interestingly, the optimum number for a given gene would depend on many factors, such as gene length or protein complexity. In any case, this property of *P. pastoris* provides an additional strategy to improve productivity levels.

1.3.5.4. Episomal expression vectors

Episomal (non-integrative) vectors consist an interesting tool for protein expression in *P. pastoris*. By including in the shuttle vector a *P. pastoris* autonomous replicating sequence (PARS) (fig. 1.13), the circular DNA can be transformed into yeast cells and by maintained for a limited number of generations. This feature can be used for transient expression of proteins or the screening of gene libraries (Lee et al., 2005).

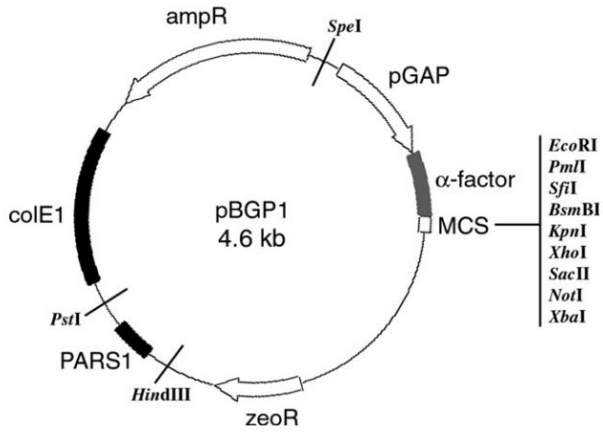


Fig. 1. 13. Map of a *P. pastoris* episomal plasmid, containing the PARS1 (*P. pastoris* autonomous replicating sequence). (Source: Lee et al., 2005).

One of the problems of *P. pastoris* episomal vectors currently available is their low stability over yeast generations, which is in contrast with *S. cerevisiae* stable

episomal vectors (e.g. pRS vectors with 2μ replication origin). Although this can be an advantage for certain uses, as will be discussed in this work, it limits its usefulness for the stable expression of heterologous proteins over a significant number of cell divisions. However, novel autonomous replication sequences have been developed for its efficient use in *Pichia* (Liackko and Dunham, 2014; Liackko et al., 2014), a significant achievement that presumably will expand the use of episomal vectors in the expression system.

1.3.5.5. Protein glycosylation in *P. pastoris* and strain glycoengineering

Protein glycosylation is one of the main post-translational modifications (PTMs) of eukaryotic proteins. The addition of glycans is not only important for proper folding, but it can have a key role in protein pharmacokinetic and pharmacodynamic properties too (Laukens et al., 2015). Two major glycan profiles are present in glycoproteins: N-glycosylation, which is achieved by covalent attachment of an oligosaccharide to certain asparagine (N) residues within the consensus sequence N-X-S/T, and the less common O-glycosylation, taking place on Ser/Thr residues. Glycosylation occurs in the secretory pathway, from the ER to the late Golgi compartments, and is driven by the enzymatic activity of several glycosyltransferases and glycosidases specifically localized throughout the pathway (Varki et al., 2009). Although the first steps of protein glycosylation taking place in the ER are highly conserved between species of the eukaryotic kingdoms, their final glycan structures can be markedly different basically due to the following enzymatic steps in the Golgi (fig. 1.14).

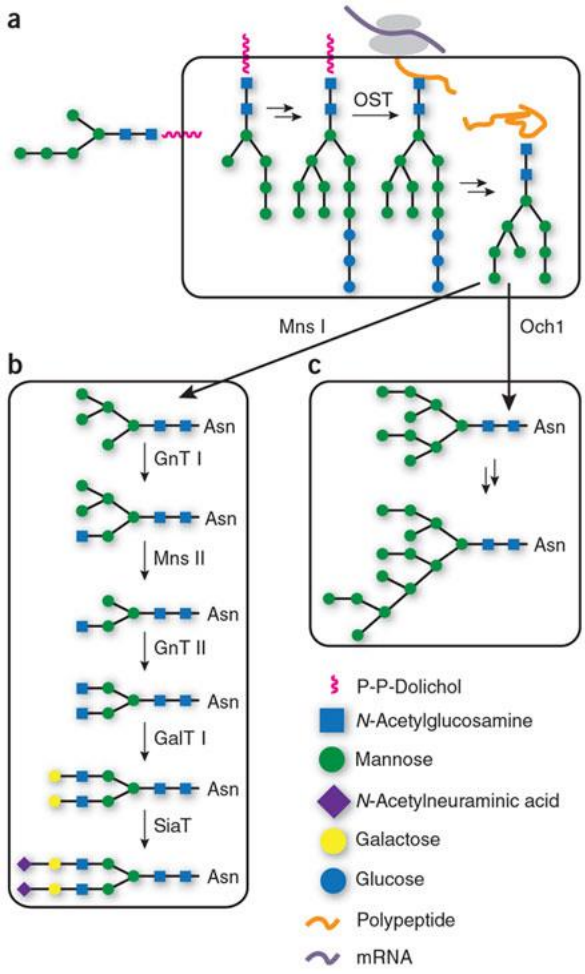


Fig. 1. 14. Overview of human and mammalian N-glycosylation pathways. OST, oligosaccharyltransferase; Mns, mannosidase; Och1, α(1-6)mannosyltransferase; GnT, N-acetylglucosaminyltransferase; GalT, galactosyltransferase; SiaT, sialyltransferase. (Source: Rich and Withers, 2009).

Mammalian glycosylation patterns are characterized for its complex composition of sugars, terminated with sialic acid (aka N-Acetylneuraminic acid) residues. On the other hand, yeasts naturally produce simpler glycans known as "high-mannose", composed only by two core N-acetylglucosamine residues and ramified mannoses (Rich and Withers, 2009). Although both types of glycosylation usually lead to bioactive proteins in which this PTM is essential, the differences between

them have a critical impact on their usability as therapeutic entities: while mammalian complex glycosylation is compatible with therapeutic glycoproteins, high-mannose glycans can entail very low plasma half-life (due to fast clearance by the reticuloendothelial system from the liver) and, what is more worrying, immunological response issues.

As a starting point, *P. pastoris* seems to be a better candidate compared to *S. cerevisiae*. While the later tends to hyper-glycosylate proteins (50-150 mannose residues), glycosylation is less extensive in *P. pastoris* (up to 20). Hyper-glycosylation can potentially lead to severe drawbacks, like hindering proper protein folding (Valero, 2012) or masking antigen epitopes (Liu et al., 2009). Moreover, *P. pastoris* lacks the α-1,3-linked mannosyl transferase responsible for the highly antigenic α-1,3-linked mannosyl terminal linkages taking place in *S. cerevisiae* (Demain and Vaishnav, 2009). As a result, *P. pastoris* was the selected yeast by many researchers, and has been submitted to intensive strain engineering to humanize both its N- and O- glycosylation profiles aiming to qualify the expression system for the production of therapeutic products. Nowadays, the number of proteins expressed in *P. pastoris* is much higher than in *S. cerevisie*, despite its importance as a host for the industrial production of biopharmaceuticals (Bill, 2014). Interestingly, the oligo-mannose glycosylation pattern represents a good starting point to achieve M6P glycans due to the structural similarity.

For the humanization of N-glycans, a set of heterologous glycosidases and glycotransferases had to be co-expressed. This could be accomplished by fusing N-terminal yeast-specific localization domains to luminal (Golgi-oriented) catalytic domains from species ranging from *C. elegans* to *H. sapiens*. A combinatorial library was used to determine the best combinations for each catalytic step (Nett et al., 2011; Wildt and Gerngross, 2005). The terminal addition of sialic acid residues required a special effort, since a whole new synthetic pathway consisting on 6 enzymes and 1 transporter had to be enabled in the yeast (Hamilton et al., 2007). Additionally, the deletion of *OCH1*, an endogenous gene encoding for the α-1,6-

mannosyltransferase responsible for the mannose structure elongation, had to be achieved (Choi et al., 2003; Vervecken et al., 2004). In order to obtain a fully humanized profile, additional gene knockouts (*Δpno1*, *Δmnn4B* and *Δbmt2*) were added (Li et al., 2006; Hamilton et al., 2006). Finally, the overexpression of a UDP-GlcNAc transporter was required to provide sufficient amounts of substrate for one of the glycosyltransferases (Choi et al., 2003). Regarding O-glycosylation profile, a humanized profile was successfully achieved too. By co-expression of a mannosidase and a set of glycosyltransferases, a TNFR2:Fc fusion protein containing mucin-type glycans (a mammalian O-linked glycosylation) could be expressed and characterized (Hamilton et al., 2013). Alternatively, the method for the complete O-linked mannose removal was recently described (Hopkins et al., 2015). All in all, the humanization of *P. pastoris* glycosylation profiles has been proved, which suggests that alternative human glycan profiles like the M6P could be also a reality.

Strains with humanized glycosylation pattern have been successfully used for the expression of proteins of therapeutic interest, such as erythropoietin (EPO, Hamilton et al., 2006), full-length antibodies (Li et al., 2006; Potgieter et al., 2009; Berdichevsky et al., 2011) or granulocyte macrophage-colony stimulating factor (GM-CSF, Jacobs et al., 2010). Surprisingly, no proteins produced with these strains can be found on the market to date. Apparently, the progressive productivity improvement for monoclonal antibodies in mammalian systems, together with the deficient growth and poor stability of such altered *P. pastoris* strains, could compromise the profitability and scalability of production processes. However, recent achievements in the improvement of humanized strains fitness (Jiang et al., 2015) could finally lever up their potential for the production of therapeutic proteins.

1.3.5.6. Cultivation techniques

Pichia pastoris is able to grow in a wide pH range (3.0 – 7.0) without significant impact in its growth profile. Therefore, this property allows to adjust this parameter according to the requirements of the protein of interest in terms of stability, productivity and protease-sensitivity (Macauley-Patrick et al., 2005). Moreover, the yeast has been cultivated in a great range of temperatures for protein expression. Even though the optimal temperature for cell growth is 28 – 30ºC (Gonçalves et al., 2013), lowering the temperature can be helpful to improve protein stability and decrease the chances of experimenting proteolytic activity over the product of interest.

Fed-batch is the most extended operational strategy for the expression of heterologous proteins in high cell density cultures of *Pichia pastoris*. The main reason is that it takes advantage of the yeast capacity to grow at very high biomass concentrations, which increases notably the volumetric productivities (Meyer et al., 2007). For cultures expressing the protein of interest under the control of AOX promoter, which is the most common option, methanol-limited fed-batch (MLFB) is the standard strategy for bioreactor cultivation. Methods have been described in detail (Tolner et al., 2006; Invitrogen Pichia fermentation guidelines, available at www.thermofisher.com), arguably contributing to the popularization of the expression system. Technically, these cultures are divided in three main stages. The two first phases (a batch phase followed by an intermediate fed-batch) are aimed to reach high cell densities by using an efficient carbon source such as glycerol or glucose. The third phase is dedicated to protein expression by switching to methanol as the sole C-source (Jahic et al., 2006). Methanol feed rate is limiting in MLFB, in order to control cell growth and decrease the main drawbacks associated to this strategy: high levels heat generation and oxygen demand. However, and as mentioned earlier in this chapter, the use of mixed-feeds (such as methanol in

combination with sorbitol) was shown to be an efficient measure to reduce significantly (40-50%) both factors (Jungo et al., 2007).

O_2-limited fed-batch (OLFB) strategy has gained interest over the last few years because this methodology helps to reduce heat generation and oxygen transfer limitations typical of MLFB. Apart from the operational advantages, the methodology can lead to better productivities and reduced release of endogenous proteins into the medium (Charoenrat et al., 2005). Moreover, a robust and scalable OLFB method was developed for the production of humanized monoclonal antibodies in glycoengineered strains. Compared to MLFB, the new process showed increased productivity, better product integrity and improved N-glycan composition (Berdichevsky et al., 2011).

Temperature-limited fed-batch (TLFB) is an alternative, less extended strategy that has been applied with success to reduce cell death and associated proteolytic activity. With temperatures as low as 12ºC, the degradation of a fusion protein was drastically decreased (Jahic et al., 2003). However, the strategy did not show significant advantages in the production of recombinant *R. oryzae* lipase (Surribas et al., 2007), and its use may be restricted to highly protease-sensitive proteins.

Regarding *P. pastoris* strains expressing proteins under the constitutive GAP promoter, limitation by controlled addition of the carbon source (usually glucose or glycerol) is the typical methodology followed in fed-batch cultures (analogous to MLFB) (Garcia-Ortega et al., 2013). GAP promoter has been used also under O_2-limiting conditions for the production of an antibody fragment. Remarkably, a 2.5-fold increase was seen in the specific productivity for this protein under these conditions (Baumann et al., 2008).

In the present work, the most described and widespread strategy, methanol-limited fed-batch strategy (MLFB) was compared to the novel and presumably advantageous O_2-limited fed-batch strategy (OLFB) for the production of α-galactosidase A in high cell density cultures in bioreactor.

1.4. *Pichia pastoris* for the production of biopharmaceuticals: current situation

The benefits of *P. pastoris* and its potential for the production of therapeutic proteins has not gone unnoticed for the biopharmaceutical industry. Yeast-based production systems have been largely known since the late 80s and early 90s, when the first recombinant insulin produced in *S. cerevisiae* (Novolin®, by Novo Nordisk) was approved for the treatment of diabetes. Since then, numerous versions of new-generation insulins have been developed and introduced to the market, and several protein-based pharmaceuticals produced in *S. cerevisiae* have been approved too (Walsh, 2014). Surprisingly, and presumably due to the incremental knowledge of *P. pastoris* as a protein production host gained over the last two decades, it was not until 2009 that the first protein-derived product expressed in this system was approved by the FDA: Ecallantide (brand name Kallbitor®, from Dyax), an inhibitor peptide of plasma kallikrein used for the treatment of hereditary angioedema (Thomson, 2010). This milestone opened the door to the approval by regulatory agencies of future therapeutic proteins produced in this expression system. Indeed, in 2012 Ocriplasmin (Jetrea®, from ThromboGenics) became the 2nd authorized product expressed in *P. pastoris*. Ocriplasmin is a truncated form of human plasmin used to treat vitreomacular adhesion (Syed and Dhillon, 2013).

Several therapeutic agents are being currently developed for its production in *P. pastoris*, including vaccines, interferons, albumins, hormones and antibodies for the treatment of a wide range of diseases, in line with the progressive interest gained by the system for the production of biopharmaceuticals (Vogl et al., 2013). Three cases are presented below as examples of the great commercial interest of *P. pastoris* for the production of therapeutic entities, partly justifying its election as an expression system for the production of lysosomal enzymes.

Case 1 - Production of full-length antibodies

Alder Biopharmaceuticals Inc. (Bothell, WA, USA) is focused on the production of therapeutic full-length monoclonal antibodies in *P. pastoris*. The antibodies are specifically engineered to eliminate glycosylation sites, improving product half-life in the blood-stream and minimizing potential immunological responses. Currently, two antibodies are undergoing clinical trials: one for migraine prevention (Dodick et al., 2014) and another for the treatment of rheumatoid arthritis and psoriatic arthritis. A third antibody aimed to treat Cushing's disease and Congenital Adrenal Hyperplasia is still at the pre-clinical stage (www.alderbio.com).

Case 2 - Production of antibody fragments

Ablynx NV (Ghent, Belgium) has developed a technology platform for the production of therapeutic nanobodies®, the smallest known antigen-binding antibody fragments. Interestingly, nanobodies are not obtained from conventional full-length antibodies consisting of light and heavy chains, but from heavy-chain only antibodies occurring naturally in *camelidae* (e.g. camels and llamas). Each one of the two heavy chains contains a single variable domain (V_{HH} or Nanobody®) (fig. 1.15), with outstanding biophysical, biochemical and pharmacological properties (Huang et al., 2010), Therefore, nanobodies have a great potential as therapeutic entities. Their small size makes them suitable to treat challenging targets, to be delivered through alternative routes (such as inhalation, oral or ocular), or to be combined to obtain multivalent specificities, among others. Although its size could lead to decreased plasma half-life by renal clearance, such effect can be avoided by multimerization of nanobodies or by targeting them to the circulating serum albumin, which acts as a carrier and improves their half-life up to 2-3 weeks (Van Roy et al., 2015).

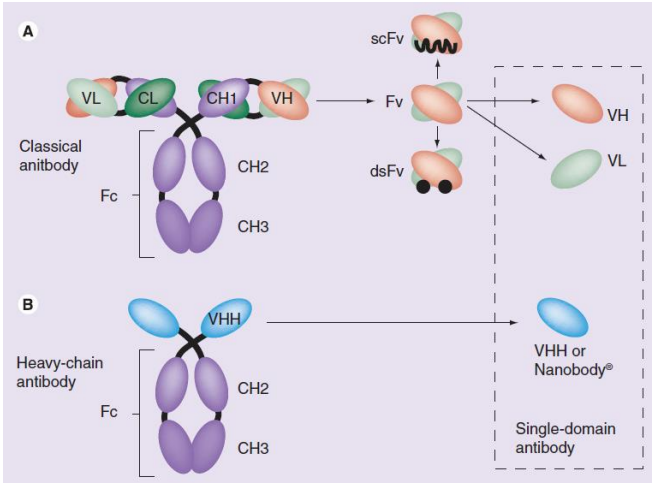


Fig. 1. 15. Schematic representation and comparison between conventional and *camelidae*-derived antibodies, the source of Nanobodies®. (Source: Huang et al., 2010).

With several products on the clinical trial stage, some of them partnered with the main biopharma companies (Boehringer Ingelheim, Genzyme, Merck & Co., Merck KGaA and Novartis, between others), Ablynx provides a great example of the potential of *P. pastoris* as a system for the production or therapeutic proteins.

Case 3 - Production of human insulin and Hepatitis B vaccine in India

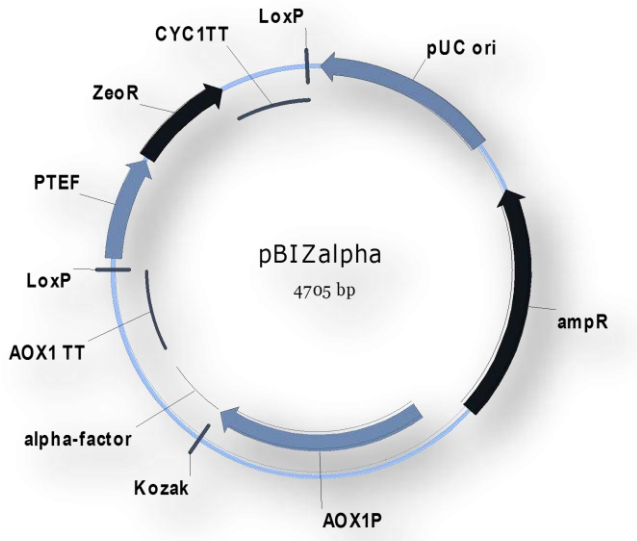
Using *P. pastoris* as expression system, Biocon (Bangalore, India) is developing the production of fast- and long-acting insulin biosimilars (rh-Insulin, Glargine, Lispro and Aspart), as well as a novel insulin conjugate suitable for oral administration. In fact, Insugen® (rh-Insulin) and Basalog® (Glargine) have been already available for diabetes type 2 treatment in India since 2004 and 2009, respectively (Owens et al., 2012). Historically produced in *E. coli* and later in *S. cerevisiae*, recombinant insulin has been a blockbuster in the biotech pharmaceutical industry for the last three decades. Biocon is taking advantage of

recent patent expiries and the numerous benefits of *P. pastoris* over the mentioned expression hosts to bounce into the market of insulins.

A final example of *P. pastoris* success is provided by Shantha Biotechnics (Medchal, India; a Sanofi-owned company). The company is on its own merits a household name in the biopharmaceutical industry of India, an emerging biotechnological market. In 1997 it launched Shanvac-B, the first recombinant health product in India and, remarkably, what is recognized as the first ever therapeutic product produced in *P. pastoris*. Shanvac-B is a vaccine for the immunization against the Hepatitis B virus, consisting of the recombinant surface antigen (HBsAg) expressed by the yeast (Chakma et al., 2011).

Chapter 2

Cloning and expression of human α-galactosidase A in *Pichia pastoris*



C h a p t e r 2

CLONING AND EXPRESSION OF HUMAN α-GALACTOSIDASE A IN *PICHIA PASTORIS*

2.1. Introduction

The cloning procedures and expression experiments to optimize the expression in *P. pastoris* of recombinant human α-galactosidase A (rhGLA) are detailed in this chapter. The first objective was to evaluate the suitability of the expression system for the production of the enzyme, in order to stablish this lysosomal hydrolase as the model protein for the study of humanized strains generated in future phases of the project. Keeping in mind that the final intention is to develop a yeast-based system for the production of ERTs, the efforts were focused from the beginning to maximize productivity titers.

Even though a great number of promoters have been described for the expression of heterologous proteins in *P. pastoris* (Vogl and Glieder, 2013), AOX1 promoter was chosen as the predetermined promoter to express rhGLA. Because AOX1 is a strong promoter tightly regulated by carbon sources, it allows the precise expression induction of the protein of interest. Moreover, the expression of rhGLA in *P. pastoris* under AOX1 promoter has been reported previously in the literature (Chen et al., 2000b).

In first place, a comparison between two of the main biotechnological *P. pastoris* strains was carried out to select the best one as the starting point. Specifically, Mut^+ (normal methanol utilization phenotype) and Mut^S (slow methanol utilization phenotype) strains were tested for the expression of rhGLA. Additionally, the co-expression of helper proteins was also explored. HAC1p (Guerfal et al., 2010) and PDI (protein disulfide isomerase) (Inan et al., 2006) were selected for the study because of their ability to increase productivity when co-expressed with the protein

of interest. As a final strategy to increase rhGLA titers, the effect of multiple GLA gene copies integrated in the yeast DNA was assessed. For this purpose, two strategies were followed: clones isolated in high Zeocin concentrations or directed generation of multi-copy clones by transformation of vectors with multiple GLA gene copies were the strategies evaluated.

The cloning and expression of different rhGLA variants is also described in the present chapter. Modifications were applied with different purposes. Firstly, the deletion of C-terminal amino acids (KDLL sequence) was evaluated aiming to increase its specific catalytic activity (Miyamura et al., 1996). Secondly, a purification tag (6xHistidine) was fused to the protein sequence. Alternatively, the RGD internalization motif was also fused to obtain a version of the enzyme potentially bioactive (Temming et al., 2005). In both cases, fusion of sequences were attempted both at the N-terminus or the C-terminus of the protein in order to define the best choice.

The last section of the chapter is dedicated to obtain a base strain expressing rhGLA as the model lysosomal enzyme. Such strain would be used in the future as a base for the co-expression of enzymatic complexes intended for the humanization of the yeast glycosylation profile. In first place, the productivity of a tagged version of the protein was optimized by co-expression of HAC1p. Additionally, the expression of the same enzyme variant was shown to be feasible under the constitutive GAP promoter. Last, a method was set-up for the efficient resistant phenotype reversion, allowing the co-expression of additional proteins by selection with any available resistance marker.

2.2. Materials and methods

2.2.1. Strains and culture media

Escherichia coli XL1-blue was used for the cloning steps and as a plasmid reservoir. *P. pastoris* strains NRRLY-11430 (CBS7435) and KM71 were used as base strains for the expression of rhGLA. The auxotrophic phenotype of KM71 (*his4*) was firstly reverted by transformation of an empty pPIC9 vector (linearized with SacI) and selection in minimal media agar plates, generating KM71H. To obtain strains expressing HAC1p, both NRRLY-11430 and KM71 were transformed with pPIC3.5K_HAC1 (in the case of KM71, the *his4* auxotrophy was reverted by the vector itself) and isolated in YPD-G418 (200µg/mL) agar plates (fig. 2.1).

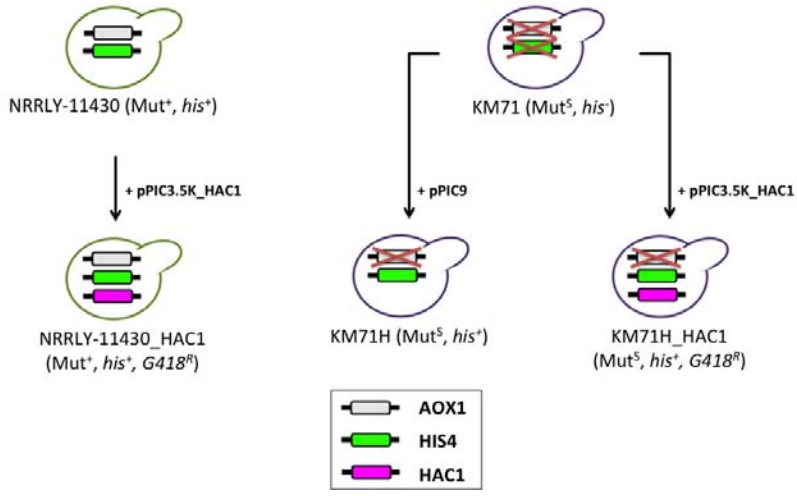


Fig. 2. 1. Generation of *P. pastoris* base strains for the expression of rhGLA.

Culture media was prepared as follows: LB (1% Tryptone, 0.5% Yeast Extract, 1% NaCl, pH 7.0; with 1% agar for plates), YPD (1% yeast extract, 2% peptone, 2% dextrose, with 2% agar for plates), BMGY (1% yeast extract, 2% peptone, 100 mM potassium phosphate, 1.34% YNB without amino acids, 4×10^{-5}% biotin, 1% glycerol, pH 6.0), BMMY (1% yeast extract, 2% peptone, 100 mM potassium

phosphate, 1.34% YNB without amino acids, 4×10^{-5}% biotin, 0.5% methanol, pH 6.0), and MD agar plates (1.34% YNB, 4×10^{-5}% biotin, 2% dextrose, 2% agar). All reagents for media preparation were purchased from Sigma-Aldrich and Panreac AppliChem. When necessary, antibiotics were added either to liquid media or agar plates: 100 µg/mL of ampicillin (Sigma-Aldrich, #A9518), 100 µg/mL of Zeocin (Invivogen, #ant-zn-1), 200 µg/mL of G418 in agar plates or 100 µg/mL in liquid culture (Sigma-Aldrich, #A1720), 200 µg/mL of Hygromycin B (in agar plates or 100 µg/mL in liquid culture Invivogen, #ant-hg-1).

2.2.2. Cloning of α-galactosidase A into a novel *Pichia pastoris* shuttle vector: pBIZα

A new vector was designed and synthetized *de novo* for the cloning and expression of rhGLA (fig. 2.2). It was conceived as a modular plasmid with interchangeable promoters, resistance markers and signal peptides by enzyme restriction.

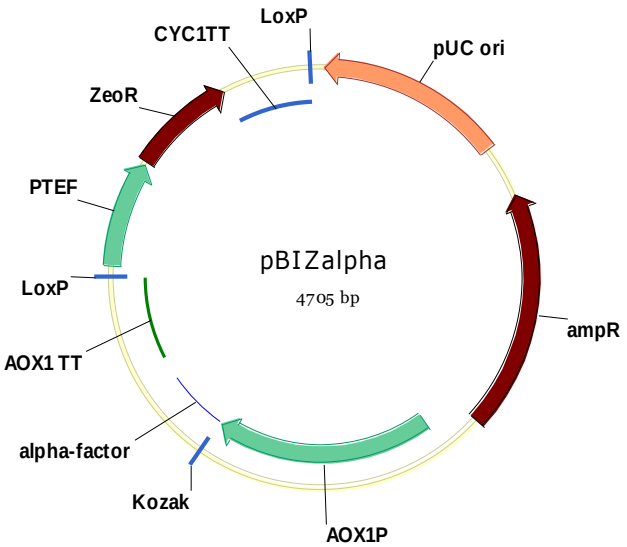


Fig. 2. 2. Bioingenium plasmid pBIZα, designed for the extracellular expression of rhGLA. The vector contains the main elements for protein expression in *P. pastoris*: AOX promoter, α-mating factor for protein secretion, multi-cloning site, AOX terminator, Zeocin marker for yeast selection (loxP-flanked for Cre recombinase-

mediated resistant phenotype reversion), pUC bacterial replication origin, and ampicillin marker for bacterial selection.

Two GLA genes were codon-optimized for *P. pastoris* and synthesized by GeneArt: one containing the RGD-motif coding sequence at 3' (pMK_GLACRGD) and the other with the motif at 5' (pMK_NRGDGLA). The genes coding for the different rhGLA variants (rhGLA, rhGLAΔKDLL, $rhGLA_{C\text{-}HIS}$, , $_{N\text{-}HIS}$rhGLA, $rhGLA_{C\text{-}RGD}$ and $_{N\text{-}RGD}$rhGLA) were generated by PCR with KOD Hot Start Master Mix polymerase (Merck Millipore, #71842) (primers listed in table 2.1, fig. 2.3). PCR products were column-purified (GE Life Sciences, #28-9034-71), digested with BsaI and re-purified. pBIZα was digested also with BsaI and purified. T4 DNA ligase (ThermoFisher, #15224-017) was used for the ligation of inserts with pBIZα (20ºC, overnight). $rhGLA_{C\text{-}HIS}$ was cloned in pBGZα too, under control of constitutive GAP promoter, using the same cloning procedure. Ligation products were purified and transformed by electroporation into *E. coli* XL1-blue electrocompetent cells (2.5 kV, 25 μF, 200 Ω, in a 2 mm electroporation cuvette). Then, 1 mL of LB was added directly into the cuvette and cells were incubated 1 h at 37ºC. Cell suspensions were then seed in LB+Ampicillin agar plates and incubated 18 – 24 h at 37 ºC. *E. coli* transformants were screened by plasmid restriction analysis and confirmed by sequencing. All restriction enzymes used were purchased from ThermoFisher)

2.2.3. Generation of rhGLA multi-copy vectors

P. pastoris expression vectors require a unique restriction site in the promoter sequence for plasmid linearization and transformation. For this reason, to generate multi-copy vectors the GLA gene had to be cloned previously in a new vector were one of the restriction sites (PmeI) is missing. Therefore, a new vector was generated with this feature (pBIZαΔPmeI), in which rhGLA was cloned (pBIZαΔPmeI_GLAwt) (fig. 2.4). Then, the whole expression cassette (AOX promoter – α-mating factor – rhGLA gene – AOX terminator) was amplified by

PCR with two different pairs of primers: the first pair including terminal NcoI sites (to generate the 2-copy vector from the single-copy one) and the second pair including terminal BglII sites (to generate the 3-copy vector from the 2-copy one, fig. 2.5) (primers listed in table 2.1). The ligation products were transformed into *E. coli* as mentioned previously.

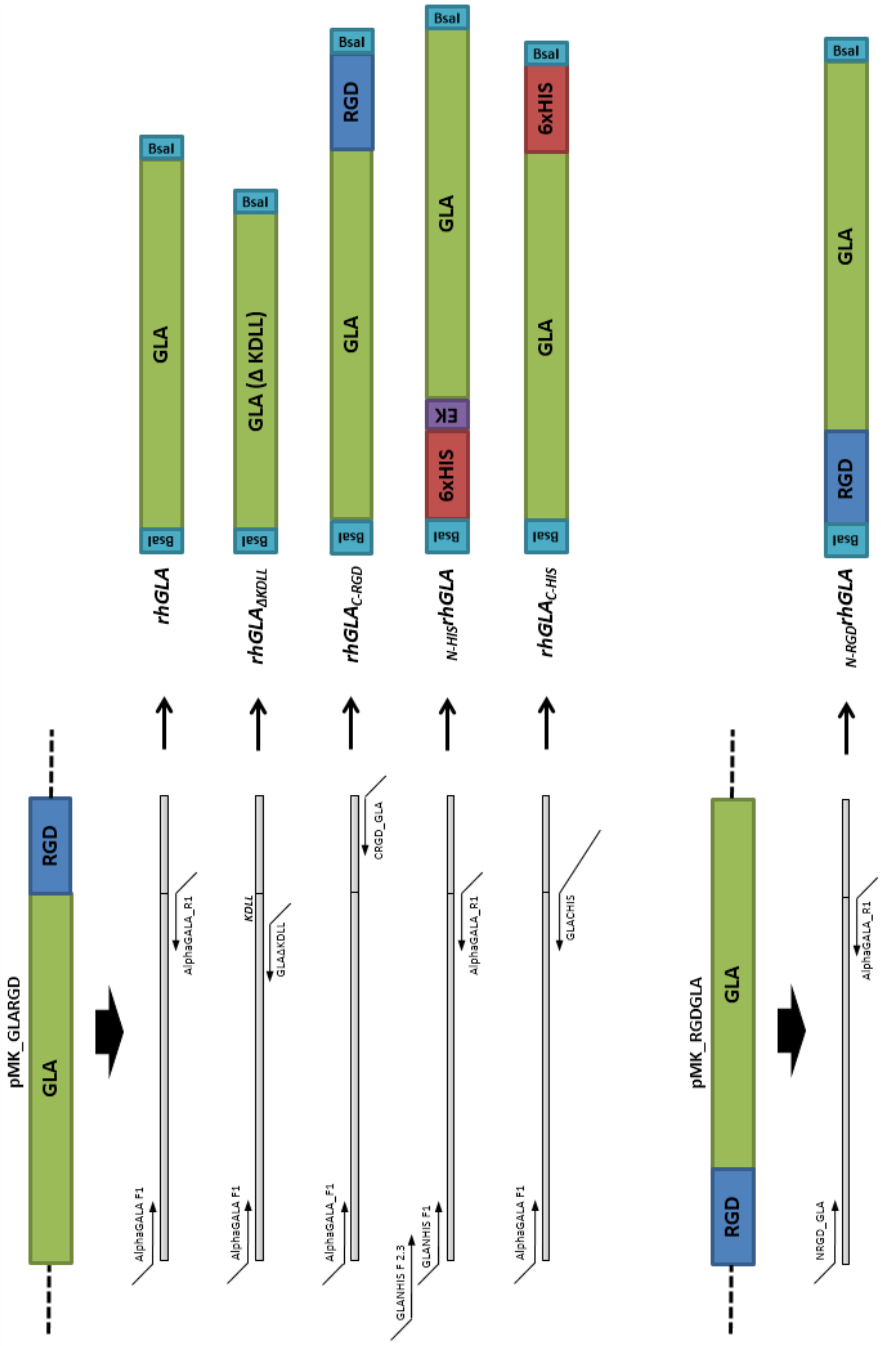


Fig. 2. 3. PCRs performed to generate rhGLA variants.

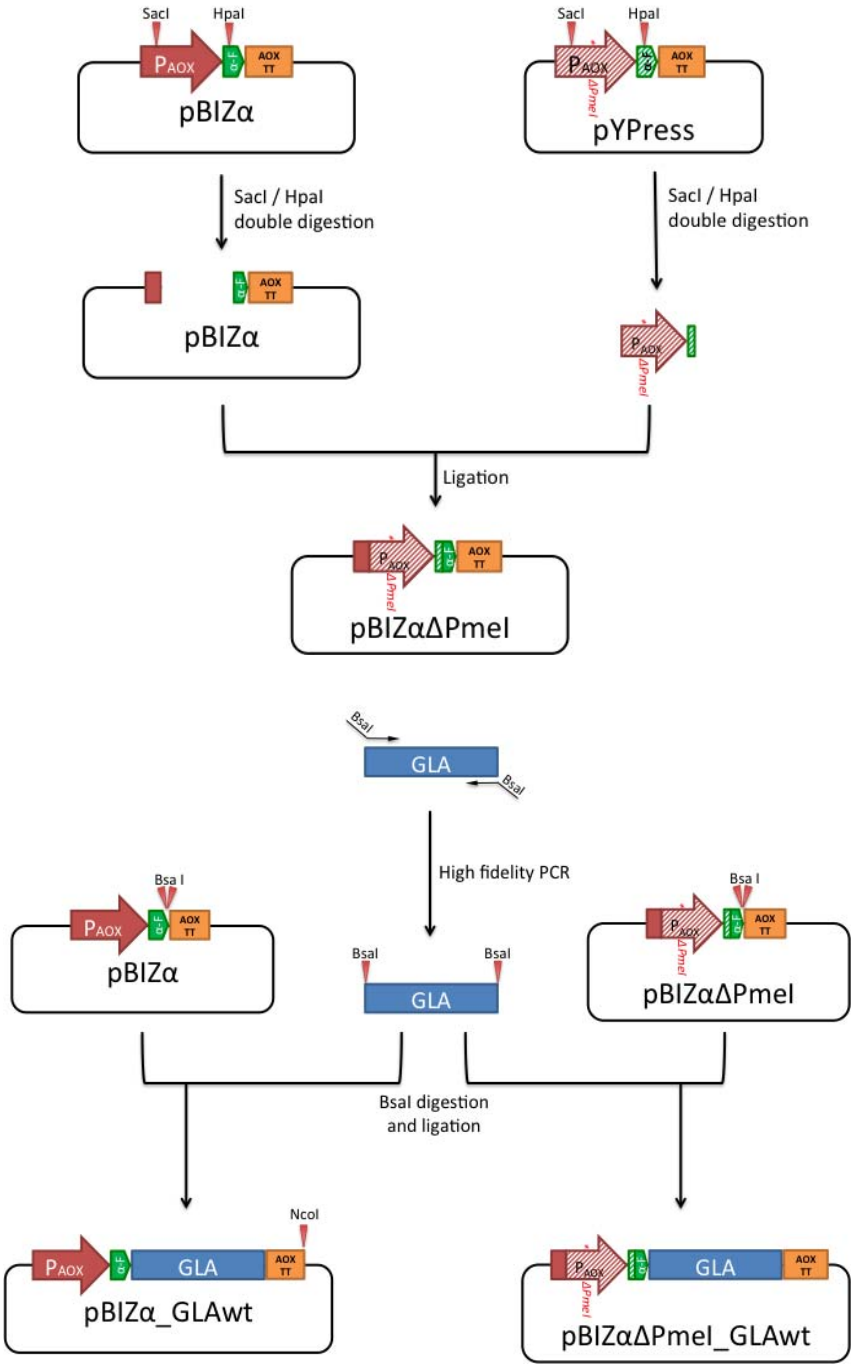


Fig. 2. 4. Depiction of the cloning procedure to obtain the single-copy rhGLA expression vector and the generation of an expression vector lacking the PmeI linearization site in its AOX promoter. GLA gene was cloned into this vector, which served as a PCR template for expression cassette amplification used for the generation of multi-copy vectors.

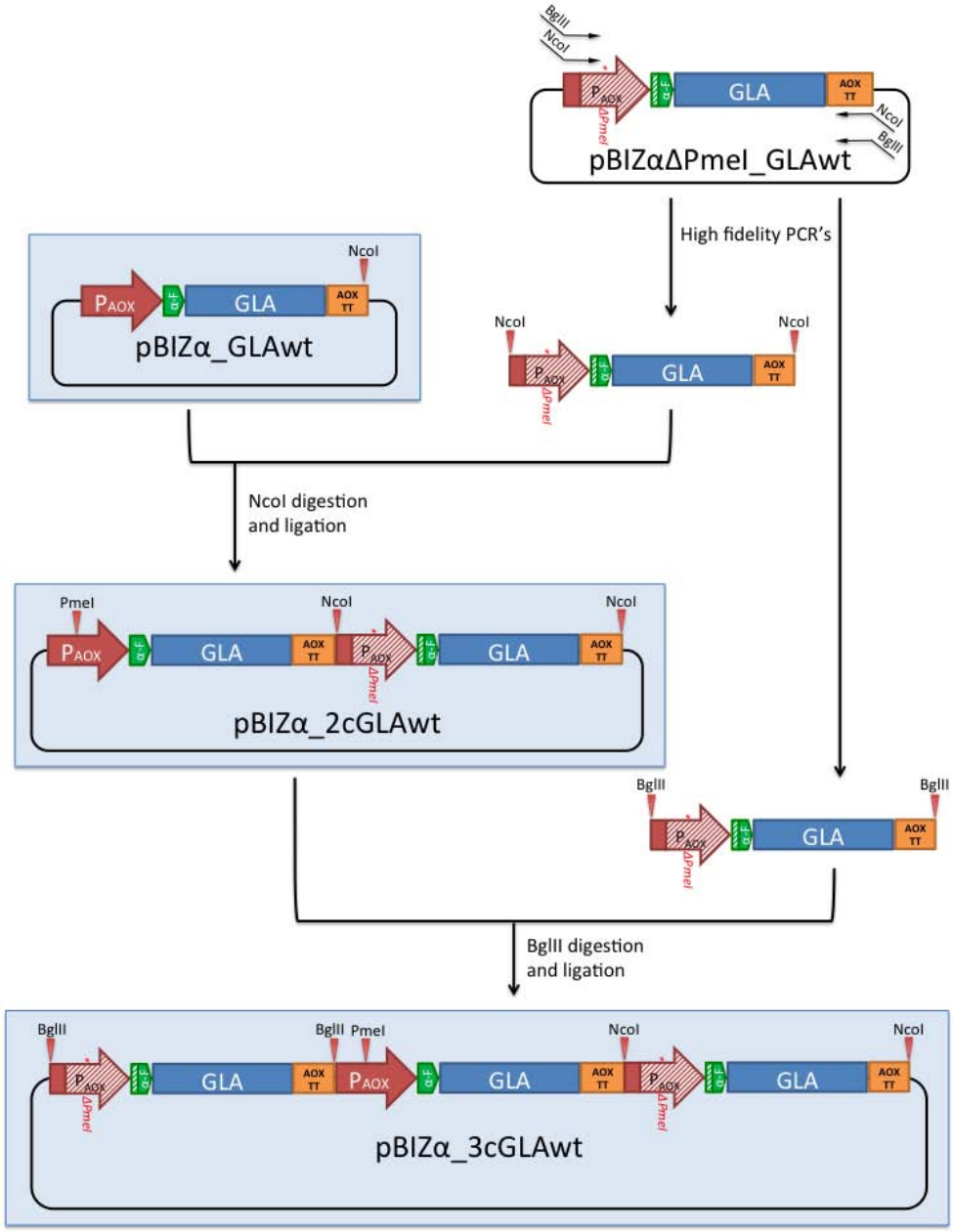


Fig. 2. 5. Depiction of cloning procedure to obtain the multi-copy rhGLA expression vectors. The GLA expression cassette was amplified by PCR from pBIZαΔPmeI_GLAwt and subcloned first into the single-copy pBIZα_GLAwt and later into the 2-copy pBIZα_2cGLAwt. Vectors transformed into *P. pastoris* are highlighted in blue shade.

Table 2. 1. List of primers used in this work.

Primer	Sequence (5' – 3')	Use
AlphaGALA_F1	ACTGGTCTCGGGCTTTGG ATAACGG	Forward, for amplification of GLA gene, and ligation into pBIZα or pBGZα.
GLANHIS_F1	TCACCATCACCACGACGA TGACGATAAGTTGGATA ACGGTTTGGCTAGAA	Forward, for amplification of GLA gene adding N-terminal 1/2 of 6xHistag + enterokinase cleavage site, and ligation into pBIZα.
GLANHIS F2.3	GATGGTCTCAGGCTCATC ATCACCATCACCACGA	Forward, for amplification of GLA gene adding N-terminal 2/2 of 6xHistag + enterokinase cleavage site, and ligation into pBIZα.
NRGD_GLA	ATCGGTCTCGGGCTACTA CTTACACTGCTTCTGCTA	Forward, for amplification of GLA gene including the N-terminal RGD motif coding sequence, and ligation into pBIZα.
AlphaGALA_R1	ATCGGTCTCGATTACAAC AAGTCCTTCAAGGACA	Reverse, for amplification of GLA gene, and ligation into pBIZα or pBGZα.
GLAΔKDLL	ATCGGTCTCGATTACAAG GACATCTGCATAGTG	Reverse, for amplification of GLA gene excluding the C-terminal KDLL coding sequence, and ligation into pBIZα.
CRGD_GLA	ATCGGTCTCCATTATGGC AAGTGTCTAGC	Reverse, for amplification of GLA gene including the C-terminal RGD motif coding sequence, and ligation into pBIZα.
GLACHIS	GATGGTCTCCATTAGTGG TGATGGTGATGATGCAAC AAGTCCTTCAAGGA	Reverse, for amplification of GLA gene adding C-terminal 6xHis-tag, and ligation into pBIZα.
GLAcass_NcoI_F	CTACCATGGCCAAAGAC GAAAGGTTGA	Forward, for amplification of GLA expression cassette, and ligation into pBIZα_GLAwt. Used also for the amplification of HAC1 expression cassette and ligation into pBIZα.
GLAcass_NcoI_R	CTACCATGGTACATTCTT CGCAGGCTTG	Reverse, for amplification of GLA expression cassette, and ligation into pBIZα_GLAwt. Used also for the amplification of HAC1 expression cassette and ligation into pBIZα.
GLAcass_BglII_F	CTAAGATCTCCAAAGAC GAAAGGTTGA	Forward, for amplification of GLA expression cassette, and ligation into pBIZα_2cGLAwt.
GLAcass_BglII_R	CTGAGATCTTACATTCTT CGCAGGCTTG	Reverse, for amplification of GLA expression cassette, and ligation into pBIZα_2cGLAwt.
HAC1_F (BsaI)	CTAGGTCTCGGATGCCCG TAGATTCTTC	Forward, for amplification of HAC1 gene, and ligation into pBIZiΔPmeI
HAC1_F (BsaI)	CGCGGTCTCGATTATTCC TGGAAGAATACAA	Reverse, for amplification of HAC1 gene, and ligation into pBIZiΔPmeI

Table 2.1. List of primers used in this work (ontinued).

Primer	Sequence (5' – 3')	Use
pBGP1 AOXTT_R (AflII)	CGAGCTTAAGGGATCCG CACAAAC	Reverse, for inverse PCR amplification of pBGP1_Cre plasmid excluding Zeocin resistance cassette.
pBGP1 ResistTT_F (NcoI)	GCTCCATGGTGCAAGCTA AGCTTCGACAAT	Forward, for inverse PCR amplification of pBGP1_Cre plasmid excluding Zeocin resistance cassette.
TEFcass_F (AflII)	AGTACTTAAGCTTGCCTC GTCCCCGCCGGGTCA	Forward, for amplification of Hygromycin resistance cassette and ligation into pBGP1_Cre (inverse PCR).
TEFcass_R (NcoI)	CTGCCATGGAGCTTGCAA ATTAAAGCCTTCGAGCGT CCCAAAA	Reverse, for amplification of Hygromycin resistance cassette and ligation into pBGP1_Cre (inverse PCR).

2.2.4. Generation of a *P. pastoris* expression vector containing HAC1 for inducible protein expression under AOX promoter

Apart from being cloned in the standard pBIZα expression vector, the $rhGLA_{C\text{-}HIS}$ variant was also inserted in a new vector containing the HAC1 gene under AOX promoter. The vector was generated by amplification of the HAC1 gene from pPIC3.5K_HAC1 and cloning into the vector pBIZiΔPmeI (generated in the company for other purposes) by BsaI restriction and ligation, following the procedures described above. The whole expression cassette of the generated vector (pBIZiΔPmeI_HAC1) was amplified by PCR and subcloned into pBIZα by NcoI restriction and ligation (fig. 2.6). The resulting vector, named pBIZαHAC1, was ready for cloning of the $rhGLA_{C\text{-}HIS}$ gene and any other gene of interest by BsaI restriction/ligation.

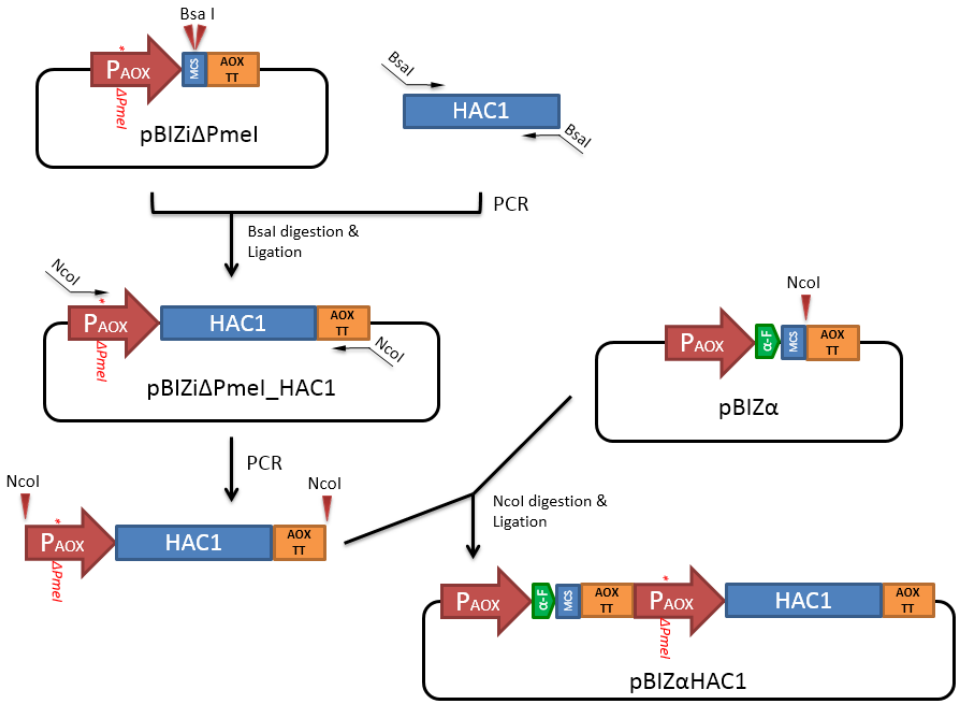


Fig. 2. 6. Depiction of cloning procedure to obtain pBIZαHAC1 expression vector, which includes the HAC1 expression cassette under AOX promoter. The generated vector was used to clone rhGLA$_{C\text{-}HIS}$ variant.

2.2.5. Transformation of vectors into *P. pastoris* and clone isolation

2 – 5 µg of shuttle vectors containing rhGLA variants were linearized with PmeI. For vectors encoding for HAC1p, different restriction enzymes were used depending on the promoter: SacI for pPIC3.5K_HAC1 and AvrII for pGCint_HAC1. Vector encoding for PDI (pPIC3.5K_PDI) was linearized with SacI, and for the pPIC9 empty vector used for HIS4 reconstitution of KM71 (to obtain KM71H), SalI was used. Digested vectors were transformed by electroporation into *P. pastoris* electrocompetent cells (2 kV, 25 µF, 200 Ω, in a 2 mm electroporation cuvette). Then, 1 mL of 1 M sorbitol was added directly into the cuvette and the cell suspension was transferred to a 50 mL tube, which was incubated 3 – 4 h at 30ºC, 200 rpm. Cell suspensions were then seed in

YPD+Zeocin or YPD+G418 agar plates, depending on the plasmid transformed and incubated 72 h at 30 ºC. *P. pastoris* transformants were isolated by re-streaking clones three times in agar plates containing antibiotic.

2.2.6. Screening of *Pichia pastoris* isolated clones: expression assays

Individual clones previously isolated on YPD+antibiotic agar plates were picked to inoculate 5 mL of BMGY supplemented with 100 µg/mL of Zeocin or G418 (depending on the vector transformed) in 50 mL centrifuge tubes. As *Pichia* cultures show high oxygen consumption rates, the caps were left partially opened to prevent oxygen depletion and ensure proper growth profiles. The cultures were incubated at 28ºC, 200 rpm, 45º tilt. After 24 h of growth phase on glycerol, OD_{600} was measured for each culture with a spectrophotometer (1240 UV, Shimadzu). With the aim to equalize the initial cell concentration at the beginning of the induction phase, the volume necessary to reach 1 OD_{600} unit in 10 mL was centrifuged at 3000 g, 3 min in 1.5 mL sterile microcentrifuge tubes. Supernatants were discarded and the pellets resuspended in 1 mL of BMMY. To initiate the induction phase on methanol, these cell suspensions were immediately added to 50 mL centrifuge tubes with vented cap (TubeSpin® Bioreactor 50, TPP #87050) pre-filled with 9 mL of BMMY, giving a final volume of 10 mL. The inoculated BMMY cultures were incubated at 25 ºC, 200 rpm, 45º tilt, and were re-induced every 24 h with 100 µL of pure methanol (1% v/v). After 72 h of induction, OD_{600} was measured and 1 mL of each culture was centrifuged at 6000 g, 3 min. Supernatants were stored at -20 ºC for further analysis.

2.2.7. Enzymatic activity assay

Activity of α-galactosidase A was measured using the fluorogenic water-soluble 4-methylumbelliferyl-α-D-galactopyranoside (MUG) (Sigma-Aldrich, #M7633). The substrate was dissolved in activity buffer (100 mM sodium acetate, pH 4.5) to a concentration of 2.46 mM. 5 µL of culture supernatant (or previously diluted culture supernatant) was dissolved in 20 µL of activity buffer in a 1.5 mL tube, and 100 µL of substrate solution was added to start the reaction. The incubation was performed at 37 ºC, 1300 rpm, 1 h, in an orbital thermoshaker. At the end of the incubation period, 1.25 mL of 0.2 M glycine buffer, pH 10.5, were added to each tube to stop the reactions. 200 µL of each reaction solution were added in a 96-well plate to measure the relative fluorescence of the samples. The plate was measured in a fluorometer (FL800, Biotek) with fixed excitation/emission wavelengths at 360/460 nm. The readings of the samples were compared to standards of 4-methylumbelliferone (500, 350, 200, 100, 50, 25 and 0 ng · mL^{-1}) (Sigma-Aldrich, #M1381), and the amount of hydrolyzed substrate was calculated. The concentration of alfa-galactosidase A is expressed in activity units: one unit of activity is equivalent to one nmol of substrate hydrolyzed per hour at 37 ºC.

2.2.8. Generation of episomal vector encoding for Cre recombinase and resistant-phenotype reversion procedure.

An episomal *P. pastoris* vector encoding for Cre recombinase under the constitutive GAP promoter was already available (pBGP1_Cre). However, it contained a Zeocin resistance marker, disqualifying it for the reversion of Zeocin-resistant strains. To change the resistance marker of the episomal vector, pBGP1_Cre was re-amplified by inverse PCR, excluding its Zeocin cassette and adding terminal AflII/NcoI sites. In parallel, a Hygromycin B resistance cassette (TEF promoter – Hygromycin B resistance gene – CYC terminator) was amplified by PCR from pBIHi, adding also terminal AflII/NcoI sites. A double digestion with

the mentioned restriction enzymes was performed independently for the PCR-amplified vector and the Hygromycin B resistance cassette. A standard cloning procedure was followed as explained previously, obtaining the desired pBGP1hygro_Cre (fig 2.7).

A selected clone over-expressing $rhGLA_{C\text{-}HIS}$ (α-galactosidase A with C-terminal 6xHis-tag) under AOX promoter was submitted to the reversion of its Zeocin-resistant phenotype. Taking advantage of the loxP sites flanking the resistance expression cassette, only the transient expression of Cre recombinase was required to catalyze its excision. Thus, the episomal pBGP1hygro_Cre vector encoding the Cre recombinase was transformed into electrocompetent cells of the mentioned strain according to the following procedure, which had to be previously optimized. 1 µg of non-linearized episomal vector was transformed, and 1 mL of 1 M sorbitol was added. The cell suspension was immediately transferred to a 50 mL centrifuge tube, and incubated 4 h at 28 ºC, 200 rpm. 200 µL were then used to inoculate 5 mL of YPD + Hygromycin (100 µg/mL), and the culture was incubated for 24 h at 28 ºC, 200 rpm. 100 µL of the previous culture were used to inoculate 10 mL of fresh YPD + Hygromycin, which were incubated for 24 h at 28 ºC, 200 rpm. Finally, 100 µL of the previous culture were used to inoculate 10 mL of fresh YPD without antibiotic. The culture was incubated for 24 additional hours at 28 ºC, 200 rpm. 100 µL of the previous culture were seeded in antibiotic-free YPD agar plates, which were incubated 48 h at 30 ºC. Finally, each isolated colony was re-streaked in an YPD, YPD + Zeocin and YPD + Hygromycin agar plates in order to confirm the correct resistant phenotype reversion.

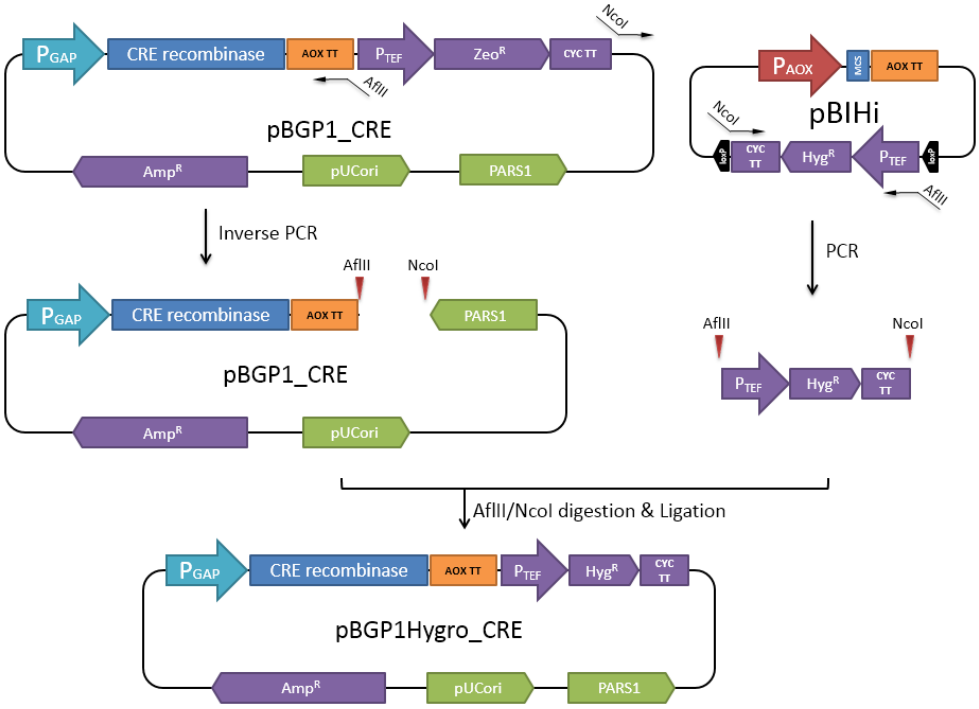


Fig. 2. 7. Cloning procedure followed to change the resistance marker cassette of the pBGP1_Cre episomal vector, from Zeocin to Hygromycin B.

2.3. Results and discussion

2.3.1. Comparison of expression levels of alpha-galactosidase A between Mut$^+$ and Mut^S strains

Based on their methanol utilization capability, *P. pastoris* strains can be divided in three different groups: Mut$^+$ (normal consumption), Mut^S (slow consumption) and Mut $^-$ (disabled methanol consumption), being the first two ones the most commonly used. As a first approach to optimize the MetOH-inducible expression of rhGLA, the performance of Mut$^+$ and Mut^S was explored. While Mut$^+$ is the wild-type phenotype, Mut^S exhibit improved productivities for certain proteins (Pla et al., 2006; Krainer et al., 2012). Moreover, the limited methanol assimilation of Mut^S strains is considered an advantage in bioreactor cultivations, as lower O_2 demand and heat generation are experienced (Arnau et al., 2011). Mut $^-$, on the other hand, is unable to consume methanol (which, in this case, is used only as inducer) and must be cultivated in presence of an alternative carbon sources. Its use is very limited compared with Mut$^+$ and Mut^S, and was not included as a third alternative. Thus, the selected strains for the initial study where NRRLY-11430 (also known as CBS7435, Mut$^+$) and KM71H (Mut^S) (fig 2.8).

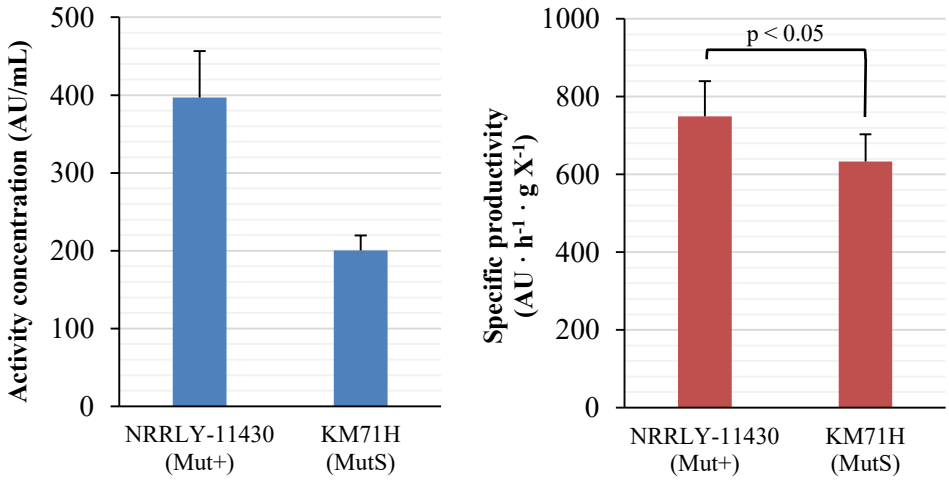


Fig. 2. 8. Comparison between rhGLA expression profiles of NRRLY-11430 (Mut$^+$) and KM71H (Mut^S) strains (6 clones per group). Left: activity concentration (AU/mL) after 72 h of induction. Right: average specific productivity (AU · h^{-1} · g DCW^{-1}).

The overall activity concentration reached in Mut^+ cultures was almost twice the measure obtained from Mut^S cultures (397 vs. 200 AU/mL). Nevertheless, Mut^S clones showed lower growth rates and, as a consequence, the measured final cell concentrations were notably lower, basically due to the reduced methanol consumption rate. For this reason, this factor had to be taken into account to compare properly the productivity performance of both phenotypes. Thereby, the comparison between their specific productivities (which is normalized according to the biomass concentration in culture) gave closer average values. Even though Mut^+ continued to be the best rhGLA producer (18% higher in terms of specific productivity), Mut^S was not discarded yet as a candidate for the following experiments.

2.3.2. Effect of the overexpression of helper proteins on the productivity of α-galactosidase A

P. pastoris is an expression system widely known for its competence for the production of secreted recombinant proteins. However, and for reasons often difficult to elucidate, some proteins face limiting factors that jeopardize their productivity. Several limitations can be found either at the transcriptional, translational or secretory level, especially for those complex proteins that require extensively the folding and post-translational cellular machinery. In order to countervail such limiting factors and boost the productivity, the overexpression of helper proteins is a strategy successfully used in *P. pastoris* (Ahmad et al., 2014). Two of the most extended, HAC1p and PDI (protein disulfide isomerase), were selected to assess their effect in the productivity of rhGLA. Both helper proteins were chosen for their contrasted success rate for the expression of different heterologous proteins in *P. pastoris* (Puxbaum et al., 2015; Inan et al., 2006).

2.3.2.1. Overexpression of HAC1p (methanol-induced) in Mut$^+$ and Mut^S strains

HAC1p is a transcriptional activator involved in the unfolded protein response (UPR) pathway. Under stress conditions in the ER, the splicing of HAC1 mRNA is induced allowing the expression of HAC1p, which contributes to the increase of ER-resident proteins levels, such as chaperones involved in protein folding and their trafficking through the ER (Kawahara et al., 1997). In this experiment, the methanol-inducible co-expression of the HAC1 transcription factor was evaluated both in Mut$^+$ and Mut^S strains aiming to determine its effect on the productivity of rhGLA. As seen for the expression of other proteins in *P. pastoris*, the over-expression of HAC1p led to a marked increase of rhGLA titers in the two strains tested (fig. 2.9). However, the effect was notably higher in the wild-type Mut$^+$ phenotype, where a 4.7 fold-change increment was seen with respect to the absence of HAC1p. Regarding the Mut^S strain, the increase was more restrained (1.4 fold-change) (table 2.2).

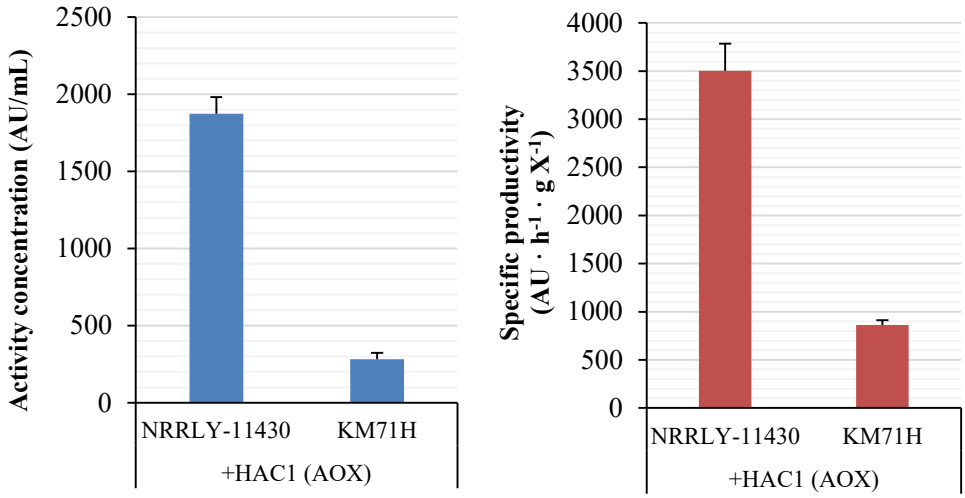


Fig. 2. 9. Comparison between rhGLA expression profiles of NRRLY-11430 (Mut$^+$) and KM71H (Mut^S) strains co-expressing HAC1p (methanol-induced). 6 clones were analyzed for each group. Left: activity concentration (AU/mL) reached after 72 h of induction. Right: average specific productivity (AU · h-1 · g DCW-1) during the

Table 2. 2. rhGLA specific productivities calculated for the co-expression of HAC1p in Mut^+ and Mut^S strains, and their respective increase compared with the productivity in absence of HAC1p over-expression.

	NRRLY-11430 (Mut^+) (+$HAC1_{AOX}$)	**Fold-change (vs. NO HAC1)**	**KM71H (Mut^S) (+$HAC1_{AOX}$)**	**Fold-change (vs. NO HAC1)**
Specific productivity (AU g X^{-1} h^{-1})	3503	4.7 x	863	1.4 x

The reasons why the inducible co-expression of HAC1p (and in fact, of rhGLA alone as seen in the previous paragraph) is more efficient in the Mut^+ phenotype remains unclear. In principle, the only genetic difference between the NRRLY-11430 (Mut^+) and KM71H (Mut^S) strains is the disruption of the AOX1 gene, which is the main responsible for the methanol assimilation. Thus, the differential effect on the productivity depending on the expressed heterologous protein could be somehow linked to the carbon central metabolism. In addition, there could be additional genetic (and potentially phenotypic) differences between the strains, caused by the random mutagenesis with nitrosoguanidine used to obtain the HIS4 auxotrophic phenotype of GS115 (Cregg at al., 1985), the parental strain of KM71H.

Following the co-expression of rhGLA and HAC1p (both under the AOX promoter), it was demonstrated that the strain of choice was the wild-type (Mut^+). Upon co-expression of HAC1p, its average clonal specific productivity of 3503 AU·g X^{-1}·h^{-1} was ~4.1 times higher than the obtained for KM71H (863 AU·g X^{-1}·h^{-1}).

2.3.2.2. Overexpression of HAC1 and PDI in a Mut$^+$ α-galactosidase A producer clone

In front of the positive results obtained thanks to the inducible co-expression of HAC1p, its constitutive expression (under GAP promoter) was explored. In addition, the co-expression of PDI was included in the assay. With 5 disulfides bonds, it was hypothesized that the expression of rhGLA could be favored by the over-expression of PDI, which catalyzes the correct formation of disulfide bonds in secretory and cell-surface proteins. In this case, only the PDI gene under control of AOX promoter (methanol-inducible) was used.

In this experiment, a selected Mut$^+$ clone already expressing rhGLA (clone GLAwt_5.1) was used as the base strain to clone independently either HAC1p or PDI. As a control, a clone co-expressing HAC1p under AOX promoter (whose positive effect was evidenced previously) and the base strain expressing rhGLA alone were included in the assay (fig. 2.10, table 2.3).

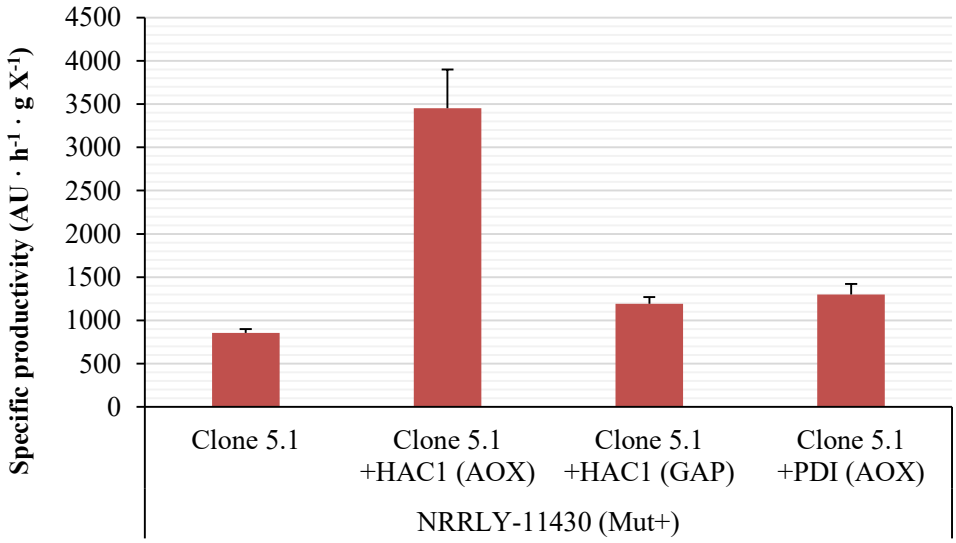


Fig. 2. 10. Specific productivity of rhGLA in NRRLY-11430 (Mut$^+$) co-expressing HAC1p (methanol-induced or constitutively) or PDI (methanol-induced). Average values correspond to culture triplicates of single clones

Table 2. 3. rhGLA specific productivities calculated for the co-expression of HAC1p (methanol inducible or constitutively) or PDI (methanol-inducible) in a Mut^+ strain, and their respective increase compared with the productivity of the original strain expressing rhGLA alone.

	Specific productivity (AU g X^{-1} h^{-1})	**Time fold (vs. Clone 5.1)**
Clone 5.1 (rhGLA alone)	855	-
Clone 5.1 +$HAC1_{AOX}$	3451	4.0 x
Clone 5.1 +$HAC1_{GAP}$	1193	1.4 x
Clone 5.1 +PDI_{AOX}	1300	1.5 x

The constitutive expression of HAC1p resulted in a 1.4 time-fold increase compared to the base clone (expressing only rhGLA). Although it was a significant increase, it was not able to approach the effect of its over-expression under AOX promoter (which increased 4 times the specific productivity of the transformed clone, a similar value to that obtained in the previous experiment). One of the reasons could be the difference in performance between promoters: AOX is considered to be more efficient than GAP when cells are grown on methanol (Cereghino and Cregg, 2000), which is required in this case for the induction of rhGLA expression. The results obtained are consistent with those reported in the literature for the expression of mouse interleukin-10 (mIL-10): while a 2.2 time-fold increase was obtained when co-expressing AOX-induced HAC1p, no improvement was seen for GAP promoter (Guerfal et al., 2010).

Meanwhile, the inducible expression of PDI was able to increase the specific activity 1.5 times. Even though the difference was significant, the result suggests

that the formation of the rhGLA disulfide bonds would not be the main expression bottleneck. Alternatively, this result could be interpreted as PDI being one of the factors contributing to the beneficial effect of the expression of HAC1, because PDI is one of the multiple ER-resident proteins whose expression levels are increased as a result of the UPR (Guerfal et al., 2010).

In any case, none of the tested options were able to reach the rhGLA productivity levels provided by the inducible co-expression of HAC1p. Although this option was not studied, it would be logical to explore if the combination of PDI with HAC1p (both under AOX promoter) can lead to an additive effect, thereby optimizing even more the rhGLA productivity capacity of *P. pastoris*. Furthermore, the positive results obtained globally with the co-expression of helper proteins mean that the over-expression of additional known expression enhancer proteins, such as Kar2 (BiP), could be a worthwhile approach.

2.3.3. Improvement of α-galactosidase A productivity by increase of gene dosage

The effect of the integration of multiple copies of the GLA gene in the yeast chromosomic DNA was assessed in parallel with the experiments described previously in this chapter. As mentioned in the general introduction, in *P. pastoris* the number of stably integrated copies is often proportional to the increase in expression levels obtained. However, there is not a consensus number of copies to maximize the productivity, and several publications show different values depending on the protein expressed (Vassileva et al., 2001; Zhu et al., 2009). Likely, the correlation between the number of copies and the productivity would be specific for each heterologous protein, probably due to different factors such as the particular product size and complexity. In fact, it has been suggested that not only an optimum copy-number would be reached (from which no improvement would

be experienced), but an excess can have a detrimental effect on the productivity (Marx et al., 2009).

Two different strategies were sequentially followed to obtain multi-copy strains expressing rhGLA. Firstly, a classical protocol based on the selection and isolation of transformants in increasing antibiotic concentrations was used. Later, the transformation of multi-copy vectors was evaluated as an improved, more directed alternative.

2.3.3.1. Evaluation of multi-copy strains isolated from agar plates of high antibiotic concentration

The selection of clones in increasing antibiotic concentrations is a simple strategy that has been efficiently used for the isolation of multi-copy *P. pastoris* clones (Sunga et al., 2008). In principle, the higher the number of vector integration events, the higher would be the resistance capacity thanks to a greater number of marker-genes. This would imply, consequently, the same copy-number of the gene of interest (GLA in this case). NRRLY-11430 (Mut^+) was transformed with the single-copy rhGLA vector, and a group of 21 clones were isolated in Zeocin concentrations 10-times higher than the standard (1 mg/mL vs. 100 µg/mL). The expression assays of these clones resulted in heterogeneous rhGLA titers (fig 2.11). Interestingly, they could be clearly grouped into three different groups according to the activity concentration reached in culture. The average value was calculated for each one, and were compared with a group of clones selected in standard Zeocin concentration (fig 2.12).

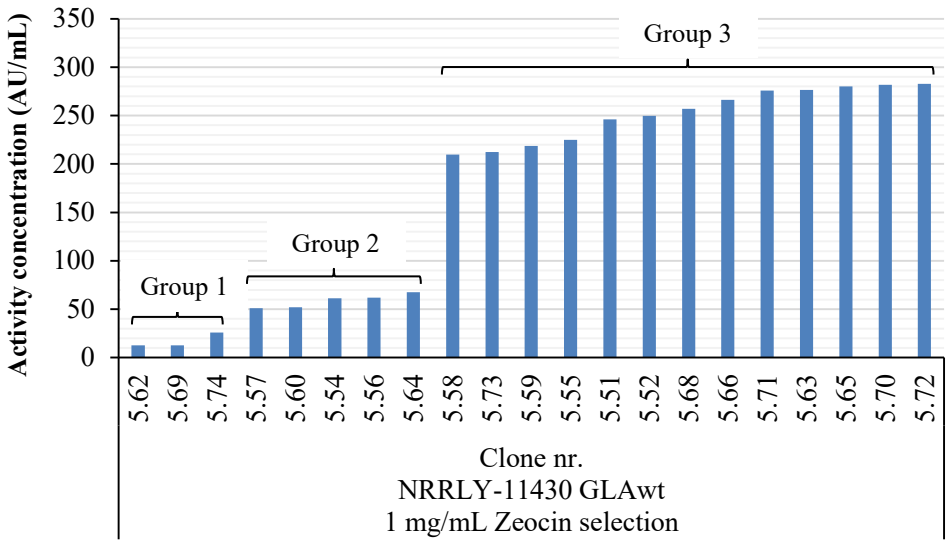


Fig. 2. 11. Clone screening of Mut^{+} clones expressing rhGLA isolated in high concentration of Zeocin (1 mg/mL). Activity concentration (AU/mL). Clones were ordered according to the activity concentration (AU/mL) reached in culture after 72 h of induction.

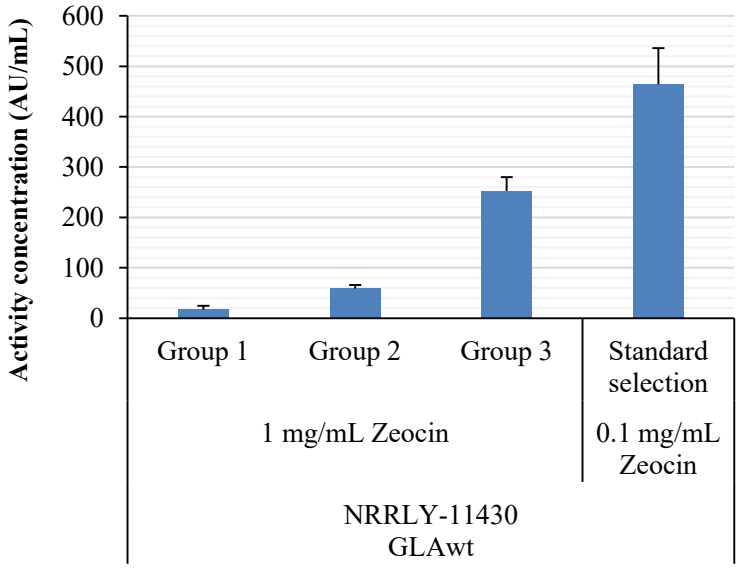


Fig. 2. 12. Comparison between the rhGLA production profile obtained from grouped clones isolated under high concentrations of Zeocin (1 mg/mL) and a group of clones isolated in standard concentration (100 µg/mL). Activity concentration average values (AU/mL) correspond to samples obtained after 72 h of induction.

Surprisingly, all of the 3 different groups of clones selected in high Zeocin concentration (in principle high-copy clones) reached GLA activity concentrations lower than the group selected using the standard antibiotic concentration (considered to be low- or single-copy). It was hypothesized that the number of plasmid integration events needed for the growth of *P. pastoris* in 1 mg/mL of Zeocin would imply a number of GLA gene copies over the bearable limit. That is to say, the different groups could correspond to different amounts of copies integrated, where the higher the amount of copies, the lower the activity due to an excess of gene dosage. The detrimental effect caused by an exceeding copy-number could be explained by the overburden of the transcriptional, translational and/or secretory machinery (Gasser et al., 2013). In order to confirm this hypothesis, it would have been useful to analyze clones from each group by real-time PCR to estimate their respective gene dosage (Abad *et al.*, 2010). Unfortunately, this option was not available. Although using intermediate Zeocin concentrations (such as 250 or 500 µg/mL) for clone selection and isolation could be an alternative to obtain more adjusted and optimized gene dosages, a different approach was followed based on the use of multi-copy vectors.

2.3.3.2. Evaluation of directed multi-copy strains generated by integration of multi-copy vectors

An alternative approach was attempted to define the required GLA gene dosage for the productivity optimization of the enzyme in *P. pastoris*. This time, two new

vectors were generated having two or three copies of the rhGLA expression cassette (i.e. AOX promoter – α-Factor – GLA gene – AOX terminator). The previously obtained single-copy vector was also included in the study as a control. All plasmids were transformed independently into a Mut^+ strain already over-expressing HAC1 under AOX promoter, as it was already defined as the best option at this point. This time, an increase of 1.51 and 1.57 time-fold was seen for clones transformed with 2- and 3-copy vectors, respectively, in comparison with those obtained with the single-copy one (fig. 2.13). As no significant differences were found between them, this copy range was considered to be the maximum for *P. pastoris*.

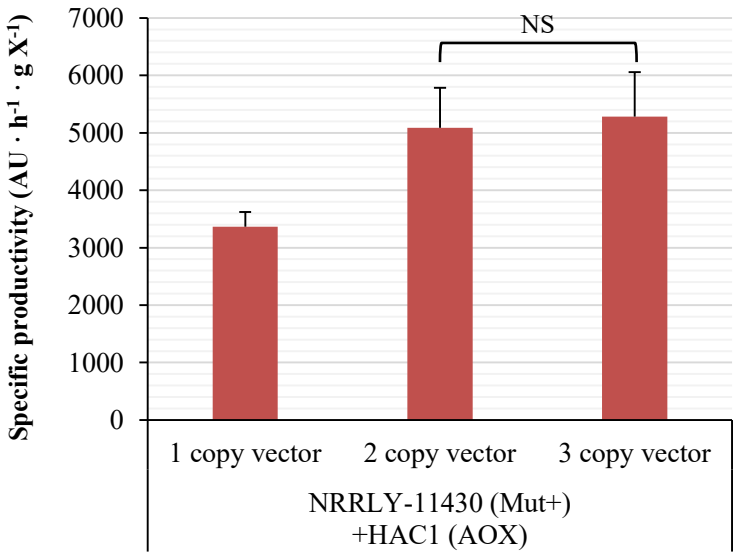


Fig. 2. 13. Specific productivity (AU · h-1 · g DCW-1) of clones transformed with vectors encoding 1, 2 or 3 copies of the GLA gene. Average values correspond to the best 6 clones from groups of 18.

While most of the clones of the single-copy group showed similar productivities, the overall clone variability of both multi-copy groups was substantially higher and contained a significant number of low or non-producers (data not shown). For this reason, the average specific productivity of the best 6 clones from every group was

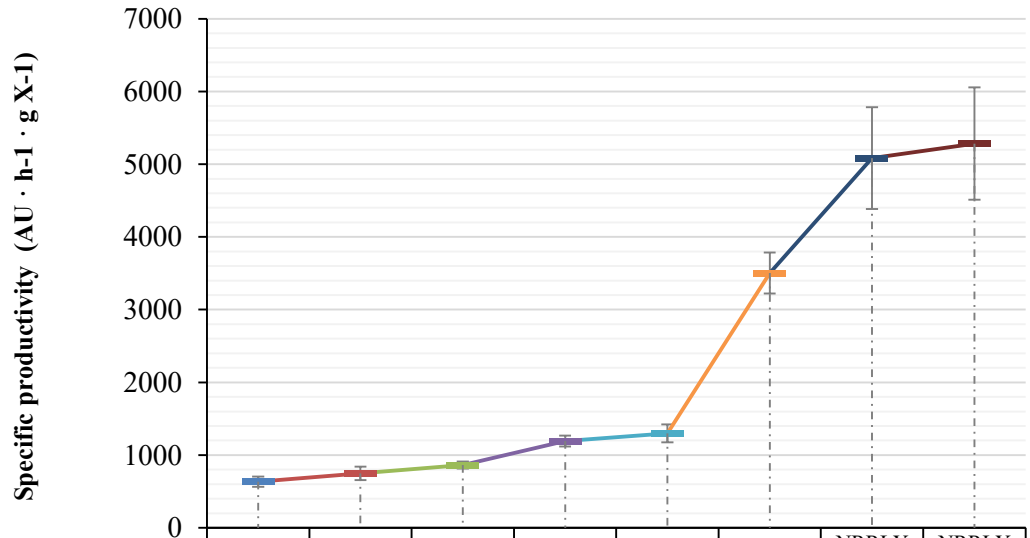
calculated to carry out the comparison. A possible explanation would be the higher susceptibility of these groups to produce multiple integration events, leading to an excessive number of GLA gene copies (as hypothesized with the selection of transformants under high antibiotic concentrations). In fact, an optimum copy number of 2 or 3 is relatively low (and thus quite easy to achieve) compared with other recombinant genes reported in the literature (Zhu et al., 2009). The reason why multiple integration could be more frequent when using multi-copy vectors is explained by the higher number of AOX promoters present in each individual vector. As mentioned above, each expression cassette of rhGLA entails its own promoter, which contains the restriction site to linearize the vector and target its integration by recombination with the endogenous AOX1 promoter. Thus, if a 3-copy vector is successfully integrated, the chromosome goes instantly from having 1 AOX1 promoter to 4 (3 for rhGLA + 1 endogenous for AOX), which instantly become DNA regions where new recombination events can potentially take place.

Even though the transformation with 2- and 3-copy vectors required the screening of a higher amounts of clones (compared to the single-copy vector), it was acceptably easy to isolate efficient clones. Therefore, the use of a 2-copy vector was considered to be the best choice given the fact that the effect of transforming the 3-copy one did not provide a significant improvement. Moreover, the 2-copy vector would presumably overload to a lesser extent the yeast secretory pathway, something to take into account because additional genes are expected to be expressed to modify the yeast glycan profile.

2.3.4. Summary of α-galactosidase A productivity improvements

In front of the results discussed above, it was established that the transformation of a Mut+ strain with a vector encoding two rhGLA copies under AOX promoter, together with the co-expression of HAC1 (under the AOX promoter too)

constituted the best option for the overexpression of rhGLA in *P. pastoris*. From the initial specific productivity obtained with the Mut^{+} strain, the optimization



	KM71H	NRRLY -11430	KM71H +HAC1 (AOX)	NRRLY -11430 +HAC1 (GAP)	NRRLY -11430 +PDI (AOX)	NRRLY -11430 +HAC1 (AOX)	NRRLY -11430 +HAC1 (AOX) 2 GLA copies	NRRLY -11430 +HAC1 (AOX) 3 GLA copies
Specific productivity (AU · h-1 · g X-1)	633	749	863	1193	1300	3503	5085	5283

Fig. 2. 14. Progressive increase of rhGLA specific productivity obtained during the optimization process.

process led a 7 fold-change increase (fig 2.14).

The titers reached were considered to be high enough to use rhGLA as a model lysosomal enzyme for the development of humanized strains. If further strain optimization would be considered in the future, there are a number of different variables that could be studied to try to increase even more the expression of rhGLA. As a first example, the screening of different promoters would be an option. Although the inducible AOX promoter is usually the first choice, the

possibility that alternative promoters could lead to better expression yields cannot be discarded. A great number of sequences have been described for the expression of recombinant proteins in *P. pastoris* (both inducible and constitutive), with different success depending on each product (Stadlmayr et al., 2010). Moreover, variant libraries based on promoter mutation and truncation have been studied for the specific cases of AOX and GAP promoters, leading to significant increases in expression levels (Hartner et al., 2008, Qin et al., 2011). Interestingly, the efficiency rankings obtained for the promoter variants wer shown to be different depending on the protein expressed, suggesting that a new promoter screening would be necessary for every new product. Another factor that is related to different expression levels is directly associated with the recombinant gene coding sequence. It has been reported that different synthetic codon-optimized sequences can be responsible of significant differences in productivity (Mellitzer et al., 2014). This effect could be due to different codon-usage frequencies within a single gene and the related tRNA availability for RNA translation. In this case, a random library of codon-optimized ORFs should be generated and compared. Furthermore, the use of different secretion signals has been a common variable to improve the expression of a given protein (Karaoglan et al., 2014). Again, there is not an ideal signal peptide for every protein, although the full version of α-factor (widely known simply as α-factor) used in this work is usually a good choice. There are different possible reasons why one signal would be better than other in every case, such as the cleavage efficiency depending on the first amino acid of the mature protein (Yang et al., 2013), but other mechanisms that are not clearly understood could be involved. Finally, the use of a wide range of helper proteins could be tested (independently or in combination) to boost the expression and secretion capability of *P. pastoris* and increase product titers. Even though two of the main ones have been studied in this work (HAC1p and PDI), the knowledge about proteins that could potentially enhance the expression performance of the host cell is constantly growing (Idiris et al., 2010; Puxbaum et al., 2015). These helper proteins can be involved in protein expression at many different levels:

transcription, translation, signaling, folding and quality control, trafficking, etc. A high-throughput screening would be useful to define not only the best combination of helper proteins, but to modulate their expression by using combinations of promoters of different strengths.

2.3.5. Expression of α-galactosidase A variants

The expression of rhGLA variants, intended for different purposes, was also assessed. In first place, a version of GLA lacking the C-terminal peptide consisting of the KDLL amino acids (GLAΔKDLL) was tested. Although not proven at the protein level, an increase in the specific activity of GLA could be expected according to the literature (Miyamura et al., 1996). If confirmed, the efficient expression of such variant could mean a catalytic improvement with respect to the rhGLA already available as ERT for Fabry disease patients. Another variant tested was rhGLA fused to the RGD motif (either to the N- or C-terminus, $_{N\text{-}RGD}$rhGLA or rhGLA$_{C\text{-}RGD}$ respectively). The RGD motif was fused with the aim to bypass the lack of human M6P glycosylation profile and achieve the efficient internalization of the enzyme into the lysosomes via the integrin-recognition pathway (Temming et al., 2005). Finally, a 6xHistidine purification tag was fused to the enzyme sequence (again, either in the N- or C-terminus, $_{N\text{-}HIS}$rhGLA or rhGLA$_{C\text{-}HIS}$ respectively). This version should allow its easy purification in a single affinity chromatography step, an advantage highly interesting for the analysis and evaluation of the protein produced in versions of humanized *P. pastoris* strains, which are expected to be generated in upcoming phases of the research developed at Bioingenium. The rhGLA variants cloned and tested are summarized in table 2.4.

All variants, which were transformed into the Mut$^+$ strain, yielded significant expression levels with the exception of those with a modified N-terminus.

Surprisingly, the addition of either the RGD motif or 6xHis tag was only viable at the C-terminus of the enzyme (fig. 2.15). This common result obtained for $_{N\text{-}RGD}$rhGLA and $_{N\text{-}HIS}$rhGLA suggests that the N-terminus of GLA may play an important role in the enzyme functionality and/or stability.

Table 2. 4. Summary of rhGLA variants expressed in *P. pastoris* in this work.

rhGLA variant	**Description**	**Purpose**
GLAΔKDLL	Deletion at C-ter (last 4 amino acids)	Increase enzyme specific activity
$_{N\text{-}RGD}$rhGLA	RGD motif fused to N-ter	Provide the enzyme with a cell internalization mechanism (integrin-mediated)
rhGLA$_{C\text{-}RGD}$	RGD motif fused to C-ter	
$_{N\text{-}HIS}$rhGLA	6xHis tag + enterokinase cleavage site fused to N-ter	Addition of a purification tag to simplify the chromatographic purification process
rhGLA$_{C\text{-}HIS}$	6xHis tag fused to C-ter	

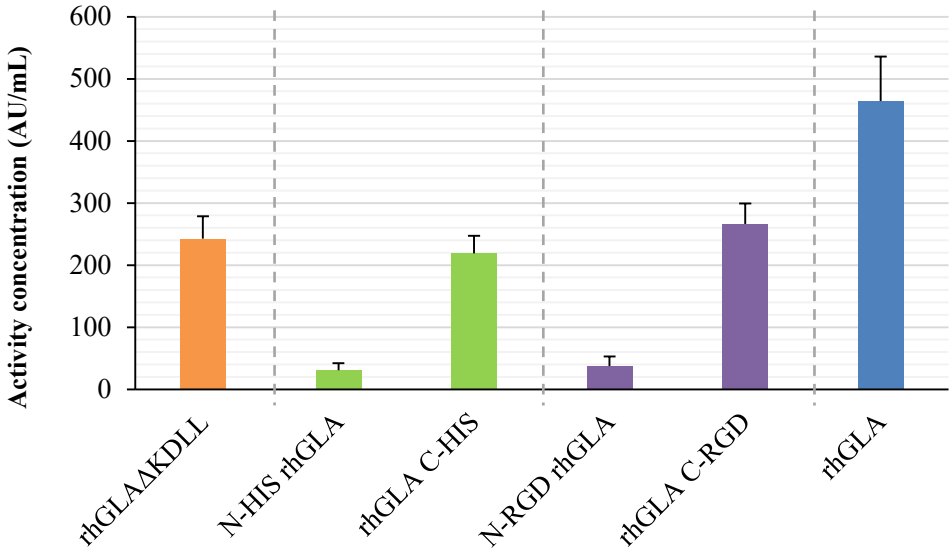


Fig. 2. 15. Activity concentration (AU/mL) reached for clones expressing variants of rhGLA. Between 24 and 48 isolated clones were analyzed for each group. Orange bar: rhGLA lacking the C-terminal KDLL sequence. Green bars: rhGLA variants containing a terminal 6xHis tag. Purple bars: rhGLA variants containing a terminal RGD motif. Blue bar: unmodified rhGLA.

When compared to the activity concentrations obtained in expression assays of the original rhGLA, all the three different viable modifications (GLAΔKDLL, $rhGLA_{C\text{-}RGD}$ and $rhGLA_{C\text{-}HIS}$) yielded lower values. At this point, it was unknown if the difference was due to an impact of the modifications on the specific activity of the variants, to a decrease in their expression efficiency, or to a combination of both factors. This question was elucidated later after the purification of each version (see Chapter 4).

Despite the lower expression levels obtained for the different versions of rhGLA, a group of clones were successfully obtained for the low-scale production, purification and evaluation of each enzyme variant. Of course, those containing modifications on the N-terminus were discarded for further characterization.

2.3.6. Generation of an rhGLA$_{C\text{-}HIS}$ producing strain as a base for the co-expression of glycosylation-related enzyme complexes.

Regarding the expression of the rhGLA$_{C\text{-}HIS}$ variant, additional experiments were performed to obtain producer clones prepared for the co-expression of different glycosylation-related enzyme complexes required for the generation of humanized stains. The process was divided in 2 stages: firstly, clones producing rhGLA$_{C\text{-}HIS}$ either under AOX or GAP promoters. If cloned under AOX promoter, the enzyme was co-expressed with HAC1p to boost its productivity. Secondly, a selected clone was used to set-up the procedure for the reversion of the resistant phenotype, enabling the re-transformation and selection of the new strain with any resistance marker available.

2.3.6.1. Selection of rhGLA$_{C\text{-}HIS}$ producing clones: constitutive expression under GAP promoter or methanol-inducible co-expressed with HAC1p

Taking advantage of the data previously obtained, it was decided to try to increase the productivity of the rhGLA$_{C\text{-}HIS}$ variant by co-expressing HAC1p under the AOX promoter. In order to use a single selection marker for the expression of the 2 proteins, the rhGLA$_{C\text{-}HIS}$ gene was re-cloned into a new vector already having the HAC1 expression cassette, which was previously generated. This way, an empty vector already encoding HAC1 would be available also to clone any other protein of interest.

In parallel, the rhGLA$_{C\text{-}HIS}$ gene was cloned into a vector under the GAP constitutive promoter. A strain producing rhGLA$_{C\text{-}HIS}$ constitutively would enable the co-efficient co-expression of other proteins controlled by the same promoter, an interesting alternative because GAP has been suggested to be a better choice for the expression of some enzymes involved in the glycosylation profile (Callewaert et al., 2001). These type of proteins are indeed the ones intended to be functionally

co-expressed in future project stages in order to obtain *P. pastoris* strains with a humanized glycosylation profile. As previous experiments suggested that the co-expression of HAC1p under GAP promoter did not lead to an important productivity increase, such improvement was omitted in this case.

Both new vectors were successfully transformed into a wild-type Mut^+ strain (NRRLY-11430). The expression assays of groups of isolated clones showed different productivities depending on the plasmid used (fig. 2.16). On one hand, the co-expression of $rhGLA_{C\text{-}HIS}$ and HAC1p under AOX promoter was able to yield similar specific productivity levels to those obtained previously for the unmodified rhGLA (3495 vs. 3503 AU · g X^{-1} · h^{-1}, respectively), meaning that HAC1p helped to void the difference seen in previous experiments between both versions in the absence of HAC1p over-expression. Regarding the expression of $rhGLA_{C\text{-}HIS}$ under GAP promoter, the productivity levels were 37% higher compared to those seen previously for the unmodified rhGLA under AOX promoter and without the co-expression of HAC1p (1027 vs. 749 AU · g X^{-1} · h^{-1}, respectively). This result confirmed the suitability of the GAP promoter for the efficient expression of GLA.

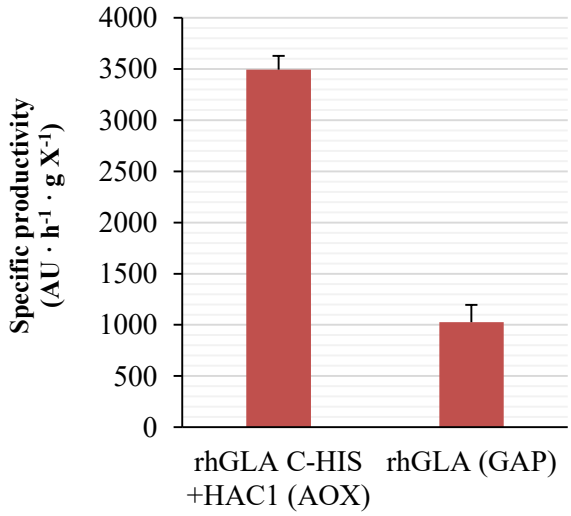


Fig. 2. 16. Comparison between specific productivity of NRRLY-11430 (Mut+) clones expressing $rhGLA_{C\text{-}HIS}$ and HAC1p under AOX (methanol-inducible) or GAP (constitutive) promoter. Average values were calculated from 12 clones for each group.

2.3.6.2. Reversion of Zeocin resistant phenotype by transient expression of Cre recombinase

After achieving the efficient expression of rhGLA$_{C\text{-}HIS}$ in *P. pastoris*, it was decided to remove the integrated resistance gene taking advantage of the flanking loxP sites. Reverting the Zeocin resistance would lead to a non-resistant rhGLA producer strain ready to be transformed with vectors containing any marker based on antibiotic selection (including Zeocin itself). Thus, the strain would become a useful tool to test different protein complexes related with the glycosylation pattern, either by co-transformation or by successive transformation of plasmids having different resistance markers and the genes of interest. The recombination of the flanking loxP sites is catalyzed by the Cre recombinase (Zhang and Lutz, 2002), whose expression should be temporary in order to perform its task precisely. A suitable approach for the transient expression of proteins in yeast is the use of episomal (non-integrative) plasmids, containing a yeast-specific replication origin (Lee et al., 2005). The instability of episomal vectors in *P. pastoris* represents an advantage in this sense, as the chromosome-independent DNA will be lost after a few generations together with the protein expression associated.

An episomal plasmid encoding the Cre recombinase under the GAP constitutive promoter was already available at that time (pBGP1_Cre). However, the vector contained Zeocin as the resistance marker for its selection, being useful to revert any *P. pastoris* resistant phenotype different from Zeocin (provided that such resistance is flanked by loxP sequences). In such case, it would be impossible to positively select those cells having the episomal vector. Aiming to enable the reversion of any resistant clone obtained by transformation with Bioingenium plasmids, a new episomal vector having Hygromycin B as an alternative selectable

marker had to be generated. In principle, the combination of both vectors should enable the sequential reversion of resistant clones, therefore allowing the transformation of the multiple plasmids required for the expression of protein complexes by using only two interchangeable selectable markers (fig 2.17).

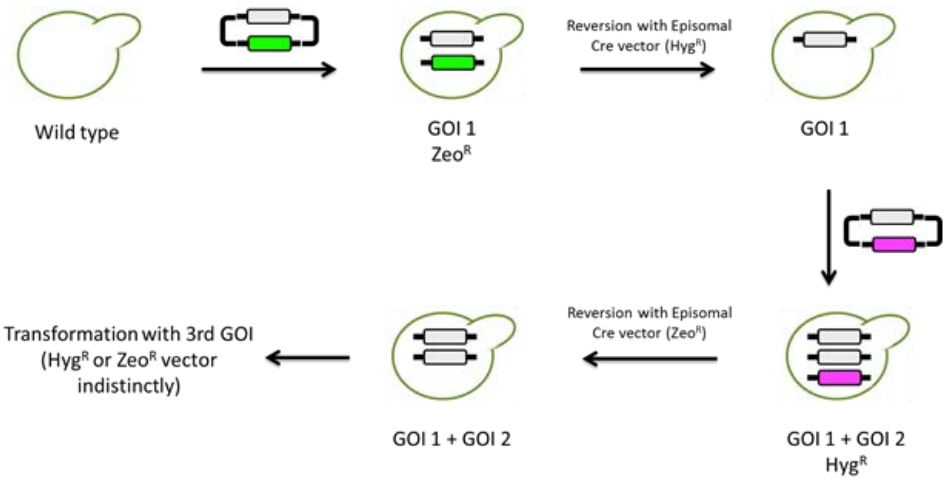


Fig. 2. 17. Process for the successive transformation of genes of interest (GOI) and reversion of resistant phenotype, allowing the co-expression of multiple proteins by using a single resistance marker (Zeocin) or alternating 2 of them (Zeocin & Hygromycin B).

In order to set-up the method to revert the resistant phenotype, the best rhGLA$_{C\text{-}HIS}$ producer clone (co-expressing HAC1, both under AOX promoter) was chosen. The new transformation and selection method was adjusted to obtain an efficient transient expression of Cre and the desired resistance phenotype reversion. This procedure was based on the transformation of a small quantity of non-linearized episomal vector (1 µg) into electrocompetent cells of the rhGLA$_{C\text{-}HIS}$ producer clone, and the successive subculture in fresh liquid media (first in presence of the antibiotic encoded by the episomal vector, i.e. Hygromycin B, and later without any antibiotic). As a final step, antibiotic-free agar plates were used to isolate clones. These clones were later screened by picking each one in three different

plates: again without antibiotic, with Zeocin (the resistance marker of the original strain), and with Hygromycin B (encoded by the episomal vector) (fig 2.18).

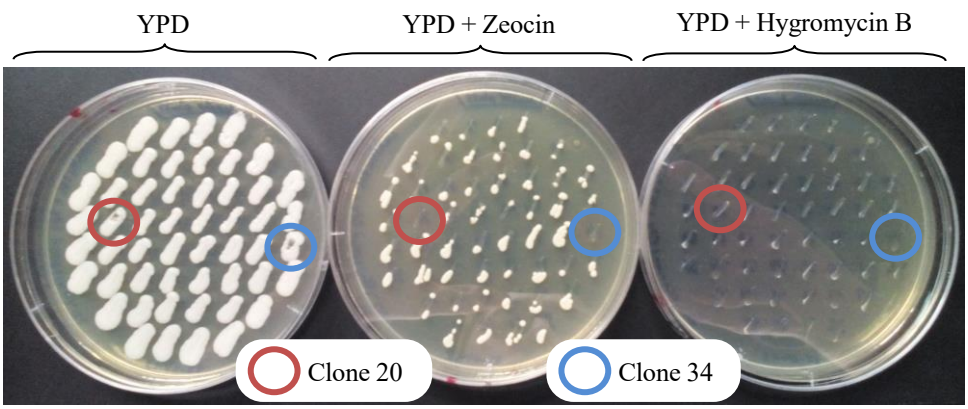


Fig. 2. 18. Agar plates with isolated clones obtained after reversion of the Zeocin-resistant phenotype on a selected clone co-expressing rhGLA$_{C\text{-}HIS}$ and HAC1p. After transient expression of Cre recombinase, each clone was picked in 3 different plates. From left to right: YPD alone, YPD + Zeocin (100µg/mL) and YPD + Hygromycin B (200µg/mL).

In principle, those clones able to grow only in the YPD agar plates (without any antibiotic) would be the ones with the expected non-resistant phenotype. On the other hand, clones growing also in presence of Zeocin would have retained the original resistance marker. Finally, growth in Hygromycin B plates would mean either the permanence of the episomal vector or its chromosomic integration by non-homologous recombination. As showed, none of the selected clones were able to grow on the Hygromycin B plate, meaning that the episomal vector was efficiently cured during the process. Moreover, an aberrant growth could be seen in Zeocin plates, typical from non-resistant clones (especially when an excess of cells is drawn into the streak). Thus, clones 20 and 34 were finally selected due to the

absolute absence of growth in the Zeocin plate (although all of them seemed to have lost the resistant phenotype).

As a final test to check if the mentioned clones had efficiently lost the resistance marker, liquid cultures of both were carried out in parallel either without antibiotic or in presence of Zeocin. As a control, the original clone transformed with the episomal vector (co-expressing rhGLA$_{C\text{-}HIS}$ and HAC1p, Zeocin-resistant) was used (fig. 2.19).

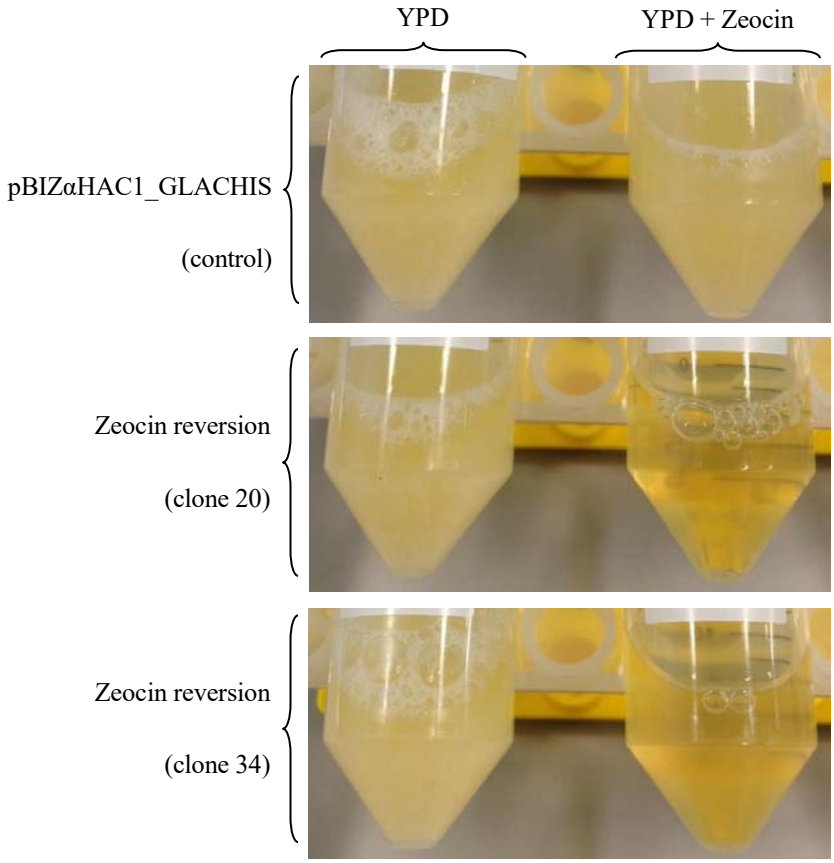


Fig. 2. 19. YPD cultures of clones with reverted resistant phenotype (20 & 34) and the original clone co-expressing rhGLA$_{C\text{-}HIS}$ and HAC1p, Zeocin-resistant. Tubes on the left side: YPD alone. Tubes on the right side: YPD + Zeocin (100 µg/mL).

The total absence of growth in the media containing Zeocin confirmed the successful excision of the resistance marker from the chromosome and the efficient resistant phenotype reversion. Moreover, both reverted clones were submitted to the usual expression assay protocol, which confirmed the retention of the $rhGLA_{C\text{-}HIS}$ productivity profile (compared to the original Zeocin-resistant clone). That means that the recombination event catalyzed by the Cre recombinase did not affect the expression cassette of the gene of interest. Clone 20 was finally selected and is being currently used as the receptor strain to test different combinations of glycosylation-related proteins, aiming to obtain newly humanized *P. pastoris* strains.

2.4. Conclusions

- Mut^+ (normal methanol utilization phenotype), is the preferred strain for the expression of rhGLA, in contrast with Mut^S (slow methanol consumption phenotype).

- Co-expression of rhGLA with HAC1p or PDI contributes to improved specific productivity. The expression of HAC1p under AOX promoter (methanol-inducible) represents the best choice, eliciting a 4.7 time-fold increase in product titers.

- By using multi-copy vectors, the beneficial effect of the multiple integration of GLA gene copies into the yeast chromosome was evidenced. Nevertheless, the maximum effect in the enzyme expression was achieved with a relatively low number of copies: 2 to 3. Together with the co-expression of HAC1p, productivity was finally raised 7 times compared to the expression of a single copy in a Mut^+ wild-type strain.

- The expression of rhGLA variants intended for different purposes was successfully accomplished. Interestingly, sequence modifications were only

viable in the C-terminus, as the expression of active enzyme was seriously impaired upon expression of N-terminal modified variants.

- The expression of a tagged version of hGLA ($rhGLA_{C\text{-}HIS}$) was optimized by co-expression of HAC1 under AOX promoter. Moreover, the suitability of the GAP constitutive promoter for the expression of this enzyme variant was confirmed.

- Resistant phenotype reversion of strains transformed with Bioingenium plasmids was shown to be feasible. No detrimental effect in rhGLA productivity titers was observed, and a non-resistant clone producing $rhGLA_{C\text{-}HIS}$ was selected as the base strain for future humanization of yeast glycoprofile.

Chapter 3

Production of recombinant human α-galactosidase A in bioreactor

Chapter 3

PRODUCTION OF RECOMBINANT HUMAN α-GALACTOSIDASE A IN BIOREACTOR

3.1. Introduction

The development of a scalable bioreactor process for the production in *P. pastoris* of recombinant human α-galactosidase A (rhGLA), aiming to implement a reproducible process to obtain significant amounts of the enzyme, is presented in this chapter.

One of the main advantages of the use of *P. pastoris* in protein production is the extraordinary cell densities that can be reached in bioreactor cultures by the yeast. As a general rule, in bioreactor processes is commonly accepted that the higher the cell concentration achieved in the vessel, the higher the volumetric productivity will be. Indeed, >10 g/L of product titers have been documented in *P. pastoris* for certain proteins (Mellitzer et al., 2015; Xiong et al., 2006).

Methanol-limited fed-batch (MLFB) has been one the most extended strategies for the protein production in *P. pastoris*. Unlike MLFB, non-limited cultures require a complex set-up for the control of cultivation parameters, representing technical and economic constraints only affordable for experimented research centers. Limitation of cell growth by controlled methanol supply (acting both as the sole carbon source and the expression inducer) is, on the other hand, easier to implement and control. Even though strict control over dissolved oxygen (dO) and temperature are still required, due the high oxygen consumption and heat generation of methanol-limited cultures, this methodology is considered a standard (Looser et al., 2015). For this reason, this was the first strategy tested for the production of rhGLA.

The methanol-limited strategy can be divided in three major phases (fig 3.1). The first two (glycerol batch and fed-batch) have as a clear objective the fast increase of cell biomass. Because methanol is not a preferred substrate carbon source for *P. pastoris*, glycerol is used instead. This substrate is characterized by a higher substrate/biomass yield compared to methanol (Looser et al., 2015), and is known to be an AOX1 promoter repressor (Valero, 2013) allowing the culture to focus their resources to cell growth instead of protein expression. When a significant cell concentration is reached (40 – 45 g/L DCW) glycerol is substituted by methanol, triggering the protein expression.

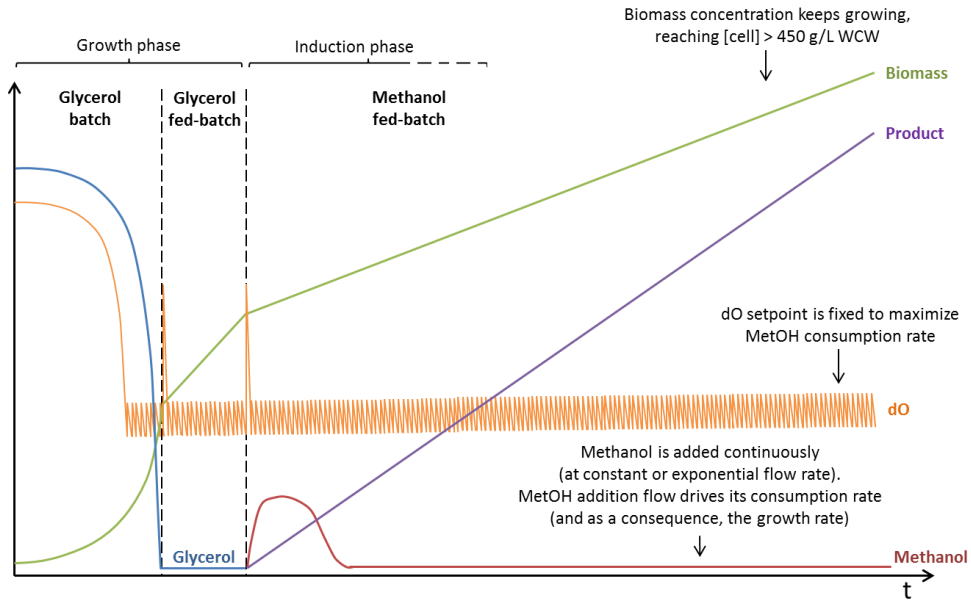


Fig. 3. 1. Typical MetOH-limited fed-batch strategy profile. The fermentation process is divided in 3 major stages. The first two ones, glycerol batch and fed-batch, are intended two achieve a high cell concentration. Once the desired cell concentration is achieved, the methanol fed-batch phase starts and triggers the protein production. During the whole process, enough aeration is guaranteed to maintain a dO set-point above 20%, keeping first glycerol and then methanol as the limiting nutrients.

As shown in the chapter, this strategy provided poor productivity levels for rhGLA. Moreover, the excess of cell biomass and the high metabolic status of the culture led to unstable processes where the cooling and oxygenation demands were higher than acceptable, compromising both its reproducibility and economic viability.

A new approach was clearly necessary for the production of rhGLA in bioreactor, and oxygen-limited fed-batch (OLFB) was proposed as an alternative. A new cultivation strategy, based on the O_2 limitation and addition of methanol in pulses during the induction phase (Berdichevsky et al., 2011), was tested. This strategy, as with the methanol-limited one, can be also divided in three phases (fig 3.2).

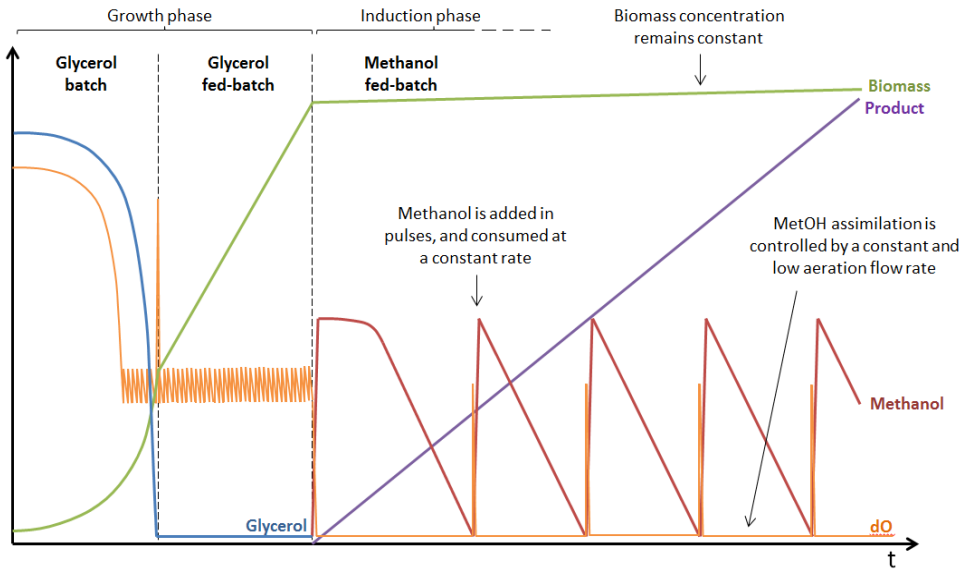


Fig. 3. 2. Typical O2-limited fed-batch strategy profile. The fermentation process is divided in 3 major stages. The first two ones, glycerol batch and fed-batch, are intended two achieve a high cell concentration. Once the desired cell concentration is achieved, starts the methanol fed-batch phase. Methanol, which triggers the protein production, is added in pulses and aeration is set to a low flow to keep O_2 as the limiting factor.

Again, the first two (glycerol batch and fed-batch) are aimed at achieving high cell densities in the bioreactor prior to the induction phase. Even though the glycerol batch is identical in both cases, the procedures differ from this point. In first place, the glycerol fed-batch is prolonged for 16 – 18 h aiming to reach higher cell densities (around 75 g/L DCW), in order to compensate the lack of cell growth experienced later during the induction phase. Meanwhile, the induction phase is radically different: methanol is added in pulses and air flow is fixed to a low value (0.2 vvm). Such low aeration profile keeps the O_2 as the limiting nutrient and drives the methanol consumption, which shows a linear decrease after every addition. Biomass concentration remains more or less constant during the induction phase.

The adaptation of this new methodology led to the production of significant amounts of enzyme, enabling the development of its purification process and further assays. The process was not only useful in terms of productivity, but also allowed to decrease dramatically the consumption of pure O_2 and improved notably the stability and reproducibility of the cultures. This novel strategy was then used to compare the bioreactor productivity of 2 different strains that showed a marked differences in specific productivity at the 10 mL-scale of the clone screening. The differences were maintained, proving the predictable behavior of this bioreactor strategy based on the results obtained at smaller scales.

3.2. Materials and methods

3.2.1. Strains

Two different clones were chosen for cultivations: clone GLAwt_5.1 (expressing rhGLA alone, under AOX promoter) and clone GLAwt_5.1[$HAC1_{AOX}$] (expressing rhGLA and HAC1p, both under AOX promoter) (see chapter 2).

3.2.2. Preparation of pre-inoculums: media culture and cell growth

For each bioreactor cultivation, 1 L baffled shake flasks with 200 mL of BMGY (10 g L^{-1} yeast extract, 20 g L^{-1} peptone, 10 g L^{-1} glycerol, 13.4 g L^{-1} yeast nitrogen base without amino acids, 0.4 mg L^{-1} biotin, 100 mM potassium phosphate buffer pH 6) with 100 μg/mL of Zeocin were inoculated with the corresponding *P. pastoris* rhGLA producer clone. Shake flask cultures were incubated for 24 h at 28 ºC, 200 rpm. For the preparation of the bioreactor inoculum, the volume from the shake flask culture necessary for an initial OD_{600} of 0.5 in bioreactor was centrifuged at 3000 g, 10 min, and re-suspended in 20 mL of sterile PBS. The cell suspension was immediately transferred to the bioreactor to initiate the culture.

3.2.3. Bioreactor media formulation and additives composition

Fermentation basal salts medium was used for all the fed-batch bioreactor cultivations (26.7 mL L^{-1} 85% phosphoric acid, 0.93 g L^{-1} calcium sulfate, 18.2 g L^{-1} potassium sulfate, 14.9 g L^{-1} magnesium sulfate · 7 H_2O, 4.13 g L^{-1} potassium hydroxide, 40 g L^{-1} glycerol, 4.35 mL L^{-1} PTM1 trace salts solution (6 g L^{-1} cupric sulfate · 5 H_2O, 0.08 g L^{-1} sodium iodide, 3 g L^{-1} manganese sulfate · H_2O, 0.2 g L^{-1} sodium molybdate · 2 H_2O, 0.02 g L^{-1} boric acid, 0.2 g L^{-1}

cobalt chloride, 20 g L^{-1} zinc chloride, 65 g L^{-1} ferrous sulfate · 7 H_2O, 0.2 g L^{-1} biotin, 5 mL L^{-1} sulfuric acid), 4 mL L^{-1} biotin solution (200 mg L^{-1})). 50% w/w glycerol solution supplemented with 12 mL L^{-1} PTM1 trace salts solution and 2 mL L^{-1} biotin solution was used to feed the fed-batch growth phase. For the induction phase, 100% methanol feed solution supplemented with 12 mL L^{-1} PTM1 trace salts solution and 2 mL L^{-1} biotin solution was used. pH was automatically controlled along the cultures with 30% ammonia and 4 M ortophosphoric acid. When necessary, 0.2% Antifoam 204 solution was added to avoid foam formation. All reagents were purchased from Sigma-Aldrich.

3.2.4. Bioreactor equipment

Cultivations were performed in a 5 L glass autoclavable bioreactor equipped with an ez-Control unit, and process data were adquired and supervised by the Bioxpert-NT software (Applikon Biotechnology, The Nederlands). A methanol sensor system (Raven Biotech, Canada) was used to monitor the methanol concentration during the induction phase. Feeding strategies were performed with 2 automatic burettes (Burette 1S, Crison Instruments). A process diagram and a picture showing the equipment setup are depicted below (figs. 3.3 and 3.4).

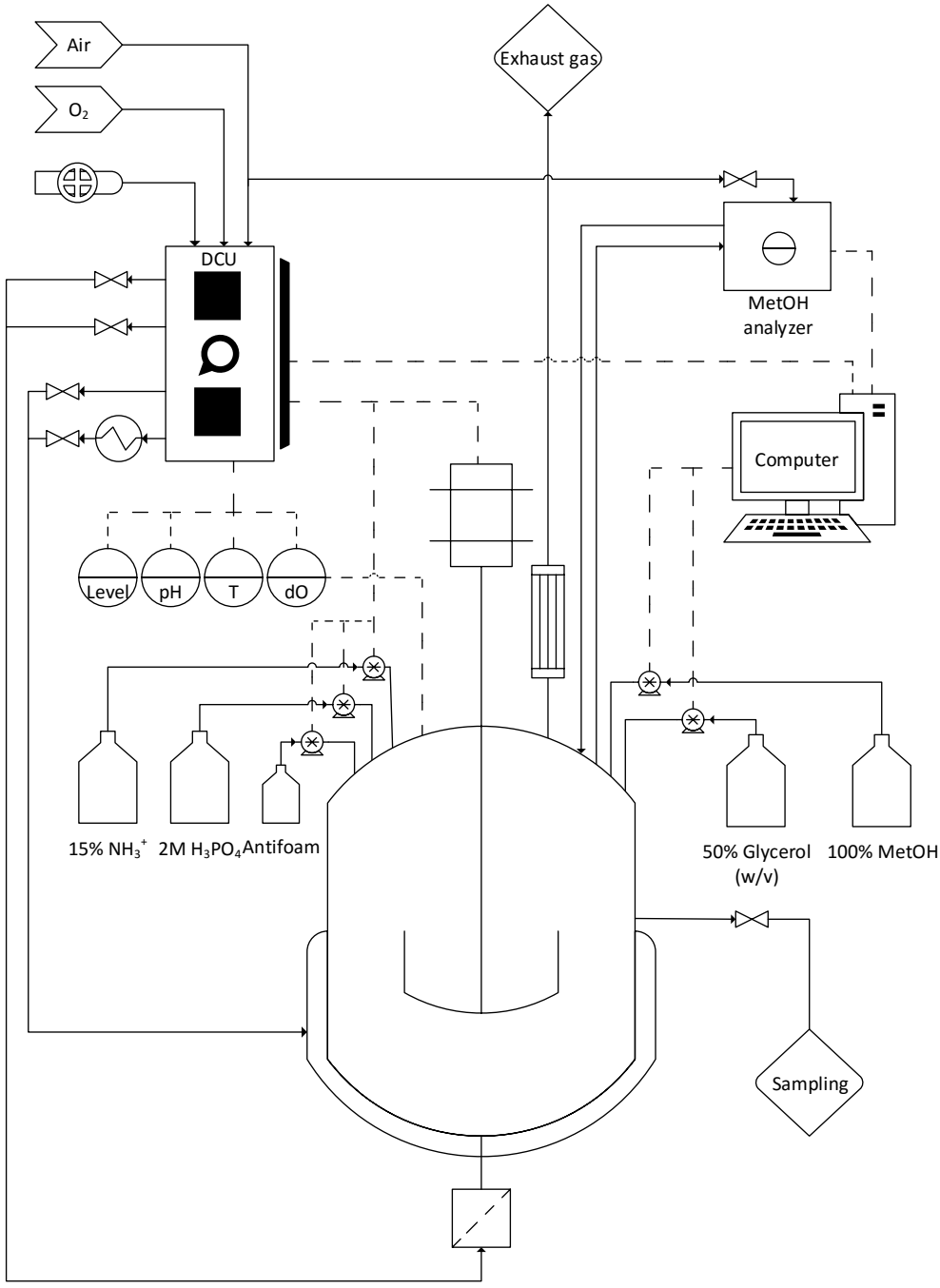


Fig. 3. 3. Process diagram of the experimental setup.

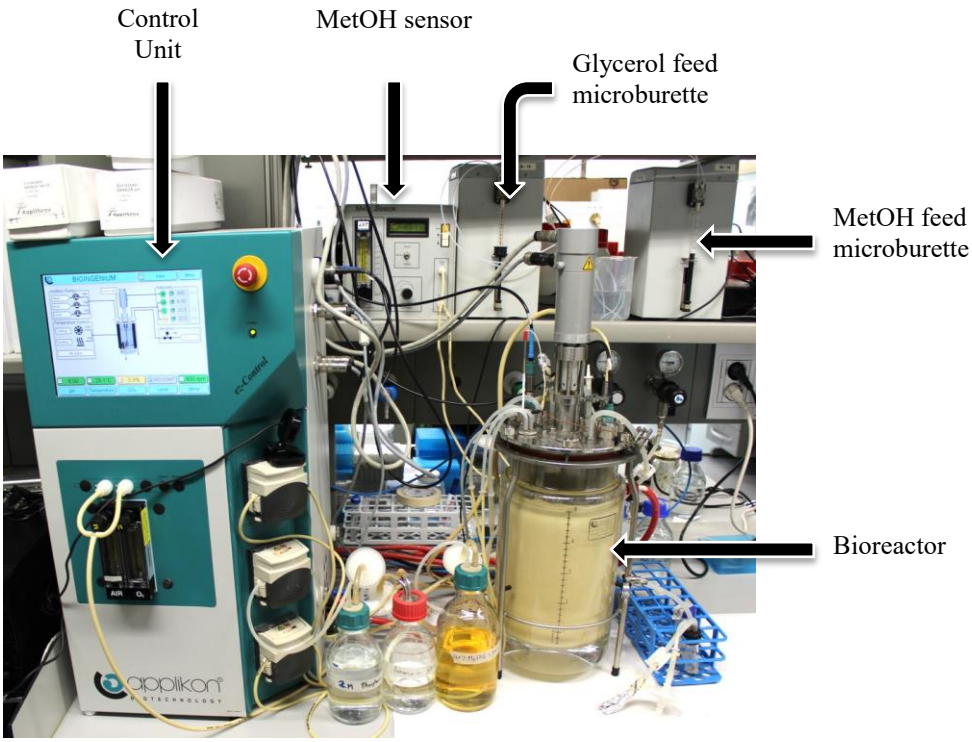


Fig. 3. 4. Equipment used for bioreactor cultivations.

3.2.5. Methanol-limited fed-batch cultivation strategy

2.5 L of basal salts Medium previously sterilized by autoclaving were inoculated with clone GLAwt_5.1 expressing rhGLA. Process temperature was 25 ºC and pH was controlled at 6 with 30% ammonia. dO was maintained at 20% of saturation atmospheric pressure by fixing the flow rate at 1 vvm and cascading agitation (800 to 1100 rpm). If maximum agitation was not enough to maintain the dO setpoint, 0.6 vvm of pure O_2 were added to the air inlet. End of batch phase was indicated by an increase in the dO measure that signaled the depletion of 40 g L^{-1} of glycerol. A constant glycerol feed of 756 µL min^{-1} was then started (glycerol fed-batch phase). Glycerol feed was maintained until a cellular concentration of 180-220 g L^{-1} wet

cell weight was reached, which required 6-7 h. After the end of glycerol fed-batch, methanol fed-batch was initiated immediately with a low initial flow rate of 150 µL min^{-1} for 2-3 h to allow the culture to adapt to the methanol consumption. After the initial adaptation, the feed was doubled to 300 µL min^{-1} for the next 2 h, and finally increased to 454 µL min^{-1}, which was maintained throughout the remainder of the fermentation. The induction was prolonged for 119 h, but a partial draining of the vessel (1 L) was done after 46 h to allow a phase extension.

3.2.6. O_2-limited fed-batch cultivation strategy

3 L of basal salts medium previously sterilized by autoclaving were inoculated either with clone GLAwt_5.1 (expressing rhGLA alone) or clone GLAwt_5.1[$HAC1_{AOX}$] (co-expressing HAC1p). During batch and glycerol fed-batch phases, the temperature was maintained at 25 ºC, pH was controlled at 6, and dissolved oxygen (dO) was maintained at 20% of saturation atmospheric pressure by fixing the flow rate at 1 vvm and cascading agitation (800 to 1100 rpm). End of the batch phase was indicated by an increase in the dO measure that signaled the depletion of 40 g L^{-1} of glycerol. It was followed by an exponential glycerol feed at a rate of 0.08 h^{-1} (starting at 5.3 g L^{-1} h^{-1}). Between 17 and 18 h of feed were necessary to reach the target biomass concentration of 300 g L^{-1} of wet cell weight. After a 30 min of substrate starvation, the dO cascade was turned off, airflow rate was set to 0.2 vvm, and agitation was set to 800 rpm. After dO decreased to less than 1%, induction phase was initiated by adding a shot of 1% v/v of methanol (considering the total volume reached at the end of the glycerol fed-batch). Subsequent methanol shots were triggered by rapid increases in the dO profile, indicating methanol depletion (volume of shots was maintained constant during the induction phase). The induction was prolonged for 4 – 5 days.

3.2.7. Biomass determination

Three 1 mL broth samples were centrifuged in pre-weighed tubes at 16000 g, 3 min. After decanting the supernatant, the residual liquid in the tube was wiped off with a cotton swab. Resulting cell pellets were weighed to determine wet cell weigh. Alternatively, cell concentration was also measured by optical density at 600 nm (OD_{600}). A bibliographical correlation between OD_{600} and dry cell weight (DCW, g L^{-1}) of 0.25 (DCW / OD_{600} unit) was used (Chan, 2005).

3.2.8. α-galactosidase A activity assay

The method described previously in Chapter 2 was used to determine rhGLA activity in supernatant samples along the bioreactor cultivations.

3.3. Results and discussion

3.3.1. Production of GLA by methanol-limited induction strategy

As a model strain for the production of rhGLA in bioreactor, a selected clone producing the enzyme under the control of the AOX promoter was used (clone GLAwt_5.1, one of the first producer clones obtained from the expression assays described in chapter 2). The fermentation profile obtained is shown in fig. 3.5. As observed, it corresponds to the theoretical behavior (fig. 3.1). Glycerol batch phase ended after 24 h of culture, reaching 24.4 g L^{-1} DCW. During the glycerol fed-batch, which lasted 5 h, the cell concentration was incremented to 47.25 g L^{-1} DCW. Finally, the methanol fed-batch phase was prolonged for a total of 137 hours (with a partial bioreactor draining after 73 hours). The maximum cell and rhGLA concentrations were achieved at the end of the fermentation (155.8 g L^{-1} DCW and 2787 AU/mL, respectively).

Usually, the induction phase in the methanol-limited cultures lasts for 48 – 72 h, generally because the maximum working volume of the bioreactor is reached. However, the activity obtained at this point (137 hours of induction) was 348 AU/mL, lower than expected regarding the productivities obtained from the screening assays (Chapter 2). At this point of the culture, the working volume was reaching its maximum. Therefore, 1 L of culture was drained to allow the process to continue for 63.5 additional hours. During this time, the activity concentration increased to 2787 AU/mL (8 time-fold increase compared to the concentration reached before the partial drainage). Interestingly, the production profile showed an exponential behavior (observed in subsequent processes too, data not shown) in contrast with the linear cell growth measured. One possible explanation would be the decreased growth rate observed at the later stages of the fermentation, which could allow the culture to dedicate its metabolic resources to protein production instead of cell growth. As methanol was added to the culture at a constant flow

rate, the specific growth rate (μ) of the culture had to decrease gradually along the whole induction phase.

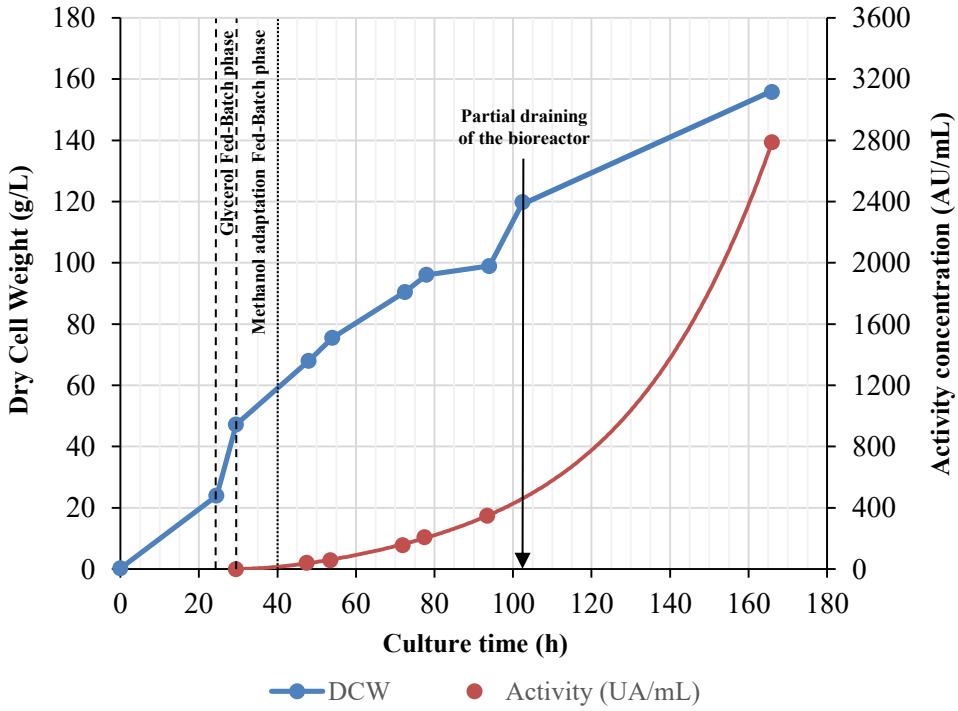


Fig. 3. 5. Fermentation profile of the GLA production in bioreactor following the methanol-limited strategy (clone GLAwt_5.1).

Even though the exponential profile in the activity concentration suggested a high process productivity in the final hours of the fermentation, the culture had to be stopped after 136.5 h of induction. At this point, the bioreactor was already evidencing serious problems to keep both temperature and dO set-points due the high cell concentration reached (155.8 g/L of DCW), as reflected by the maintained maximum stirring speed limit (1100 rpm) and the continuous supply of pure O_2. This behavior was also observed in additional fermentations carried out, and in one occasion the culture even collapsed, as indicated by a rapid decrease in the cooling and aeration demand, together with the sudden drop of all the

accumulated activity (likely due to the massive release of proteases to the medium). Moreover, the centrifuged biomass had a muddy appearance that was attributed to cell lysis. Probably, the continuous supply of pure O_2 contributed to a detrimental oxidative environment in the bioreactor, until a toxic boundary was reached. It has been described that autophagic processes can be related to peroxisomal proteins and components damaged by reactive oxygen species (ROS), generated through methanol oxidation during the induction phase (Vanz et al., 2012).

Furthermore, the amount of pure O_2 consumed in this strategy was higher than affordable, and together with the operational limitations experienced at the final stages of the cultures, the potential scalability of the process was in compromise. Therefore, an alternative strategy was explored to overcome the limitations of the methanol limited strategy.

3.3.2. Production of GLA by oxygen-limited induction strategy

As an alternative to the methanol-limited strategy, a novel O_2-limited strategy recently described (Berdichevsky et al., 2011) was tested. Even though some similar approaches based on O_2 limitation had been reported (Charoenrat et al., 2005; Trentmann et al., 2004), this methodology was attractive because of its simplicity, ease of implementation and reported scalability.

The clone GLAwt_5.1 was used again, in order to compare properly the outcome of both strategies (MLFB vs. OLFB). Figure 3.6 shows the fermentation profile obtained with the O_2-limited strategy, which was very similar to the expected (fig 3.2). This time, glycerol batch phase ended after 26.5 h of culture, reaching 20 g L^{-1} DCW. The glycerol fed-batch lasted 17.5 h, and the cell concentration was incremented to 71 g L^{-1} DCW. Finally, the methanol fed-batch phase was prolonged for a total of 169.5 hours. The cell concentration remained more or less

constant during the whole induction phase (oscillating between 77.25 and 69.25 g L^{-1} DCW). On the other hand, product concentration increased linearly for the first 115.5 hours of induction, reaching a plateau that was maintained for the last 54 hours of culture. The maximum rhGLA concentration was reached at the end of the fermentation (3901 AU/mL).

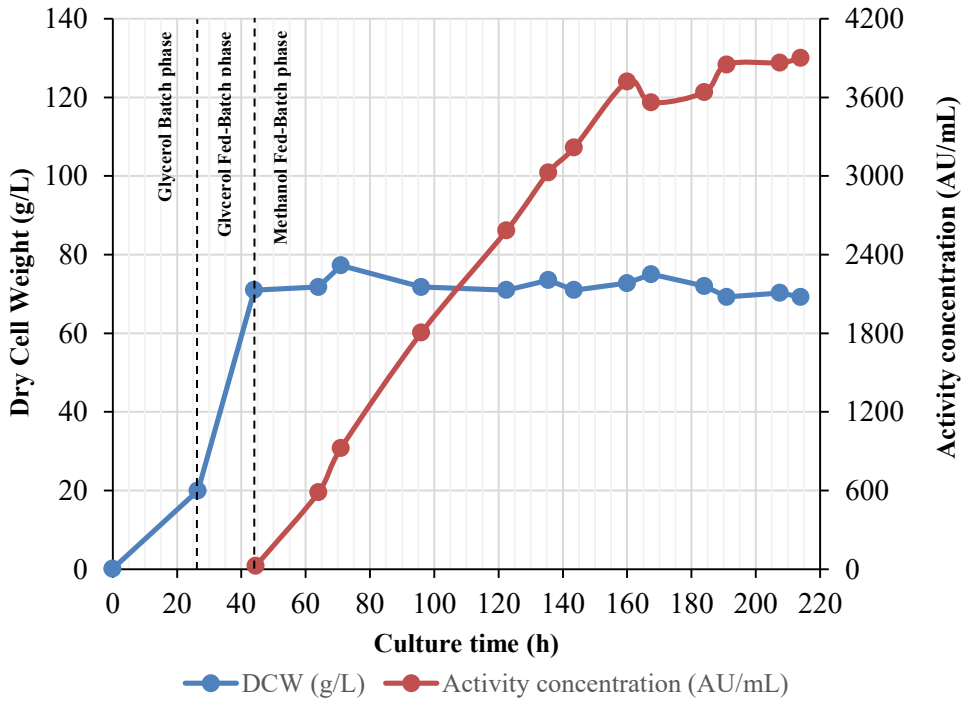


Fig. 3. 6. Fermentation profile of the GLA production in bioreactor following the O_2-limited strategy (clone GLAwt_5.1).

The maximum activity concentration reached represented a 40% increase compared to the maximum value obtained in the methanol-limited process (3901 AU/mL vs. 2787 AU/mL). In contrast with the previous strategy, where product concentration increased exponentially, the profile of rhGLA activity was linear (likely due to the stable cell concentration). As mentioned, the increase in rhGLA activity stopped almost completely after 115.5 hours of induction. Presumably, the

reason could be some nutrient limitation, potentially being nitrogen or phosphate. In standard procedures, nitrogen and phosphate supply are guaranteed by the addition of a base solution such as NH_4^+ or NH_3OH and a phosphate-containing compound in the media (like phosphoric acid, used in the pH control too). Because *P. pastoris* (and yeasts in general) tend to acidify the medium while growing, the need of base solution (and therefore nitrogen source) supply is always present. However, this strategy represents an exception mainly for two reasons. First, base and acid demands observed during the induction phase were almost null, which agrees with the very low growth rate. Secondly, the extension of the induction phase is notably long (4.8 days). These two factors can lead to the depletion of some critical nutrient essential for the protein expression. In the conditions experimented (strain, culture conditions, and media composition), the optimal duration of the induction phase was estimated to be 4.5 – 5 days (108 – 120 hours). As mentioned, the cell concentration remained virtually constant during the induction phase, with values fluctuating between 69 and 77 g/L of DCW. Predictably, the specific growth rate (μ) would be close to 0. However, a minimum growth had to be present as the cell concentration was maintained while the culture volume increased during the whole phase.

The use of the O_2-limited strategy resulted in important operational advantages were experienced with the O_2-limited strategy: the stirring speed required and the cooling demand were much lower, methanol consumption was reduced, there was no need for pure O_2 supply during the induction phase, and fermentations were more stable and reproducible.

According to the literature, one additional advantage of the methodology is the robust and reliable process scale-up based on the oxygen uptake rate (OUR) (Berdichevsky et al., 2014). Specifically, this parameter should be maintained constant between scales in order to obtain the same culture profile. Unfortunately, the efficient online measure of OUR requires the use of gas sensors in the bioreactor gas exhaust to determine the O_2 and CO_2 composition, which were not

available in the laboratory. Aiming to define an alternative "scalable" parameter, methanol specific consumption rate (q_S) was proposed. OUR, expressed as mmol O_2 L^{-1} h^{-1}, is directly related to the respiratory state of the culture. Considering the conditions during the induction phase, where methanol is the only carbon source available and whose consumption is directly coupled to the O_2 supply, a direct correlation of both parameters should exist. The relationship was evidenced by the dO spike observed every time a methanol shot was depleted (in fact, spikes were used as a trigger for the next shot). Therefore, the methanol profile during the induction phase was analyzed. As expected, the initiation of methanol consumption after the first shot was delayed for 2 – 3 hours due to the typical adaptation to methanol consumption (alcohol oxidase, responsible for the methanol oxidation into formaldehyde, is expressed by induction of its AOX promoter by methanol itself). Once adapted, the methanol consumption was resumed immediately after every shot. (fig 3.7).

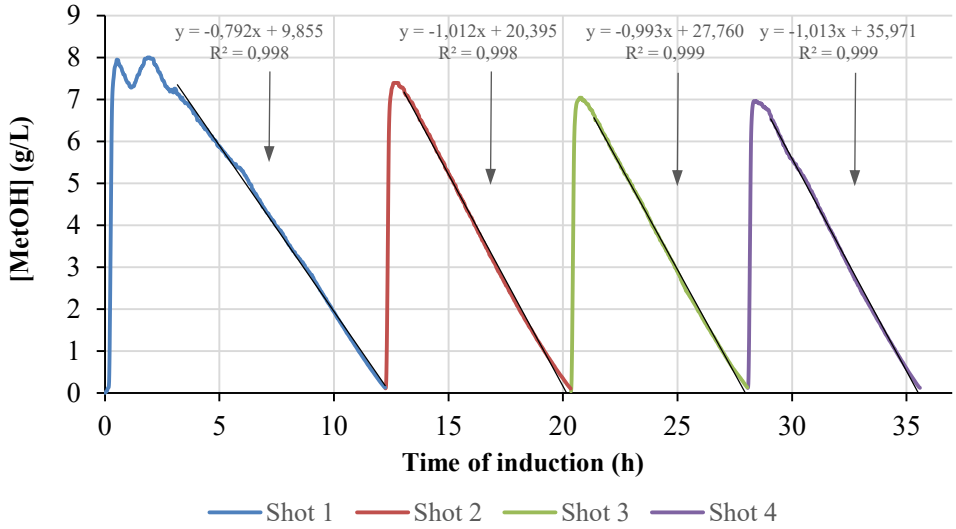


Fig. 3. 7. Methanol consumption profile during the induction phase of the O_2-limited strategy. The methanol concentration evolution of the first 4 shots are depicted.

In all cases, the decrease was linear, in line with the constant aeration flow and cell concentration. Considering that the volume of culture remains constant in every shot, the volumetric consumption rate was calculated by linearizing the methanol concentration drop of every pulse, and the specific consumption rate was calculated using the cell concentration at the end of every shot (table 3.1).

Table 3. 1. Methanol consumption rates (volumetric and specific) calculated for the first 4 shots during the induction phase.

MetOH Shot	**Substrate volumetric consumption rate [g MetOH L^{-1} h^{-1}]**	**Biomass concentration (X) [g DCW L^{-1}]**	**Substrate specific consumption rate (q_s) [mg MetOH g X^{-1} h^{-1}]**
1	0.792	71.75	11.04
2	1.012	77.25	13.10
3	0.993	74.25	13.37
4	1.013	71.75	14.12

As it can be seen, the methanol specific consumption rate (q_s) remained practically constant from the 2nd methanol addition on (13 – 14 mg MetOH g X^{-1} h^{-1}). Even though no scale-ups were attempted, this parameter could be used in principle as a reference to reproduce the cultivation profile at higher scales. Evidently, in such case some methanol measure would be necessary (e.g. by HPLC or methanol sensor), but represents an alternative to the use of O_2 and CO_2 measures in the gas exhaust. Specifically, the methanol specific consumption rate (as with the OUR measure) should be tuned during the first 3 – 4 pulses by fixing the aeration rate to 0.2 vvm (a constant value between scales) and adjusting the stirring speed as required according to the bioreactor volume and design. Once adjusted, it could remain fixed for the entire process.

The limitation of the methanol consumption rate by means of O_2 supply led to a profile that suggests the potential applicability of the strategy for any other non-fermentative carbon source. If true, this would be an interesting strategy also for constitutive protein expression of proteins (e.g. under GAP promoter). Under carbon source-limiting conditions, one of the limitations is the fast reaching of huge cell densities, mainly due to the higher substrate/biomass yields of preferred carbon sources. Instead, if the application of O_2-limiting conditions lead to a growth profile similar to that seen for methanol, the fermentations could be notably prolonged with all the associated operational strategies advantages, potentially leading to higher process productivities and lower costs. Therefore, it would be worthwhile to study the suitability of the strategy for the production of proteins under GAP promoter with a non-fermentative substrate such as glycerol (Macauley-Patrick et al., 2005). In case of using a fermentative carbon source such as glucose (Baumann et al., 2008), the effect of the metabolism shift under O_2-limiting conditions should be evaluated.

3.3.3. Comparison between MetOH-limited and O_2-limited strategies

Data obtained from the methanol-limited and O_2-limited fermentations were used to analyze in-depth both strategies. The use of the same clone, medium formulation and bioreactor setup enabled a reliable performance comparison (table 3.2). O_2-limited induction strategy clearly outperformed the methanol-limited one in all the parameters directly related to process productivity. Specifically, the total production was estimated to be 1.7 fold higher, while total methanol consumption was 60% lower. Moreover, process productivity was also increased (from 1.86 to 4.03 mg L^{-1} day^{-1}).

Table 3. 2. Performance parameters calculated from methanol-limited and O_2-limited cultivations of clone GLAwt_5.1 producing rhGLA. For the calculation of productivities, the total time of induction was taken. Green dots highlight those parameters where O_2-limited induction stands out on methanol-limited.

	Induction strategy		**Fold-change (O_2-lim. vs. MetOH-lim.)**
	MetOH-limited*[1]	**O_2-limited*[2]**	
Time of induction (h)	137 (partial draining at 73 hours)	115.5	0.84 x ●
Broth volume (L)	4.9	4.6	0.9 x
Supernatant volume (L)	2.5	3.6	1.4 x
Cell concentration (DCW, g/L)	155.8	72.8	0.47 x
Activity concentration (AU/mL)	3135	3720	1.2 x ●
rhGLA concentration (mg protein/L)	20.8	24.7	
Total production (mg protein)	52	88.9	1.7 x ●
End-point productivity (mg protein L^{-1} h^{-1})	$7.8 \cdot 10^{-2}$	0.17	2.2 x ●
End-point specific productivity (mg protein g X^{-1} h^{-1})	$5.0 \cdot 10^{-4}$	$2.3 \cdot 10^{-3}$	4.6 x ●
MetOH consumption (L)	2.2	0.85	0.4 x ●

Product/Substrate yield, $Y_{P/S}$ (g protein / g MetOH)	$2.4 \cdot 10^{-2}$	0.10	4.4 x ●

*[1] Parameters were calculated adding the enzyme obtained both in the partial bioreactor draining and at the end of the cultivation.
*[2] Parameters were calculated taking the point at which maximum activity concentration was reached in culture.

Apart from the improvement of performance parameters shown above, several additional advantages were observed with the O_2-limited induction over the MetOH limited strategy:

- Lower culture cooling demand, thanks to a moderate heat generation due to low methanol consumption rates.

- No need of pure O_2 supply (eventually, it can be necessary only at the end of the glycerol fed-batch phase), which improves the process economy and decreases the risk of operation.

- Induction phase shows a more stable and easy to control profile. Moreover, foam generation issues are minimized due to lower and constant stirring speed and aeration flow.

- Induction phase can be extended thanks to the lower methanol accumulation rate and to the process stability. This advantage can improve productivity of continuous processes, decreasing the frequency of dead times.

- Rational consumption of methanol, limited by the oxygen transfer rate to the culture. Overall, the total amount of methanol consumed is notably lower without compromising the culture productivity, resulting in better product/substrate yields ($Y_{P/S}$).

- Productivity would not be relevantly affected. In the case of GLA, it is even improved.

- Predictably, the product quality would be better (in terms of protein integrity and glycosylation profile homogeneity). Even though this was not studied for GLA, it has been reported in the literature (Berdichevsky et al., 2011).

- The low methanol consumption rate leads to significantly lower cooling requirements, allowing to decrease notably the temperature set-point during the induction phase (which is an interesting measure especially for those unstable or protease-sensitive proteins).

- Process scalability, based on the OUR or methanol specific consumption rate, is relatively simpler.

It is widely accepted that a minimum dO set-point over 20% had to be always guaranteed in *P. pastoris* bioreactor cultivations (especially during methanol induction) to avoid a hypoxic environment (Darby et al., 2012). However, it has been shown that lowering the dO set-point can have a clear impact in productivity, as shown by Baumann and coworkers for the production of an antibody fragment under the GAP promoter (Baumann et al. 2008). Noteworthy, and even under O_2-limiting conditions, cultivations were performed under relatively high aeration conditions: between 1 and 0.5 vvm and cascading agitation between 600 and 1200 rpm in a 40-L bioreactor. Similar parameters can be inferred from other publications where the use of OLFB was described: between 0.6 and 1.5 vvm, with agitation at 1000 rpm in a 10-L bioreactor for the production of Thai Rosewood β-glucosidase (Charoenrat et al., 2005), and 0.35-1.6 vvm cascading agitation between 300 and 1200 rpm in a 10-L bioreactor for the production of a single-chain antibody fragment (Trentmann et al., 2004). In all cases, better productivities

were obtained. In our process, aeration conditions were notably lower with fixed 0.2 vvm and 800 rpm in a 5-L bioreactor.

According to the results showed in this chapter and data reported in the literature, a dramatic decrease in the oxygen transfer could somehow induce a great shift in the cell behavior at the metabolic level and benefit the expression of the recombinant protein. A thorough study at the metabolic and proteomic level would be worthwhile to explain the observed profile and elucidate its differences compared to highly-aerated conditions. The outcomes could help to define a model to predict and maximize the productivity of *P. pastoris* bioreactor processes.

Considering all the advantages experienced and described above, the O_2-limited induction strategy became the standard methodology at Bioingenium for the production of a great number of recombinant proteins in *P. pastoris*, obtaining high productivities for most of them.

3.3.4. Production of GLA co-expressed with HAC1 under O_2-limited conditions

Once established the O_2-limited strategy as the preferred method for the production of rhGLA in bioreactor, a new available strain was used. Specifically, a clone co-expressing rhGLA (single-copy) with the HAC1p transcription factor, both under AOX promoter, was selected (clone GLAwt_5.1[$HAC1_{AOX}$]). As showed in chapter 3, this strain modification led to a substantial increase in the specific productivity of the enzyme (4 fold-change increase, with respect to the same clone in the absence of HAC1p over-expression). The primary objective of this experiment was to obtain larger quantities of enzyme from a single fermentation, allowing at the same time to evaluate the behavior of a different strain under the new conditions defined. Additionally, the results would confirm if different

productivity values observed in screening assays (50-mL tube scale) correlated with the strain performance in bioreactor.

The fermentation profile (fig 3.8) was almost identical to the one observed previously for clone GLAwt_5.1. Glycerol batch phase ended after 22 h of culture, reaching 19.25 g L^{-1} DCW. During the glycerol fed-batch, which lasted 18.5 h, the cell concentration was incremented to 72.75 g L^{-1} DCW. Finally, the methanol fed-batch phase was prolonged for a total of 143 hours. The cell concentration remained more or less constant during the whole induction phase (oscillating between 72.75 and 67.25 g L^{-1} DCW). On the other hand, product concentration increased linearly for the first 102 hours of induction, and from this point on a gradual decrease was observed. The maximum rhGLA concentration was 12228 AU/mL).

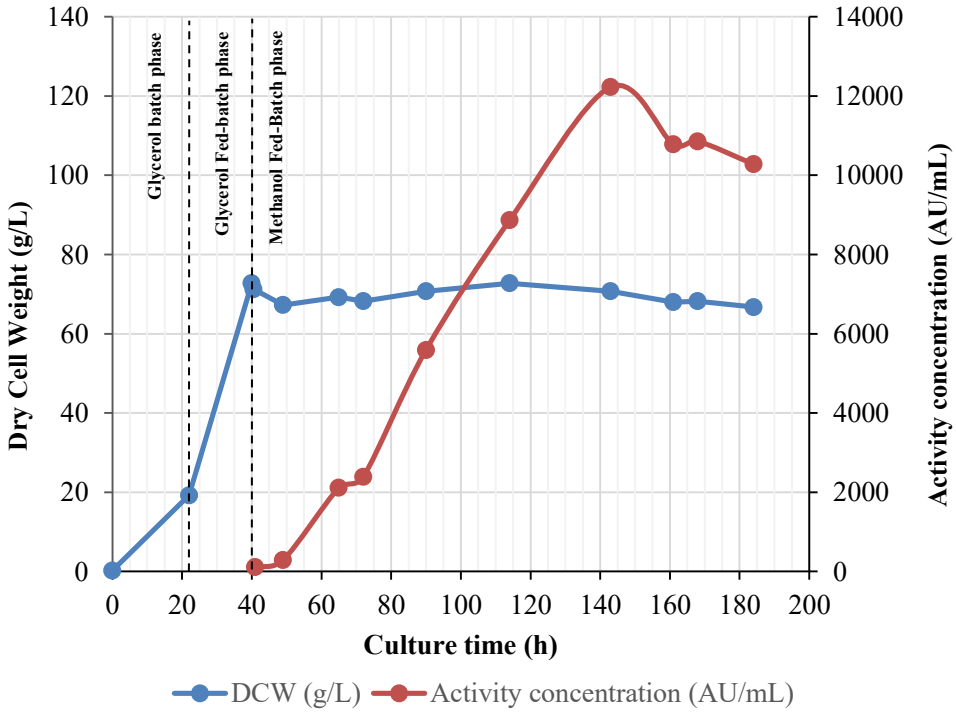


Fig. 3. 8. Fermentation profile of the GLA production in bioreactor following the O2-limited strategy (clone GLAwt_5.1[$HAC1_{AOX}$]).

The main difference observed for the clone used in this fermentation (GLAwt_5.1[HAC1$_{AOX}$]) was, of course, the amount of GLA activity accumulated. As mentioned, a maximum value of 12228 AU/mL was reached after 102 h of induction, which represents a 3.1 time-fold increase compared to the maximum activity concentration measured for the clone GLAwt_5.1 (lacking over-expression of HAC1p, cultivated using the same strategy). From this point on, the increase of rhGLA activity stopped and even decreased substantially. This time was similar to the one observed in the previous fermentation for clone GLAwt_5.1 (102 vs. 115.5 hours). The decrease could be explained by nutrient limitations after long induction phases, combined with the intrinsic volume increase of fed-batch processes (adding a dilution effect) and a certain level of proteolytic activity or enzyme inactivation. Again, ~110 hours of induction was considered to be the optimum time under the tested conditions.

The performance parameters were calculated from the fermentation and compared in detail with the GLAwt_5.1 clone, cultivated under O_2-limitning conditions too (table 3.3). The over-expression of HAC1p led to a 3.1x increase in the total enzyme production, while the specific productivity experimented a 3.9 fold-change increase compared to the clone expressing rhGLA alone (which is notably close to the difference seen at the 10-mL culture scale, 4 time-fold). Nevertheless, when specific productivities from 10-mL cultures and bioreactor were independently compared for each clone, a consistent decrease was seen in both cases for the bioreactor culture (2.5x decrease for GLAwt_5.1 and 2.6x decrease for GLAwt_5.1[HAC1$_{AOX}$]). This result suggests that the potential productivity of the strains scale might be improvable at the bioreactor by means of process optimization.

The results shown confirm that the O_2-limited induction strategy is an efficient methodology for the production of rhGLA in bioreactor. Moreover, a direct

correlation between the specific productivity values obtained in the expression assays and the strain performance in the fermenter could be seen. In fact, the O_2-limited induction profile mimics to the methodology used classically to carry out expression assays at 10-mL or shake flask-scale, where methanol is added in pulses once or twice daily and the consumption is driven by the air transfer capacity provided by the vessel shape and agitation speed.

Therefore, the cultivation of selected strains co-expressing HAC1p with 2 or 3 copies of GLA (which showed the best titers at the 10-mL screening assay scale) would likely yield even higher rhGLA titers than the one described in this paragraph for a single-copy strain.

Table 3. 3. Performance parameters calculated from O_2-limited cultivations of clones GLAwt_5.1 (producing rhGLA alone) and GLAwt_5.1[$HAC1_{AOX}$]. For the calculation of productivities, the total time of induction was taken. Green dots highlight those parameters where clone co-expressing rhGLA+HAC1 stands out on clone expressing rhGLA alone.

	Clone GLAwt_5.1*	**Clone GLAwt_5.1[$HAC1_{AOX}$]***	**Fold-change (rhGLA+HAC1 vs. rhGLA alone)**
Time of induction (h)	115.5	102	0.9 x
Broth volume (L)	4.6	4.3	0.9 x
Supernatant volume (L)	3.6	3.4	0.9 x
Cell concentration (DCW, g/L)	72.8	70.8	~ 1 x
Activity concentration (AU/mL)	3720	12228	3.3 x ●
rhGLA concentration (mg protein/L)	24.7	81.2	

Total production (mg protein)	88.9	276.1	3.1 x ●
End-point productivity (mg protein L^{-1} h^{-1})	0.17	0.63	3.7 x ●
End-point specific productivity (mg protein g X^{-1} h^{-1})	$2.3 \cdot 10^{-3}$	$8.9 \cdot 10^{-3}$	3.9 x ●
MetOH consumption (L)	0.85	0.82	~ 1 x
Product/Substrate yield, $Y_{P/S}$ (g protein / g MetOH)	0.10	0.34	3.4 x ●

* Parameters were calculated taking the point at which maximum activity concentration was reached in culture.

Regarding the bioreactor process itself, there are different measures that could be tested in order to improve both the productivity and the protein stability during the long induction phases of O_2-limited fed-batch processes. First, a different media composition could lead to better performance thanks to a rational macro- and micro-nutrient supplementation. The media used for the fermentations described in this report is the standard one detailed in the manual "Pichia Fermentation Process Guidelines" (Invitrogen), which was described for its use in methanol-limited fermentations. This media contains high salt concentrations, which could be detrimental for the production of some proteins due to its high ionic strength (Fickers, 2014). As an alternative, media compositions with reduced concentrations of salts (Gonçalves et al., 2013) or semi-defined media containing yeast extract and peptone (Potgieter et al., 2009) were shown to be efficient for the production of recombinant proteins in *P. pastoris*. Moreover, the addition of casamino acids is a specific media modification that could be used to prevent nitrogen limitation and/or to minimize protease activity (Wang et al., 2005). Lowering the temperature of the induction phase could be also an alternative measure to improve the product stability along the fermentation (Jahic et al., 2003). As the observed growing rates

during the induction are very low due to the O_2-limitation, a temperature of 20 ºC or even lower would not be, in principle, a limiting factor. Last but not least, the process parameters could be optimized specifically for the production of alfa-galactosidase using a DoE (design of experiments) approach. The O_2-limited process parameters described in the literature (Berdichevsky et al., 2011) and used in the experiments described in this chapter for the production of alfa-galactosidase A were originally optimized for the cultivation of glycoengineered *P. pastoris* strains in semi-defined medium for the production of recombinant monoclonal antibodies. As no other similar O_2-limited strategy has been found in the literature to date, it is unknown if different strains, culture media and/or heterologous protein require different optimal process parameters, including OUR and initial cell concentration for the induction phase.

3.4. Conclusions

- A recently described fermentation strategy, based on O_2-limitation during the induction phase, has been successfully implemented for the efficient production of human α-galactosidase A.

- O_2-limited induction showed better productivity yields than the classical methanol-limited induction strategy, as well as important operational advantages.

- The differential productivity profile observed for two different strains at the clone screening-scale (10-mL) was conserved using the O_2-limited strategy described in this chapter. Therefore, one would be able to reproduce at bioreactor scale the clonal productivity differences observed at shake flask scale.

Chapter 4

Purification and characterization of human recombinant α-galactosidase A

C h a p t e r 4

PURIFICATION AND CHARACTERIZATION OF HUMAN RECOMBINANT α-GALACTOSIDASE A

4.1. Introduction

After the study of strain improvement for the production of rhGLA in *Pichia pastoris* and the development of strategies for its overproduction in bioreactor, the methods developed for the purification and characterization of the recombinant enzyme are described and discussed in this chapter.

Every protein purification process can be divided in early downstream (product recovery) and late downstream (chromatographic purification). While the early downstream phase involves usually the separation of the target protein from the cellular fraction, its concentration and the consequent buffer exchange, the late downstream is classically performed by chromatographic methods. In turn, the chromatography-based purification stage can be divided in different steps: capture, intermediate purification and polishing, having each one of these different objectives within the process that will be discussed later in this chapter (Strategies for protein purification Handbook, GE Life Sciences). Performing *in vitro* and *in vivo* studies often require high amounts of the protein to be tested, ranging usually from milligrams to a few grams. Anticipating this necessity for forthcoming assays, a chromatographic process was designed based on a classical 3-step purification strategy, allowing to obtain high amounts of pure rhGLA previously produced in a lab-scale bioreactor. In order to achieve this objective, the full scalability of the process was taken as the central premise for the selection of the column dimensions, chromatographic media and buffer formulation.

Additionally, affinity chromatography procedures were set-up for the purification of different variants of rhGLA in a single step, taking advantage of the high selectivity provided by the type of chromatographic media. Specifically, a galactose (substrate)-affinity media was used to purify the active variants of rhGLA seeking for the determination of differences between their specific activities. On the other hand, an immobilized metal ion affinity chromatography (IMAC) was the choice to perform the purification of the His-tagged version on rhGLA ($rhGLA_{C\text{-}HIS}$). The aim was to define a simple process to purify small amounts of enzyme for further characterization, which would be useful to characterize the glycosylation patterns of future modified *P. pastoris* strains.

Finally, methods were defined for the characterization of the purified enzyme, with the aim to describe its properties when produced in a wild type *P. pastoris* strain (i.e. NRRLY-11430). Therefore, the starting point for future strain modifications (especially regarding the glycosylation profile) was established.

4.2. Materials and methods

4.2.1. Sample recovery from bioreactor and conditioning

Culture broth from bioreactor cultivation was centrifuged at 6000 g, 20 min, 4 ºC. 1 L of supernatant were mixed with 0.1 volumes of storage buffer (110 mM NaHPO4, 0.55 mM PMSF), and stored at -20 ºC. For secondary clarification of the bioreactor harvest, two types of microfiltration techniques were tested: tangential flow filtration (Pellicon 2 Mini GVPP 0.22 µm 0.1 m^2, Millipore, #P2GVPPV01), or depth filtration (Millistak+® HC Pod Depth Filter, X0HC media 0.054 m^2, Millipore, #MX0HC054H1). The filtered sample was then ultrafiltered (Pellicon 3 cassette 10 kDa 0.11 m^2, Millipore, #P3B010A01). By this diafiltration step, supernatant was concentrated to 250 mL and buffer exchanged to 10 mM Bis-tris, pH 6 (8 ºC). Finally, the sample was degasified prior to the first chromatography step. All reagents for buffer preparation, including those used in chromatography steps, were purchased from Sigma-Aldrich.

4.2.2. Chromatography equipment

All the chromatographic experiments detailed below were performed with an ÄKTA Purifier 10 FPLC system (GE Life Sciences) equipped with a sample application pump and an automated fraction collector (fig. 4.1). During chromatography runs, protein concentration was monitored by absorbance at 280nm and salt concentration by conductivity (measured in mS/cm). All chromatographic steps and sample handling were performed at 8 ºC. All the buffers described below were filtered (0.22 µm) and degasified.

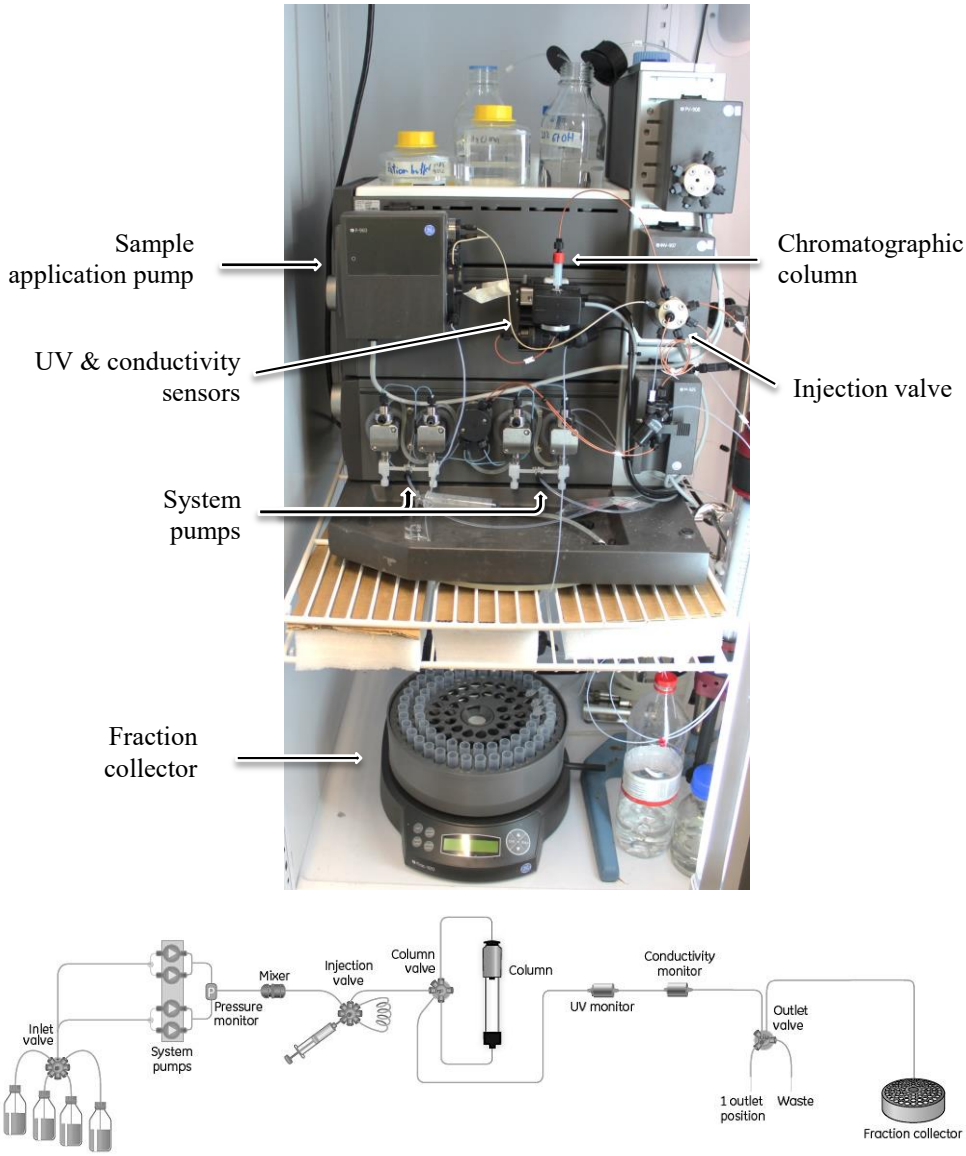


Fig. 4. 1. ÄKTA Purifier 10 FPLC system used for the set-up of chromatographic steps for the purification of rhGLA. The main parts are arrowed and the system scheme is depicted.

4.2.3. Purification of un-tagged rhGLA by chromatography

4.2.3.1. Step 1 – Anion exchange chromatography (AIEX)

The conditioned sample (clarified, concentrated and diafiltered against 10 mM Bis-tris, pH 6 at 8 ºC) was applied to a Capto Q resin. In the set-up experiments, a HiScreen Capto Q (4.7 mL, GE Life Sciences, #28-9269-78) was used. For the purification of the whole bioreactor volume, sample was applied to 50 mL of Capto Q chromatographic media (GE Life Sciences, #17-5316-10) packed into an XK26/20 column (GE Life Sciences, #28-9889-48). In all the chromatographic runs, binding buffer was 10 mM Bis-tris, pH 6 at 8 ºC, and the elution buffer had the same composition plus 1 M NaCl.

In first place, three runs were performed to optimize the elution conditions. In all of them, the column (HiScreen Capto Q) was equilibrated with 5 column volumes (CV) of binding buffer and 100 mL of sample were subsequently loaded at 1.2 mL/min. Then, the column was washed with 3 CV of binding buffer and the elution was started. During the elution phase, 1 mL aliquots were taken. The three experiments differed in the elution profile. In the first one, a 20 CV linear gradient was applied for the elution of rhGLA (from 0 to 1 M NaCl). In the second, a step-wise elution scheme was followed, and 0.05, 0.1, 0.15, 0.3, 0.5 and 1 M NaCl steps were consecutively applied. Finally, in the third run the elution steps were slightly adjusted to 0.1, 0.2, 0.3 and 1 M NaCl. In both cases, each elution step was prolonged for 3 CV. An additional experiment with the HiScreen Capto Q column was performed, in order to determine the breakthrough (saturation) point. In this case, the same process parameters were used, but 250 mL of conditioned sample were applied to the column. The flow through was collected in 10 mL aliquots for further activity analysis.

For the purification of a whole bioreactor batch, the sample was loaded in the XK26/20 column packed with Capto Q resin. Therefore, 580 mL of clarified, concentrated and buffer-adjusted supernatant was loaded at 3 ml/min on the Capto

Q column previously equilibrated with binding buffer. The column was washed with 3 CV of the same buffer, and the enzyme was eluted applying a linear gradient from 0% to 20% (200 mM NaCl) of elution buffer in 1 CV followed by a step at 200 mM NaCl for 3 CV. Finally, the column was washed with 5 CV of elution buffer. 10 mL fractions were collected, and those with enzyme activity were pooled (189 mL).

4.2.3.2. Step 2 – Hydrophobic chromatography (HIC)

A HiScreen Capto Butyl (4.7 mL, GE Healthcare, #28-9924-73) was used for all the experiments detailed below. In all cases, the samples were conditioned adjusting their salt concentration to 0.66 M $(NH_4)_2SO_4$, by adding dropwise a concentrated solution of 4 M $(NH_4)_2SO_4$. For the chromatographic runs, the column was equilibrated with 5 CV of binding buffer (10 mM MES-Tris, 0.66 M $(NH_4)_2SO_4$, pH 5.6 at 8 ºC). Then, the sample was loaded at 1.2 mL/min. The column was washed with 3 CV of binding buffer, and the enzyme was eluted with a linear gradient from 0% to 100% of elution buffer (10 mM MES-Tris, pH 5.6 at 8 ºC) in 14 CV. 3 mL fractions were collected, and the column was finally washed with 5 CV of elution buffer.

In the first experiment, aimed to validate the parameters specified above, 60 mL of pooled active sample from HiScreen Capto Q chromatography (AIEX), already adjusted to 0.66 M $(NH_4)_2SO_4$, was purified. For the purification of a whole bioreactor batch, the 189 mL pooled from the Capto Q (packed in XK26/20) were adjusted to 0.66 M $(NH_4)_2SO_4$ and loaded in 2 runs into the HiScreen Capto Butyl, in order to avoid the saturation of the column. In both cases, fractions with enzymatic activity were pooled and dialyzed overnight at 4 ºC against 5 L of stability buffer (25 mM sodium phosphate monobasic, 150 mM NaCl, pH 6.0) using a 12-14 kDa dialysis membrane (Spectra/Por2, 12-14 kDa, 45 mm, Spectrum Laboratories, #132680).

4.2.3.3. Step 3 – Gel filtration (GF) or Size exclusion chromatography (SEC)

The third chromatographic step was performed with a HiLoad Superdex 200 Prep Grade (GE Life Sciences, #28-9893-35). The dialyzed enzyme solution obtained from the HIC step was concentrated to 8.5 mL by centrifugal ultrafiltration (Vivaspin 20, 10 kDa, Sartorius, #VS2001). The sample was divided in 3 fractions of 3, 3 and 2.5 mL and loaded in 3 different runs to SEC column. An isocratic profile was applied to the chromatography, using a single mobile phase along the process (25 mM sodium phosphate monobasic, 150 mM NaCl, pH 6.0) with a flow rate of 0.5 mL/min. 2 mL fractions were collected, and the ones with enzymatic activity were pooled and concentrated by centrifugal ultrafiltration to 10.5 mL (Vivaspin 20, 10 kDa, Sartorius, #VS2001).

For long-term preservation of samples, 0.05% Tween 20 was added. Then, it was separated into 1 mL aliquots and stored at -20ºC.

4.2.4. Purification of rhGLA variants by galactose-affinity chromatography

For the purification of rhGLA variants by galactose-affinity chromatography, 5 mL of D-Galactose agarose (ThermoFisher, #20372) packed into a XK16/20 column (GE Life Sciences, #28-9889-37) were used. Samples were obtained from shake flask cultures of selected clones in Chapter 2 (100 mL, 72 h of induction with methanol). After centrifugation at 6000 g, 20 min, supernatants were dialyzed overnight against binding buffer (100 mM acetate-acetic acid, pH 4.6). Then, samples were filtered (0.45 μm) and loaded into the column previously equilibrated with 5 CV of binding buffer. After sample loading, the column was wash with 3 CV of binding buffer. For the set-up of the elution conditions with rhGLA,

different conditions were applied step-wise (with 3 CV for each condition: 1. pH was increased to 5.8 and 0.1 M galactose was added as a competitor, 2. Increase to 0.5 M Galactose (to improve the competitive conditions inside the column), 3. Increase to pH 8 (to reach a value where GLA is inactivated), and 4. Application of 0.1 M glycine, pH 2.3 (CiP (Cleaning in Place) buffer, according to the instructions supplied by the resin manufacturer). In order to stabilize the enzyme eluted, aliquots (0.9 mL) were collected in tubes were preloaded with stability buffer 10x concentrated (250 mM sodium phosphate monobasic, 1.5 M NaCl, pH 6.0).

For the purification of all the rhGLA variants (rhGLA, rhGLAΔKDLL, $rhGLA_{C\text{-}HIS}$ and $rhGLA_{C\text{-}RGD}$), the enzyme was eluted with 0.5 M galactose, pH 5.8. The active elution fractions were pooled and concentrated by centrifugal ultrafiltration to ~0.5 mL (Vivaspin 20, 10 kDa, Sartorius, #VS2001) for sample analysis.

4.2.5. Purification of $rhGLA_{C\text{-}HIS}$ by immobilized-metal affinity chromatography (IMAC)

A HisTrap FF column (1 mL, GE Life Sciences, #17-5319-01) was used for the purification of the $rhGLA_{C\text{-}HIS}$ variant. Firstly, an assay to define the proper sample conditioning was carried out. Samples were obtained from shake flask cultures of a selected clone in Chapter 2 producing the mentioned variant (100 mL, 72 h of induction with methanol). After centrifuging the cultures at 6000 g, 20 min, three aliquots of supernatant (30 mL each) were conditioned differentially by adding dropwise 5x concentrated binding buffers (100 mM sodium phosphate, 2.5 M NaCl, 100 mM imidazole) adjusted to different pH: 7.5, 7 or 6.5. Finally, samples were filtered (0.45 μm). For each run, the HisTrap FF column was equilibrated with 5 CV of binding buffer (20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole). Samples were then loaded at 1 mL/min and, after a washing step with 3 CV of binding buffer, eluted in a single-step with elution buffer (same as binding,

but with 500 mM imidazole). For every run, the pH of buffers was adjusted according to the loaded sample (7.5, 7 or 6.5). Eluted fractions (0.9 mL) were collected in tubes pre-loaded with 10x stability buffer (250 mM sodium phosphate monobasic, 1.5 M NaCl, pH 6.0), and stored at -20ºC until analysis.

After selecting pH 6.5 as the most convenient pH, an optimization assay was made by testing the effect of additives in the buffers. A new sample was obtained by shake-flask culture as explained above. The supernatant was divided in three aliquots (~30 mL each), which were conditioned differentially by adding dropwise 5x concentrated binding buffers (100 mM sodium phosphate, 2.5 M NaCl, 100 mM imidazole, pH 6.5) containing different additives: 10% glycerol, 0.1% Tween 20, or without additive. Then, the samples were filtered (0.45 μm) and purified according to the procedure detailed above (in each run, binding and elution buffers contained the corresponding additive).

4.2.6. Analytic procedures for sample characterization

4.2.6.1. SDS-PAGE and Western-Blot

Purified samples analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were loaded in 12% acrylamide gels and ran in a Mini-Protean II (BioRad). The standard settings recommended by the manufacturer and the procedure described by Laemmli et al. (1970) were followed. Samples were prepared by mixing with 5x Lane Marker Reducing Sample Buffer (ThermoFisher, #39000), and heating at 95ºC, 10 min. 5 μL of Spectra™ Multicolor Broad Range (ThermoFisher, #26634) was used as the protein ladder. Gels were stained with Coomassie Blue (either with Brilliant Blue R Concentrate, Sigma-Aldrich, #B8647, or with PageBlue™ Protein Staining Solution, ThermoFisher, #24620) or by Silver staining (PageSilver Silver Staining Kit, ThermoFisher, #K0681).

For Western-blot analysis, unstained gels were transferred to a nitrocellulose membrane (ThermoFisher, #88018) using a Semy-dry Trans-Blot device (BioRad), according to the manufacturer specifications. Membranes were blocked overnight in blocking buffer (phosphate buffered solution (PBS), pH 7.4, 0.1% Tween 20, 2% skim milk). Then, membranes were incubated with primary antibodies for 1 h. A rabbit polyclonal anti-GLA (SantaCruz, #sc-25823) was used for the detection of all rhGLA variants, while a mouse monoclonal anti-6xHis tag (ThermoFisher, #MA1-21315) was used for $rhGLA_{C\text{-}HIS}$ detection. In both cases, primary antibodies were diluted 1:1000 in blocking buffer. After 3x washing steps of 10 min (in washing buffer, same as blocking buffer without skim milk), membranes were incubated with the secondary antibody (goat anti-rabbit IgG (H+L), HRP conjugate, ThermoFisher, #G-21234, or goat anti-mouse IgG (H+L), HRP conjugate, ThermoFisher #G-21040). Secondary antibodies were diluted 1:20,000 in blocking buffer. After the incubation, membranes were washed again 3x 10 min in washing buffer. Finally, membranes were incubated with SuperSignal™ West Pico chemiluminescent substrate (ThermoFisher, #34078), and used for developing of photographic films.

4.2.6.2. Protein quantification

The enzymatic activity of samples was analyzed according to the α-galactosidase A activity assay described in chapter 2 (materials and methods paragraph).

Total protein concentration was determined by Bradford (ThermoFiser, #23200) assay for non-purified and partially purified samples. Absorbance was measured in 96-well pates at 595 nm in a Powerwave XS spectrophotometer (Biotek). For pure rhGLA samples, concentration was measured by UV_{280nm} absorbance in a quartz cuvette. For the protein concentration calculation, an extinction molecular

coefficient (ε) of 111.100 M^{-1} cm^{-1} was inferred from its peptide sequence with the ProtParam tool (www.expasy.org).

4.2.6.3. Protein identification by peptide mass fingerprinting and molecular weight determination by mass spectrometry

Both analyses were performed by the Proteomics Platform of the Scientific Park of Barcelona, according to the procedures described below. In both cases, a purified rhGLA sample was previously ran on a SDS-PAGE and the gel was Comassie Blue stained. The gel band corresponding to rhGLA was cut and submitted for analysis.

For the analysis by peptide mass fingerprinting, the band was washed with 50 mM NH_4HCO_3 and acetonitrile. Then, it was reduced with 10 mM DTT for 30 min, 56 ºC, and alkylated with 55 mM iode acetamide for 30 min in darkness. Finally, the sample was digested with 100 ng of porcine trypsin (Trypsin GOLD, Promega, # V5280) at, 37 ºC, overnight. The resulting peptides were extracted from the gel matrix through 5% formic acid and acetonitrile, and were dried with a centrifugal evaporator. For the mass spectrometry analysis, the digested peptides were resuspended in 5 µL of 0.1% trifluoroacetic acid. Then, a mixture of 0.5 µL of resuspended sample and 0.8 µL of α-ciano-4-hydroxycinamic acid (CHCA) diluted in 50% acetonitrile/H2O mQ, 0.1% trifluoroacetic was prepared and deposited onto a MALDI plate and was left to air dry before being analyzed. A MALDI-TOF/TOF - Proteomics Analyzer 4700 (Applied Biosystems) was used for the analysis of the samples with a laser power of 3000-4500, and the data was acquired according to the MS Reflector Positive method. The data from the 7 most intense fragmented ions were analyzed. The MS and MS/MS data were used to perform a search in on-line MASCOT against the NCBI nr database. The parameters used for the search were trypsin as the enzyme used for the sample digestion, 2 missed cleavages, carbamidomethylation of cysteine as a fixed modification, oxidation of methionine

and pyro-Glu (N-term Glutamine) as a variable modification, and a peptide tolerance of 50 ppm and 0.25 Da (respectively for MS and MS/MS spectra).

The molecular weight was determined by MALDI-TOF. The protein contained in the gel band was extracted and desalted, and 0.5 to 4 µL were deposited onto a MALDI plate and was left to air dry before being analyzed. The matrix solution was 10mg/mL of α-ciano-4-hydroxycinamic acid (CHCA) diluted in 70% acetonitrile/H2O mQ, 0.1% trifluoroacetic acid. A MALDI-TOF/TOF - Proteomics Analyzer 4700 (Applied Biosystems) was used for the analysis of the samples with a laser power of 3000-5000, and the data was acquired according to the MS Linear Positive method. The molecular weight range analyzed was 2000-80000 m/z.

4.2.6.4. Protein size determination by dynamic light scattering (DLS)

The presence of aggregates of rhGLA on the purified sample was determined by dynamic light scattering (DLS) with a Zetasizer Nano ZS (Malvern Instruments). 30 µL of sample were introduced into a Low-volume quartz batch cuvette (ZEN2112, Malvern Instruments) for molecular size determination. As a disaggregant substance, different concentrations of Tween® 20 were added to the sample (Sigma-Aldrich, #P1379): 0.002%, 0.01% and 0.02% (v/v). Zetasizer Software (version 6.32, Malvern Instruments) was used for data analysis.

4.2.6.5. Glycan analysis

The analysis was performed by the Department of Analytic Chemistry of the University of Barcelona, according to the procedures described below. After reducing and alkylating a purified sample of rhGLA, it was subjected to a tryptic

digestion. Then, peptides and glycopeptides were analyzed by capillary liquid chromatography coupled to high-resolution mass spectrometry with fly-time analyzer (μLC-TOF-MS). The average mass for each peptide containing from 1 to 20 mannose residues was calculated and compared to glycan standards.

4.3. Results and discussion

4.3.1. Early downstream

Prior to the chromatographic steps, the culture broth obtained in the bioreactor must undergo through several steps collectively known as the early downstream. Because rhGLA was produced extracellularly; supernatant was the culture fraction of interest and was obtained by discontinuous centrifugation. Even though this type of centrifugation is practical only for the processing of batches up to 5-10 L, yeast cells are suitable for continuous centrifugation, which could be a good solution for eventual pilot-scale productions.

The second step was a microfiltration stage, used to remove particles that can potentially damage the chromatographic media packed within the columns, which can include, between others, remaining cells, cell debris, protein aggregates ad precipitated salts. Tangential flow filtration (TFF) was the first choice, because it generally allows the processing of large amounts of sample. However, the yield obtained with this technique was lower than expected. Therefore, depth filtration (DF) was tested alternatively. A comparison between both techniques was carried out to evaluate their performance for the microfiltration of rhGLA supernatant samples (fig. 4.2). DF gave a better yield in terms of activity recovery (80%) compared to TFF (60%). However, the most remarkable advantage of using DF was seen after analyzing the later ultrafiltered (buffer exchanged and concentrated) sample: unexpectedly, the activity of this fraction increased to a 98% of the initial supernatant activity value, which means a great overall early downstream yield. This incremental effect was not seen with TFF. A plausible reason for this phenomenon is the presence of small enzyme aggregates in the supernatant sample.

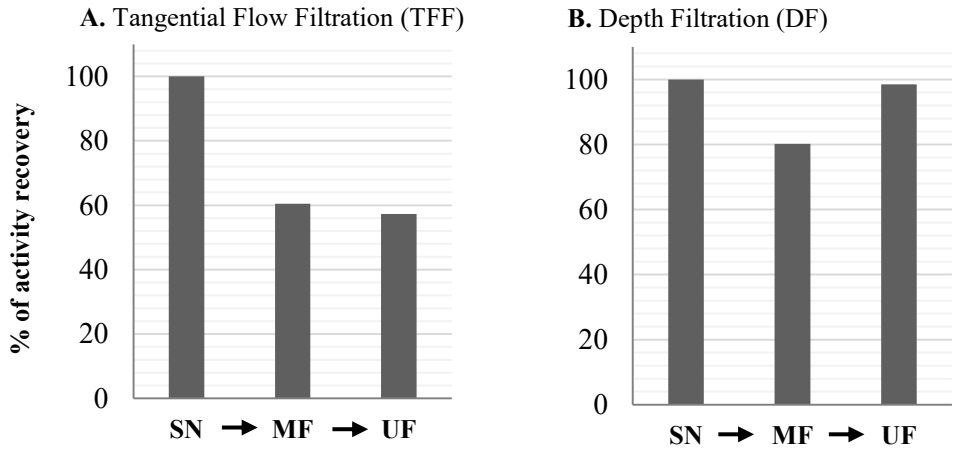


Fig. 4. 2. Recovery yield comparison between using (A) tangential flow filtration (TFF) and (B) depth filtration (DF) for the microfiltration of rhGLA bioreactor supernatants. SN: supernatant. MF: microfiltered sample. UF: ultrafiltered sample.

In later stages of pure product characterization, as will be discussed at the end of this chapter, it was confirmed that the produced rhGLA tends to form active reversible aggregates. In fact, the high ionic strength of the culture media used in bioreactor could even promote aggregation of those proteins that are prone to form clusters. Such effect would explain, firstly, the decrease in the activity measures after TFF. Assuming that this type of filter is more efficient in retaining aggregates, part of the product would not be able to cross the membrane. This was confirmed by the analysis of the TFF retentate fraction, which showed high activity values confirming that the enzyme was not only retained, but also concentrated (4x with respect to the original supernatant). Secondly, the aggregation hypothesis would give an explanation to the increase in the activity experienced by the ultrafiltered samples after DF. If these small aggregates are able to cross through the filter

(possibly due to the small differences in nominal pore size, 0.2 µm of TFF vs. 0.5 µm of DF), activity could be later recovered after altering both the product concentration and its buffer in the ultrafiltration process.

Ultrafiltration was the technique selected for the last recovery step. This type of filtration, which is tangential by definition, allowed in first place the concentration of the previously clarified sample, and secondly, its buffer exchange by successive steps of dilution (with the desired buffer) and concentration. This buffer exchange was essential to adjust the sample to the conditions required in the first chromatographic step. A 12-14 kDa pore-size ultrafilter cassette was used because at the time was the only one available on-site. However, manufacturers often recommend to use 50% of the target protein, which means that assuming an approximate rhGLA size of 100 kDa (corresponding to the glycosylated enzyme dimer), a 50 kDa pore-size cassette would have been a better choice to reduce an undetermined amount of low Mw contaminant proteins.

4.3.2. 3-step chromatographic process for the purification of rhGLA

A chromatographic process based on the classical 3-step set-up process (Capture + Intermediate Purification + Polishing) was designed for the purification of rhGLA. For each stage, a different technique was chosen according to the suitability of the different chromatographic media available. Specifically, IEX (Ion Exchange Chromatography) was used for the capture step, HIC (Hydrophobic Chromatography) for the intermediate purification step and GF (Gel Filtration) for the polishing step.

It is important to underline that this purification process was defined using the untagged rhGLA as the model protein. For this reason, eventual purifications of the different enzyme variants described in previous chapters may require further tuning of different parameters (such as ionic strength or pH) due to the small but

potentially relevant differences that exist between the amino acidic sequences of their terminal regions. However, the same purification scheme could be in principle used as a starting approach.

4.3.2.1. Capture step: anion exchange chromatography

Capto Q was the selected resin for the capture step. Capto base matrix, which is based on a mixture of highly rigid agarose and dextran, is characterized for its high capacity, pressure/flow properties, chemical stability, scalability and low cost, according to the manufacturer. Therefore, is especially suited the objectives pursued in capture chromatography steps: process speed and resin capacity (fig. 4.3). A quaternary amine ligand ($-N^{+}(CH_3)_3$) gives the strong anion exchanger specificity to the Capto Q matrix, which means that is able to maintain its positive charge (and thus its function) over a wide pH range.

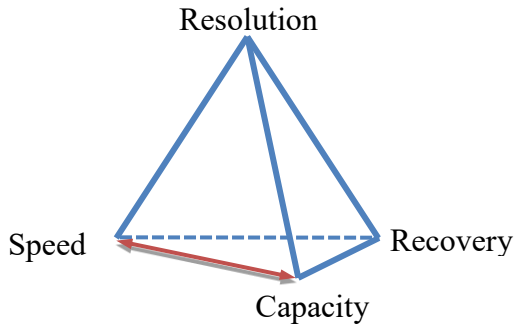


Fig. 4. 3. Representation of the balance between speed, capacity, resolution and recovery in chromatography purification steps. The objective of Capture steps is a good balance between speed and media capacity.

It is often recommended to perform anion exchange chromatography (AIEX) at a pH between 0.5 units higher than the pI of the protein of interest (to give it enough negative net charge and thus ensure its binding) and 1 unit higher than the pI (to maintain a minimum specificity, avoiding the binding of high amounts of contaminant proteins). For this reason, and given that GLA has a theoretical pI of

5.14, pH 6 was the chosen pH for the tests made with Capto Q media. Also, this pH allowed to work inside the pH stability range of GLA (between 4 and 7) (Ishii et al., 2000). Finally, Bis-Tris was the buffer selected due to its compatibility with anion exchange media and pH buffering capacity range (pH 6.0 – 7.0).

There are two main strategies to perform the elution step in ion exchange chromatography (IEX). The most extended one consists of an increase of salt concentration (usually NaCl) in the elution buffer, either by applying a linear or a step-wise gradient. The increase of a counter-anion concentration (Na^+ for CIEX and Cl^- for AIEX, in case of using NaCl) gradually displaces the proteins bound to the column, competing for the binding sites in the column. An alternative is the pH elution strategy, which is based on the change of pH in the elution buffer (again by applying a linear or step-wise gradient) with the aim to change the net charge of the protein of interest. The differential pI of the proteins bound to the column determines their elution profiles in a pH gradient. In this work, the salt elution strategy was used.

In the first experiment, a linear gradient was applied for the elution. The chromatogram showed relevant information related with the chromatography performance (fig. 4.4). Firstly, the low UV_{280nm} signal in the binding stage suggests that many of the proteins in the loaded sample bind to the column. Secondly, the activity measures of the elution fractions showed a wide distribution of rhGLA over the gradient (between 0.1 and 0.4 M NaCl in the elution buffer). For this reason, it is unlikely that the clear peak seen at the beginning of the elution gradient could correspond only to rhGLA. Finally, the UV_{280nm} signal profile of the elution step indicated a poor resolution capacity of the Capto Q chromatographic media, as no isolated peaks could be seen. However, this is a frequent and accepted drawback in capture steps, where high bead size resins are used and the main objectives are the concentration of the desired protein and discrimination of main contaminant species.

The activity measure of the elution fraction pool in comparison with the loaded sample activity showed a 90% recovery yield (table 4.1), suggesting that not only the conditions tested but also the chromatographic media used was efficient in the preservation of the enzymatic properties of rhGLA. Moreover, the loss of activity in the flow through (FT) and wash-out (WO) fractions were acceptably low (4% of the loaded activity).

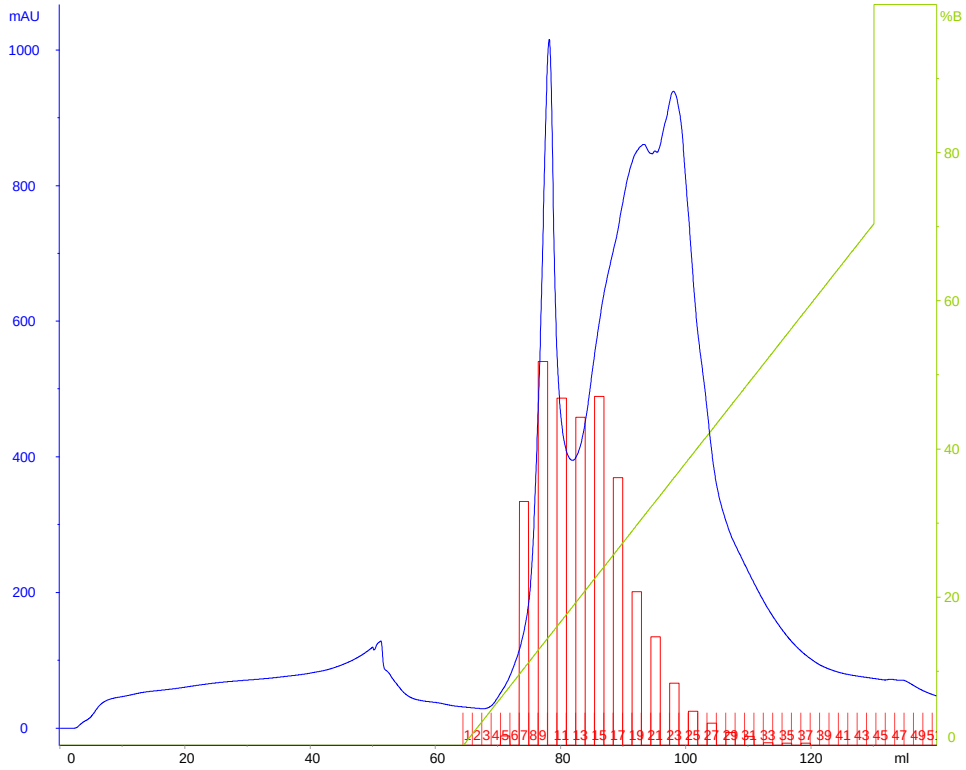


Fig. 4. 4. Chromatogram corresponding to the whole AIEX capture purification step. A linear gradient elution was performed. Blue line: UV_{280nm} signal. Green line: % of elution buffer (being 100% equal to 1 M NaCl). Red bars: GLA activity (arbitrary units).

Table 4. 1. Total activity and percentage recovered (with respect to the loaded sample) in the AIEX capture purification step eluted by linear gradient.

Fraction	Total AU (nmols MU/h)	% recovered
Load	105502	
FT+WO	4043	4
Elution	94832	90

As previously anticipated by the UV_{280nm} signal of the chromatogram and the associated activity profile, the purification factor observed was low: the fractions with high activity concentration (F7 → F17) contained high amounts of contaminant proteins too (assessed by SDS-PAGE, data not shown). One possible cause for the broad elution profile of rhGLA could be the presence of a heterogeneous glycosylation profile in rhGLA. It is known that protein glycosylation structures can have a direct impact on isoelectric point (pI) of proteins (Solá and Griebenow, 2010), and different glycoforms can cause a pI shift in a given protein (Kuzmanov et al., 2009) thereby affecting its chromatographic behavior. Aiming to establish which would be the approximate ionic strength necessary to elute all the bound rhGLA from the AIEX column, a multiple step-wise elution profile was carried out. By adjusting the exact ionic strength, an increase of the purification factor was expected due to a better discrimination of contaminant proteins.

Single UV_{280nm} peaks were obtained for each step (fig. 4.5). Enzymatic activity quantification of fractions showed that the protein was distributed throughout different elution steps, thus confirming the initial presumption that rhGLA could have different net charges at a certain pH (maybe due to a heterogeneous glycosylation profile). Specifically, rhGLA activity was observed when elution was

performed applying concentrations of 50, 100, 150 and 300 mM NaCl. However, it was clear that the last 300 mM step concentration could be further reduced and optimized to improve even more the contaminant separation, as the total activity obtained in this step was very low (table 4.2).

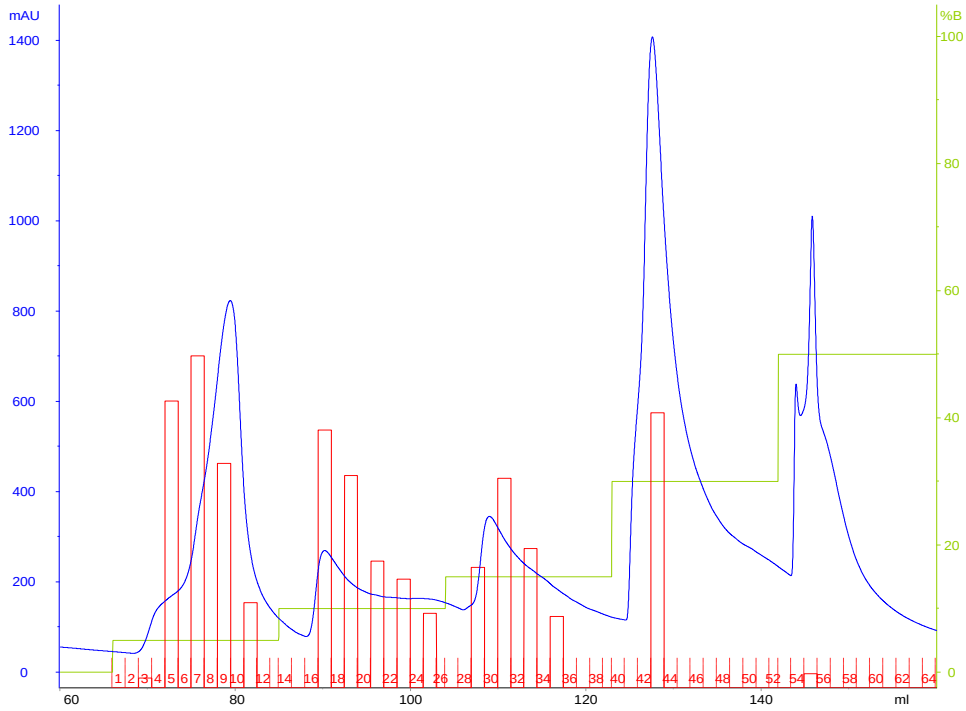


Fig. 4. 5. Chromatogram corresponding to the elution of the AIEX capture purification step. A step-wise elution profile (50, 100, 150, 300 and 500 mM NaCl) was performed. Blue line: UV_{280nm} signal. Green line: % of elution buffer (being 100% equal to 1 M NaCl). Red bars: GLA activity (arbitrary units).

Table 4. 2. Total activity and percentage recovered (with respect to the total activity recovered in the elution) in every elution step of the AIEX capture purification eluted by a step-wise strategy (1st approach).

Peak (mM NaCl)	Total AU (nmols MU/h)	% of total eluted activity

50	10313	37
100	8572	31
150	5835	21
300	3021	11

These results showed that a maximum concentration of 300 mM NaCl was enough to elute all the rhGLA bound to the column. However, and as long as the enzyme presented different net charges, a single-step elution strategy using an adjusted NaCl molarity able to de-attach every single "sub-population" of charged GLA would lead an increase of both the purification factor and the enzyme concentration. For this reason, a third AIEX test was performed to screen the accurate ionic strength necessary to elute all the GLA loaded and bound to the column. From the results obtained in the previous experiment, it was presumed that a concentration of 200 mM NaCl could be enough to elute all the protein of interest. This time, almost all the enzymatic activity was found in the fractions collected in the first 2 steps (100 and 200 mM NaCl), and only negligible amounts of GLA were eluted in the 3rd step, proving that 200 mM NaCl was sufficient to elute rhGLA (fig 4.6, table 4.3).

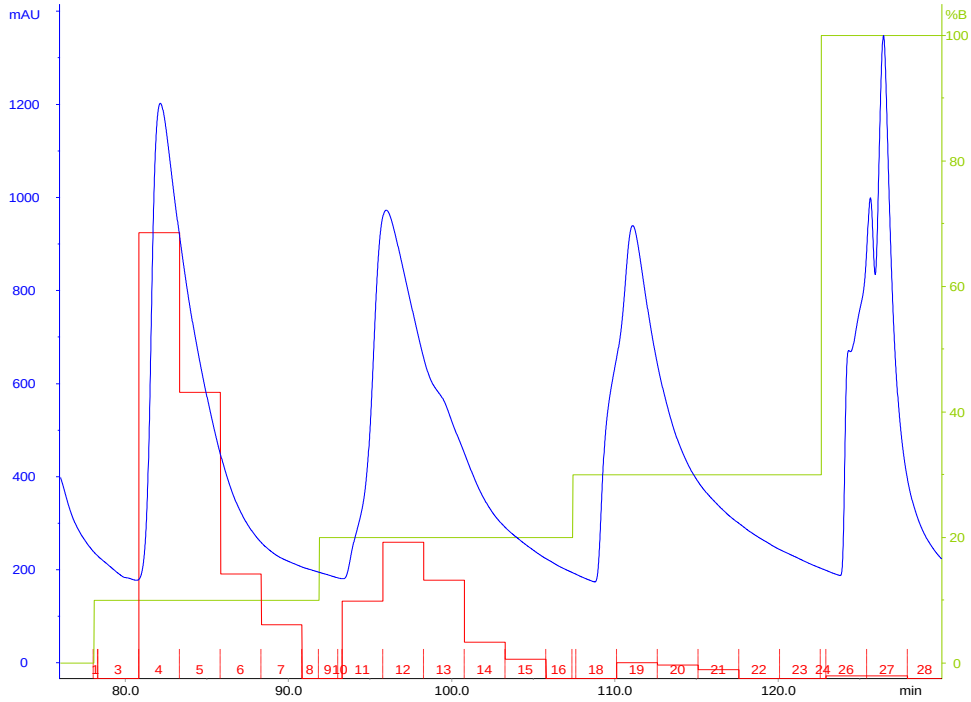


Fig. 4. 6. Chromatogram corresponding to the elution of the AIEX capture purification step. A step-wise elution profile (100, 200 and 300 mM NaCl) was performed. Blue line: UV_{280nm} signal. Green line: % of elution buffer (being 100% equal to 1 M NaCl). Red bars: GLA activity (arbitrary units).

Table 4. 3. Total activity and percentage recovered (with respect to the total activity recovered in the elution) in every elution step of the AIEX capture purification eluted by a step-wise strategy (2^{nd} approach).

Peak (mM NaCl)	**Total AU (nmols MU/h)**	**% of total eluted activity**
100	79307	68
200	32819	28
300	4479	4

Even though a remarkable amount of contaminants were still co-eluted with the protein of interest, the separation was notably improved with respect to the linear

gradient elution performed in the first experiment. However, the main aim of this this capture chromatographic step, apart from concentrating the target protein, was to eliminate interfering substances and major contaminants that could somehow affect subsequent steps.

Once the chromatographic media and the AIEX capture step conditions were established, a breakthrough capacity test was performed in order to scale up the process. The chromatogram obtained showed an increase of both UV_{280nm} signal and activity concentration starting from fraction 8, and from fraction 12 the activity measure was equivalent to the one corresponding to the loaded sample, meaning that all binding sites in the column had already been occupied (fig 4.7). Even though the complete saturation was achieved after loading 120 mL of sample, 80 mL of concentrated sample was considered the maximum loadable volume in order to avoid a decrease in the chromatography yield. This is known as the breakthrough point: the capacity limit from which protein loss is observed in the flow through due to column saturation.

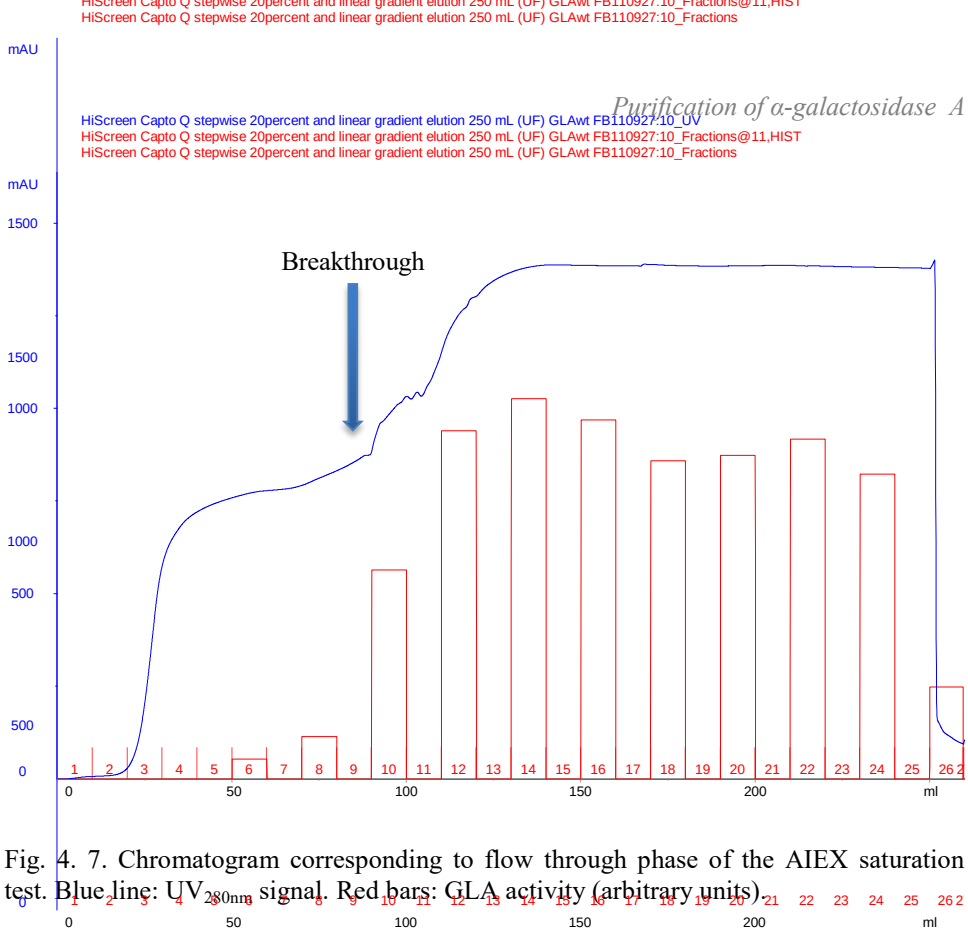


Fig. 4. 7. Chromatogram corresponding to flow through phase of the AIEX saturation test. Blue line: UV_{280nm} signal. Red bars: GLA activity (arbitrary units).

Therefore, this value was taken to calculate the amount of clarified and concentrated supernatant that could be loaded in 4.7 mL of chromatographic media:

$$8\ fractions \cdot \frac{10\ mL}{1\ fraction} = 80\ mL\ (clarified, ultrafiltered\ sample)$$

Considering that the sample was initially concentrated from 1L to 250 mL, the amount of supernatant that could have been loaded without saturating the column can be calculated:

$$80\ mL_{concentrated\ sample} \cdot \frac{1000\ mL_{supernatant}}{250\ mL_{concentrated\ sample}} = 320\ mL_{supernatant}$$

Assuming that the average of supernatant volume obtained from a 5 L bioreactor process is around 3 L (due to the volume corresponding to biomass), the amount of chromatographic media is calculated as follows:

$$3000\ mL_{supernatant} \cdot \frac{4.7\ mL_{chromatographic\ media}}{320\ mL_{supernatant}} = 44\ mL_{chromatographic\ media}$$

According to the calculations, approximately 44 mL of chromatographic media were necessary for the purification of a whole 5 L-scale production process in a single run. It has to be noted that the breakthrough value depends not only on the media and sample properties, but also on the dynamic profile of the process. Thus, the calculated capacity depends as well on the flow rate. For this reason, this parameter had to be properly up-scaled to keep the linear flow constant between scales. So, the linear flow rate used in the 4.7 mL scale had to be calculated from the volumetric flow value (1.2 mL/min) and the column dimensions (diameter of 0.77 cm):

$$F_{linear}\ (cm/h) = \frac{F_{volumetric}\ (mL/min) \cdot 60\ (min/h)}{A_{section}\ (cm^2)}$$

$$\frac{1.2\ mL/min \cdot 60\ min/h}{\pi \cdot \left(\frac{0.77}{2}\right)^2\ cm^2} = 154.6\ cm/h$$

Taking the calculated linear flow and considering the new dimensions of the XK26/20 column (diameter of 2.6 cm), the volumetric flow of the bigger scale could be estimated:

$$F_{volumetric}\ (mL/min) = \frac{F_{linear}\ (cm/h)}{60\ (min/h)} \cdot A_{section}\ (cm^2)$$

$$\frac{154.6\ cm/h}{60\ min/h} \cdot \pi \cdot \left(\frac{2.6}{2}\right)^2 cm^2 = \mathbf{13.7\ mL/min}$$

4.3.2.2. Hydrophobic chromatography

Setting up a hydrophobic chromatography (HIC) can be tricky due to the inaccuracy of hydrophobicity profile predictions of a globular protein in solution, but if the right conditions are determined it is common to observe high purification factors since few proteins are capable to bind to the column. Therefore, the protein of interest can be usually recovered at high concentrations and purity in a single step. On the other hand, the yield obtained in HIC can be low compared to IEX, probably due to instability issues caused by the high salt concentrations necessary for the binding step (the sample is normally adjusted to 0.5 - 1 M of salt, usually NaCl or $(NH_4)_2SO_4$), which can be responsible for certain inactivation and/or aggregation of some proteins. At the same time, the high salt concentration required represents one of the main advantages of using HIC, since is not necessary in principle to dialyze or ultrafilter the sample but only increase its salt content.

The linear decreasing gradient of salt concentration is the most extended strategy to elute the bound proteins. Typically, the chromatographic parameters (basically the type of ligand and the salt concentration of the binding buffer) are adjusted to obtain the elution of the desired protein at the middle of the gradient. There are different ligands available for HIC, whose hydrophobicity profiles differ one from the other. Therefore, the ligand type turns to be the first parameter to be determined in HIC, and the choice will depend basically on the hydrophobicity and stability of the protein of interest at high salt concentrations. With the aim to evaluate HIC as the capture step, Capto Butyl was selected as the chromatographic media to perform the corresponding tests. As mentioned previously, Capto-based media is specially recommended for capture and intermediate purification. In this case, butyl ($-O-(CH_2)_3-CH_3$) is the hydrophobic ligand coupled to the resin, which is considered a group with an intermediate relative hydrophobicity according to the manufacturer.

The first experiment aimed to determine if HIC could be used as the capture step to take advantage of its high selectivity (instead of IEX). The recovery yield obtained was an impermissibly poor 18%. Unfortunately, the specific reasons for this low yield could not be determined. It was hypothesized that undefined substances present in the supernatant could be interfering with the binding sites of the column and/or causing protein aggregation or inactivation under the high salt concentrations required for this technique. To confirm this suspicion, a pool of samples recovered from the purification by HIC was re-purified with the same column. The rhGLA eluted in the first run was supposed to be free of most of the species present in the supernatant, and a remarkable increase in the purification yield was expected. Indeed, this time a 76% of the loaded activity was recovered (table 4.4), confirming that HIC could be efficiently used for pre-purified samples. In front of these results, it was established that this technique had to be used as an intermediate purification step, where the main objective is to find a balance between resin capacity and resolution (fig. 4.8).

Table 4. 4. Total activity and percentage recovered (with respect to the activity loaded into the column) in the HIC purification. Comparison between the purification of a supernatant and a pre-purified sample.

Fraction	**Sample from supernatant**		**Pre-purified sample**	
	Total AU (nmols MU/h)	**% recovered**	**Total AU (nmols MU/h)**	**% recovered**
Loaded sample	146529		114950	
Elution	26294	18 %	87190	76 %

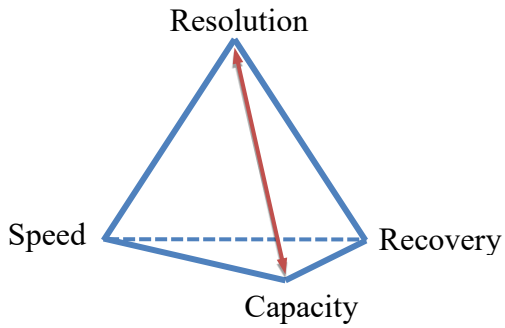


Fig. 4. 8. Representation of the balance between speed, capacity, resolution and recovery in chromatography purification steps. The objective of intermediate purification steps is a good balance between media capacity and resolution.

Following the proposal to use HIC as intermediate purification step, a sample purified by AIEX using the conditions previously optimized was loaded into the HIC column. Again, the elution was performed with a linear gradient. The results obtained confirmed the high specificity of the chromatographic media and conditions selected: the chromatogram showed a low UV_{280nm} signal in the elution step compared to the high signal measured during the sample loading (fig. 4.9), meaning that most of the proteins of the loaded sample did not bind to the column. The high selectivity of this step was later confirmed by SDS-PAGE analysis of the eluted fractions (fig. 4.10), where only few proteins were recovered including GLA. Unfortunately though, a substantial overlap between the proteins eluted was seen (including a main contaminant of ~40 kDa). This effect would be caused by the high bead size of the resin (90 μm) and its associated low resolution, and even though the use of a column with lower bead size (and therefore higher resolution) could solve that, no such product was available for the Capto Butyl format.

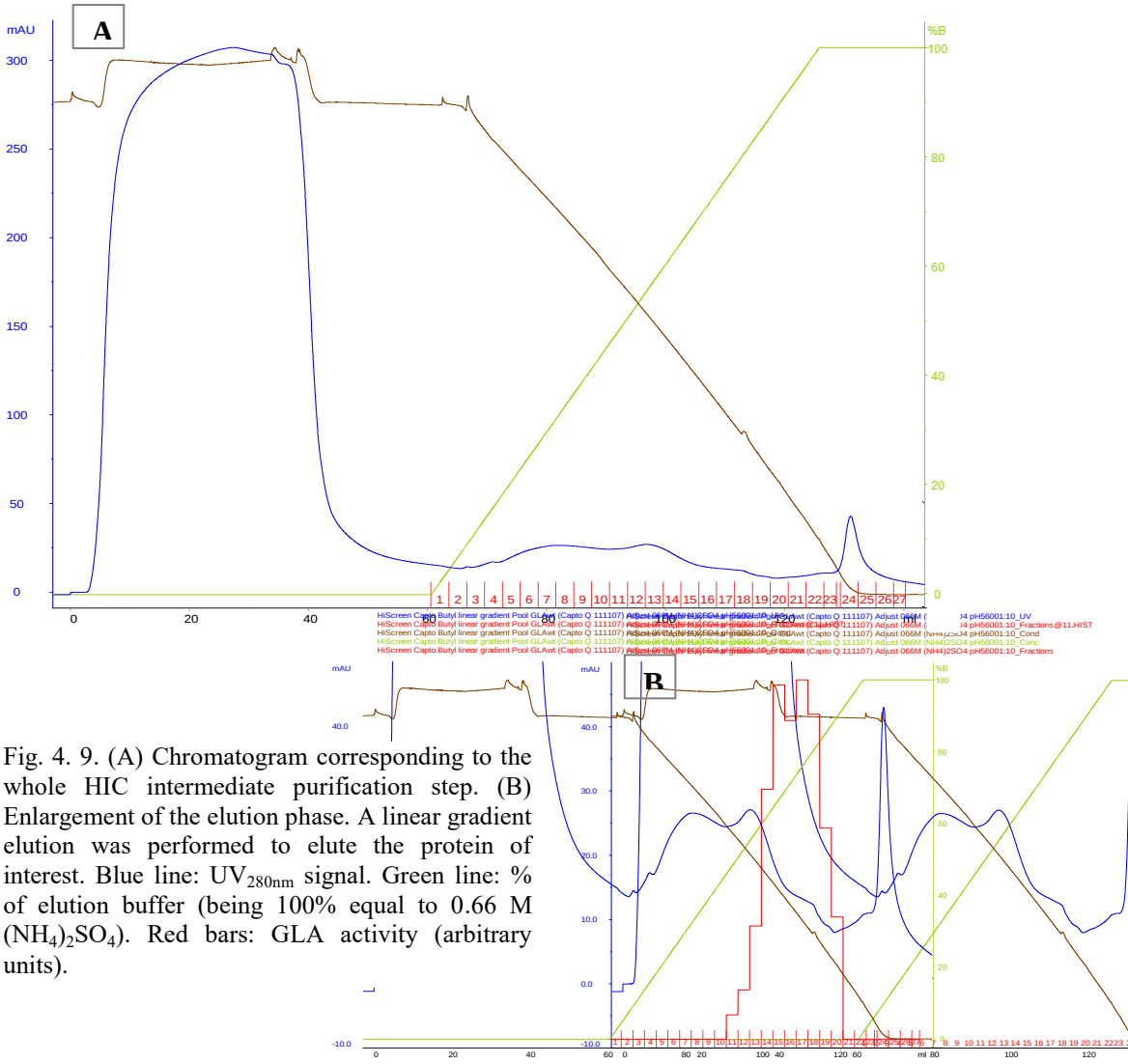


Fig. 4. 9. (A) Chromatogram corresponding to the whole HIC intermediate purification step. (B) Enlargement of the elution phase. A linear gradient elution was performed to elute the protein of interest. Blue line: UV_{280nm} signal. Green line: % of elution buffer (being 100% equal to 0.66 M $(NH_4)_2SO_4$). Red bars: GLA activity (arbitrary units).

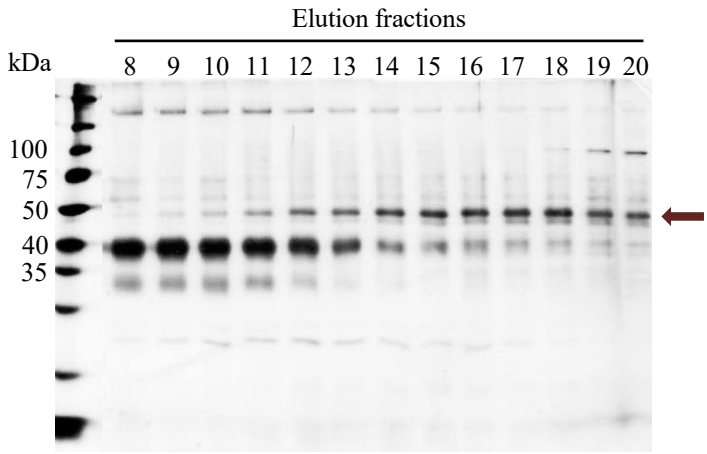


Fig. 4. 10. SDS-PAGE (silver-stained) analysis of the elution fractions obtained from the HIC purification. The red arrow shows the ~50 kDa band corresponding to the GLA, which is coincident with the activity measures of the loaded samples.

Regarding the yield of this chromatographic step, 68% of the total loaded activity was estimated to be recovered in the elution fractions (table 4.5). This value was considered acceptable to integrate HIC as the intermediate chromatographic step in the purification process, especially considering that the purification factor was high (according to the SDS-PAGE).

Table 4. 5. Total activity and percentage recovered (with respect to the loaded sample) in the HIC intermediate purification step eluted by linear gradient.

Fraction	**Total AU (nmols MU/h)**	**% recovered**
Load	71822	
Flow through	296	0 %
Elution	48661	68 %

3.1.1. Size-exclusion chromatography

The last step of the chromatographic process was a SEC (Size Exclusion Chromatography, also known as Gel Filtration). This technique is based in the separation of protein populations according to their size, without any chemical interaction between the chromatographic media and the sample. These columns are especially suited for polishing steps (fig. 4.11), due to their advantages and limitations. Gel filtration can be used to separate samples with a small number of different proteins provided that they clearly differ in their molecular weights. It is also very efficient when separating aggregates, multimers and/or proteolysed fragments from a single protein, which is critical to maximize the purity of therapeutic proteins. Another great advantage is the possibility to perform a buffer exchange simultaneously during the chromatography, thanks to the fact that salts are eluted earlier than any small protein. On the other hand, the technique has an important limitation: the correlation [bed height] / [column diameter] is very high compared to other types of chromatography. This geometrical relationship, necessary to ensure enough distance for protein separation, is responsible for the low [sample volume] / [bed volume] factor (e.g. 5 mL of sample in a 125 mL column). At the same time, it causes the flow rate to be very low (0.5 to 1 mL/min in a 125 mL column) to avoid overpressure issues in such a high bed. On the whole, samples have to be highly pure and concentrated before being loaded in a SEC column (which is typically achieved only in the last chromatography steps), and the processes are very resolutive but slow (Berek, 2010).

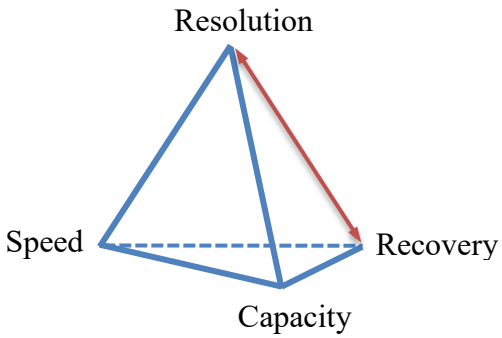


Fig. 4. 11. Representation of the balance between speed, capacity, resolution and recovery in chromatography purification steps. The objective of polishing steps is a good balance between media resolution and product recovery (yield).

For the polishing of rhGLA samples coming from the HIC step, Superdex 200 Prep Grade was the chosen resin. Two peaks were clearly separated, the first one corresponding to the total amount of rhGLA activity recovered (fig. 4.12). This peak showed an asymmetric profile, with a belly on its left side that could be due to the presence of multimeric aggregates in the sample (the activity profile measured was absolutely coincident with such peak shape). The pool was analyzed by Western-blot (fig 4.13) to confirm that the band of ~50 kDa corresponded to the purified rhGLA.

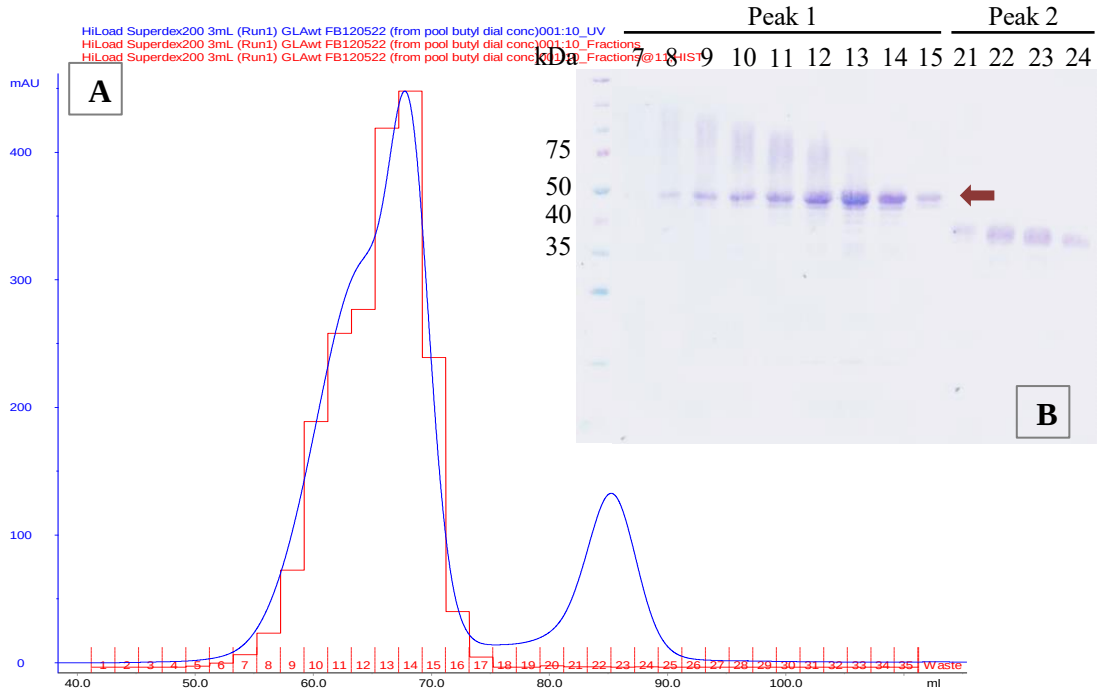


Fig. 4. 12. (A) Chromatogram corresponding to the SEC polishing purification step. An isocratic elution profile was performed to separate the protein of interest from the contaminants still present in the sample. Blue line: UV280nm signal. Red bars: GLA activity (arbitrary units). (B) SDS-PAGE (Coomassie-stained) analysis of the elution fractions obtained. The red arrow shows the ~50 kDa band corresponding to the GLA, which is coincident with the activity measures of the loaded samples.

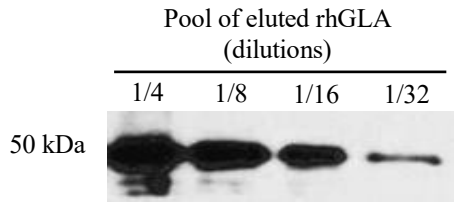


Fig. 4. 13. Western-blot analysis of the rhGLA purified by size exclusion chromatography. Several dilutions of the sample were tested to determine the detection limit of the rabbit anti-GLA antibody used.

The purity was calculated by densitometry to be over 98%. The second peak corresponded to the ~ 40 kDa main contaminant identified as the main protein co-purified in the HIC intermediate step. In view of these results, keeping all the active fractions from the HIC elution regardless of its purity would have been a better choice, as the SEC was capable of separating the main contaminant. This measure could be taken in future purifications in order to improve the final yield of the process.

4.3.2.3. Overall process evaluation and discussion

The optimized conditions and parameters defined in the previous sections were used to perform an entire purification procedure from a whole bioreactor batch, in order to evaluate and detect process weaknesses and potential improvements. In the case of the AIEX step, a 50 mL column was used in accordance with the up-scaled process, allowing the purification of a whole bioreactor supernatant in a single run. The overall purification parameters are detailed in table 4.6.

Table 4. 6. Overall purification summary. SN: supernatant. MF: microfiltration. UF: ultrafiltration. AIEX: anion exchange chromatography. HIC: hydrophobic chromatography. SEC: size exclusion chromatography.

STEP	Volume (mL)	Activity (UA/mL)	Total activity ($UA \cdot 10^6$)	Protein concentration (mg/mL)	Total protein (mg)	Specific activity (UA/mg P)	Yield (%)	Purif. Factor
SN	3152	1031	3.25	0.244	768	4232	100	
MF	4072	658	2.68	0.183	743	3607	82	0.9
UF	580	5325	3.09	0.545	316	9778	95	2.3
Capto Q (AIEX)	189	15933	3.01	0.842	167	18023	93	4.3
Capto Butyl (HIC)	8.5	184379	1.57	1.373	11.7	134188	48	31.7
Superdex 200 PG (SEC)	10.5	136366	1.43	1.310	9.5	150526	44	35.5

As can be deduced from table 4.6 and the associated figure (fig. 4.14), the critical step that allowed a high increase in the purification factor was the hydrophobic chromatography. According to the final purified sample, the specific activity of rhGLA produced in *P. pastoris* was $1.5 \cdot 10^5$ AU/mg of protein.

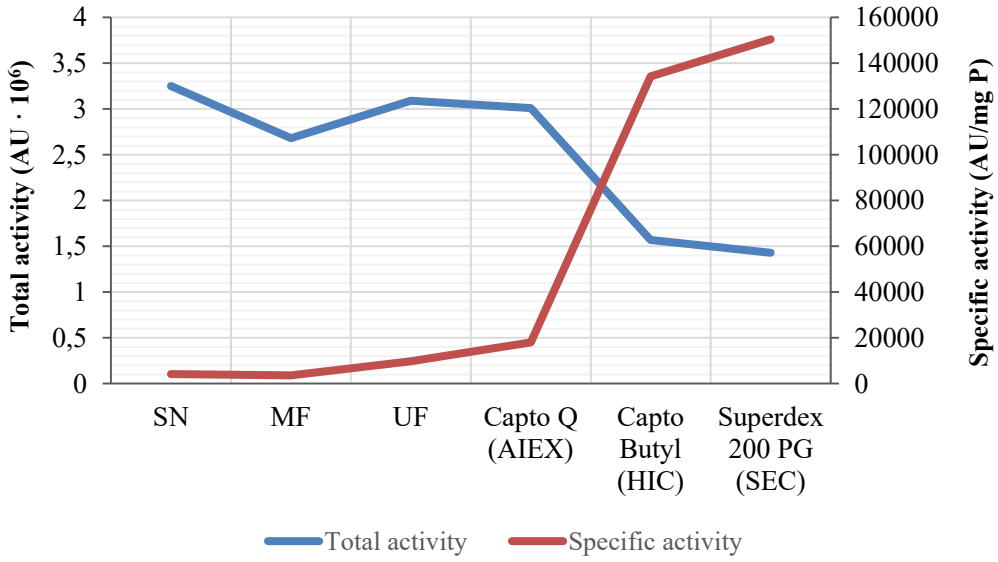


Fig. 4. 14. Evolution of the total activity (as a reference of accumulated yield) and specific activity (as a reference of purity) throughout the whole rhGLA 3-step purification process. SN: supernatant. MF: microfiltration. UF: ultrafiltration. AIEX: anion exchange chromatography. HIC: hydrophobic chromatography. SEC: size exclusion chromatography.

Two major improvable points could be deduced from the previous data. In first place, a low purification factor (1.87x) was calculated for the AIEX step. Such reduced value, previously anticipated by the chromatogram UV_{280nm} signal, indicates that rhGLA was not being as efficiently separated from contaminants as desirable. The main strategy to improve this parameter in this capture step would be to minimize the amount of contaminant proteins bound to the column during the sample loading by modifying the buffer composition. Another parameter with clear room for amelioration was the yield of the HIC step (52%). Because of the overlap

between the elution of GLA and some contaminating proteins, some tail fractions were discarded to avoid an excessive contaminant transference to the SEC step. In this case, an increase in the column resolution would allow to separate efficiently the eluted proteins and take all the eluted GLA instead of wasting contaminated elution tails.

In view of these results, the following improvements would be proposed for the described purification process:

- ***Ultrafiltration step: increase nominal pore size to ~50 kDa.*** Although the Mw of the GLA monomer is around 50 kDa, it acquires a dimeric conformation in its native state. Filter suppliers usually recommend to use ultrafiltration cassettes with a pore size of maximum a 50% of the Mw of the protein of interest, in order to avoid any unexpected leakage during the diafiltration. In the process described, a 10-12 kDa filter membrane was used due to availability issues. However, the increased pore size would help to get rid of small proteins before the chromatographic steps, helping to increase both their purification factors and increasing the sample capacity of the columns.

- ***AIEX step: decrease pH of mobile phase from 6 to ~5.5.*** Assuming that the real pI of the recombinant GLA produced in *P. pastoris* is similar to the pI of 5.14 inferred from its amino acid sequence, a fine tuning in the binding pH between 5.5 and 6 could avoid the binding of some proteins whose pI is inside this range. In fact, a considerable amount of yeast proteins have been classified with a pI between 5 and 6 (Kiraga et al., 2007). An increased selective binding could result, again, in higher sample capacity of this column and a greater purification factor.

- ***HIC step: column of lower bead size (from ~ 90 µm to ~ 40 µm).*** After having developed the purification process, a new Capto Butyl chromatographic media with reduced bead average size (40 µm) was commercially launched (Capto Butyl ImpRes). This bed size, more common in intermediate purification steps, would allow a greater peak resolution and protein separation.

- ***Overall process: use of Tween 20 as an antiaggregant additive.*** Aggregates can easily lead to product loss during the chromatographic steps, either because the protein gets stuck into the column or due to a deficient interaction with the solid phase. These aggregates, as shown below in section 4.3.5.3, were efficiently dissolved without activity loss by using Tween 20, an additive compatible both with chromatographic media and the use of the purified protein as a therapeutic product.

4.3.3. Purification of rhGLA variants (rhGLA, rhGLAΔKDLL, $rhGLA_{C\text{-}HIS}$, $rhGLA_{C\text{-}RGD}$) by galactose-affinity chromatography

The 3-step chromatographic process already described proved to be useful to obtain GLA samples with high purity and concentration. However, the process was complex and time-consuming, meaning that its usefulness is specially indicated for the purification of high amounts of enzyme. Such quantities were not required for the purification and characterization of the enzyme variants expressed and discussed previously in chapter 2, nor for eventual characterization of the glycosylation profile of modified yeast strains. As an alternative, α-galactose affinity chromatography was seen as a promising alternative for the fast and easy purification of these GLA versions. D-galactose agarose media was initially designed for the purification of lectins, galactosidases and other galactose-binding molecules but has been already used for the purification of α-galactosidase A

(Yasuda et al., 2004). Specifically, it consists of α-D-galactose residues immobilized onto crosslinked 6% beaded agarose through a thiol bond (fig 4.15).

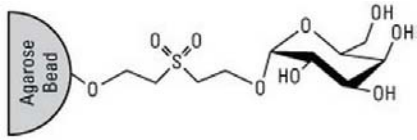


Fig. 4. 15. Chemical structure of D-Galactose Agarose. (Source: Pierce Thermo Scientific).

The technique relies on the interaction between the galactose substrate (covalently-bound to the agarose) and the active site of the enzyme. To achieve an efficient elution of the target protein, free galactose is used as a competitor substance for the active sites of the enzyme. Alternatively, a sudden increase/decrease of the pH outside its catalytic activity range can also be used to avoid the interaction with the chromatographic media.

4.3.3.1. Set-up of galactose-affinity chromatography with rhGLA

In order to define the chromatography parameters for the galactose affinity chromatography, rhGLA was chosen as the model variant. Taking into account that this technique is based on the activity profile of the enzyme, it was assumed that the results obtained should be valid for any other active enzymatic variants.

In a first attempt, default conditions provided by the supplier were tested. Unfortunately, with such conditions only a 20% of the total loaded activity was recovered in the elution fraction. The fact that only 2% was lost in the flow through fraction meant that most of the protein remained bound inside the column and that, most probably, the elution conditions were not enough to properly de-attach the enzyme from it.

In a second try, different elution conditions were tested sequentially in the same chromatography. Unfortunately, this experiment did not led to any improvement in the recovery yield (again 22%, between all the elution conditions tested), and showed that even those conditions specially used to avoid enzymatic substrate recognition (3rd and 4th elution steps) were unable to elute the remaining activity (table 4.7).

Table 4. 7. Total activity and percentage recovered (with respect to the total activity loaded into the column) in every elution step of the D-galactose affinity chromatography.

Step	**Elution conditions**			**Recovered activity in the elution fraction (% of loaded activity)**
	pH	**[Galactose] (M)**	**Others**	
Binding	4.6	-	-	2 %
Elution 1	5.8	0.1	0.1 M NaCl	17 %
Elution 2	5.8	0.5	0.1 M NaCl	2 %
Elution 3	8	0.5	0.1 M NaCl	1 %
Elution 4	2.3	-	0.1 M Glycine	0 %
Total				22%

To confirm that the decrease in the activity percentage calculated for each elution condition coincided with a decrease in the protein amount (and not to an enzymatic inactivation), elution fractions were analyzed by SDS-PAGE (fig. 4.16). The decrease in the intensity if the bands matched indeed with the decrease in the activity measures. Interestingly, cleaning-in-place (CiP) conditions did not elute any additional protein, not even inactive enzyme (fig. 4.16, fraction 15). This

would mean that there was not any rhGLA interacting with the resin, and that the missing enzyme loaded could be trapped inside the column.

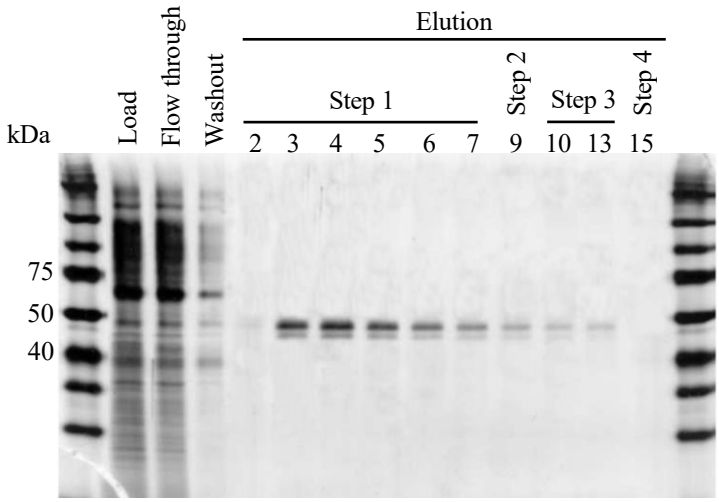


Fig. 4. 16. SDS-PAGE (silver-stained) analysis of the elution fractions obtained from the D-galactose affinity purification.

Even if the loaded product is properly active, there are a number of different reasons that could explain the poor rhGLA recovery obtained. First of all, and as it is being discussed along this chapter, product aggregation was a major concern that could explain why the enzyme recovered in the elution fraction was lower than expected: enzyme clusters could easily get stuck within the column, explaining the activity loss. Keeping in mind that yield was not the main objective of this process, but the high purification factor in order to obtain small amounts of pure product in a single step, no additional improvements were attempted. Even if a high recovery would had been obtained from the beginning, the high cost of the chromatography media (> €3000/100 mL) makes it unsuitable for its scale-up and use as the purification method of high amounts of enzyme. Nevertheless, if additional variant purification and characterization has to be performed in the near future, it would be worthwhile to check the effect of anti-aggregating substances to confirm or discard this hypothesis. Other reasons could be directly related to the chromatographic

media structure (which was not originally designed for the purification of GLA), basically the ligand density and/or the linker length. These factors can be related to the media efficiency and can be optimized in other to reach the maximum binding and capacity, allowing the correct product interaction without steric hindrance.

4.3.3.2. Purification of rhGLA variants by galactose-affinity chromatography

Using the conditions set up previously for the purification of rhGLA (GLA with the native peptide sequence), supernatants from cultures expressing rhGLA, rhGLAΔKDLL, $rhGLA_{C\text{-}HIS}$ and $rhGLA_{C\text{-}RGD}$ were submitted to the galactose affinity chromatography. Eluted fractions showed a high level of purity, and the presence of rhGLA was confirmed by western-blot (fig. 4.17).

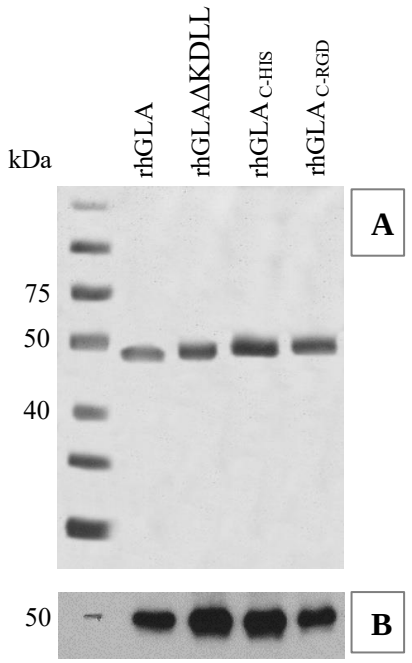


Fig. 4. 17. SDS-PAGE (silver-stained) (A) and Western-blot (B) analyses of the rhGLA variants purified by D-galactose affinity chromatography.

After determination of GLA activity and protein concentration in each sample, the specific activity of the different variant could then be calculated (fig. 4.18).

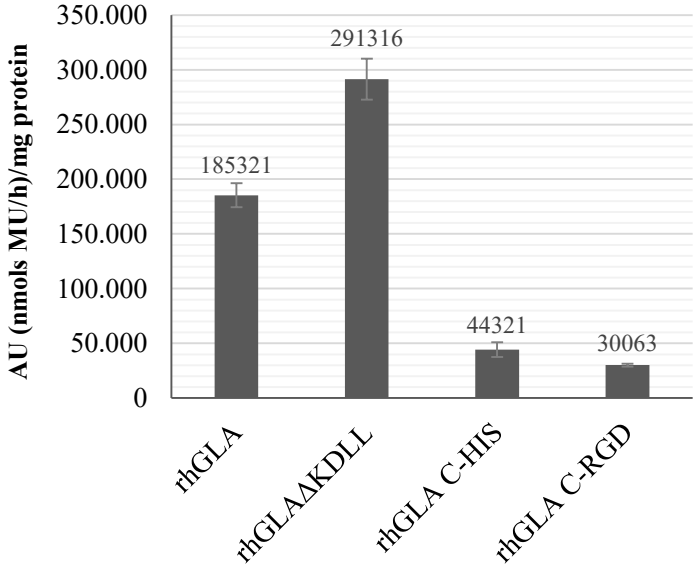


Fig. 4. 18. Specific activity measured for the different rhGLA variants purified by D-galactose affinity chromatography. rhGLA: native peptide sequence. rhGLAΔKDLL: rhGLA lacking the C-terminal KDLL residues. $rhGLA_{C\text{-}HIS}$: rhGLA containing C-terminal 6xHistidine tag. $rhGLA_{C\text{-}RGD}$: rhGLA containing C-terminal RGD internalization motif.

These results confirmed the hypothesized increase on the specific activity conferred by the deletion of the 4 C-terminal amino acids (KDLL). Specifically, a 60% improvement was measured in comparison with the rhGLA with the native sequence. It has been suggested that modifications in the C-terminal sequence of the protein would not imply an increase in its enzymatic performance, because the predicted structure shows that the terminal residues have no electron density and are probably disordered (Garman and Garboczi, 2004). However, the data obtained here confirms the improvement at the protein level for the first time. This interesting finding would mean that, if used as a therapeutic product,

rhGLAΔKDLL could reach an equivalent catalytic performance to rhGLA with less amount of protein administered, reducing potential immunogenic responses and minimizing the costs per dose.

The specific activity obtained for the rhGLA is similar to the one measured in the enzyme purified by the 3-step chromatography process ($1.86 \cdot 10^5$ vs. $1.51 \cdot 10^5$ AU/mg of protein, respectively). On the other hand, the measures of the $rhGLA_{C\text{-}HIS}$ and $rhGLA_{C\text{-}RGD}$ variants revealed notably lower values ($0.44 \cdot 10^5$ AU/mg and $0.30 \cdot 10^5$ AU/mg, respectively). Interestingly, this result was in accordance with the lower productivities measured for *P. pastoris* clones producing these variants (see Chapter 2) when compared to clones expressing rhGLA: cells could be producing similar levels of enzyme in terms of total protein, but seemingly the lower activity measures were caused by the lower catalytic performance. In conclusion, it seems clear that the addition of any C-terminal sequence to rhGLA is detrimental to its activity profile, maybe due to some interaction with functional sites and/or the correct conformation of the homodimer.

Despite having a very low product recovery profile, this technique proved to be an efficient way to purify small amounts of rhGLA, regardless the modifications made to the different variants generated. However, the chromatographic media used would be useful to isolate any catalytically active rhGLA, so that its binding site is able to recognize and bind specifically to the α-galactose bound substrate.

4.3.4. Purification of 6xHis-tagged rhGLA ($rhGLA_{C\text{-}HIS}$) by nickel affinity chromatography

As seen in chapter 2, an rhGLA variant with a C-terminal 6x Histidine tag was successfully expressed in *P. pastoris*. This version of rhGLA ($rhGLA_{C\text{-}HIS}$) has a great potential as a tool for glycosylation analyses of future humanized yeast strains. The main advantage of the tagged enzyme is the possibility to purify it in a

single step (2 at most, if a polishing step is required) without a complex sample conditioning, thereby allowing the fast processing of multiple strain variants and their corresponding clone collections. Even though galactose-affinity purification was the first choice for this purpose, its low purification yield led to the development of a more efficient purification strategy.

Ni Sepharose™ 6 Fast flow was the chosen chromatographic media. As part of the immobilized metal ion affinity chromatography (IMAC) family, this media takes advantage of the interaction between chelated metal ions and side-chains of some amino acids (mainly histidine). Ni^{2+}, usually the preferred metal ion for the purification of histidine-tagged proteins, is tightly bound to a chelating group coupled to the resin and is able to interact specifically with exposed groups of histidine. Imidazole is the compound used both in the binding stage (at low concentrations, to avoid unspecific binding of contaminant proteins) and in the elution (at high concentrations, used to achieve the efficient de-attachment of the target protein) (Bornhorst and Falke, 2000).

According to the His-tag affinity media supplier, samples to be purified should be adjusted to a pH between 7 and 8 to ensure a proper binding. However, rhGLA is unstable in solution at pH > 7 (Ishii et al., 2000), which represents a limitation when working with this affinity media. Such limitation can be overcome, at least partially, by minimizing the time between the sample conditioning and neutralizing collected elution samples in tubes preloaded with some "stability buffer" (i.e. pH below 7). The purification of an rhGLA expressed in HEK 293F cells by IMAC was shown to be feasible, providing samples with degree of purity (Corchero et al., 2011).

In first place, the suitability of different pH values for the purification of $rhGLA_{C\text{-}HIS}$ by IMAC was assessed. In this experiment, the whole process of sample conditioning (which involves not only the adjustment of pH and salts concertation, but also a filtration step) was evaluated prior to loading the samples into the chromatography column (fig. 4.19).

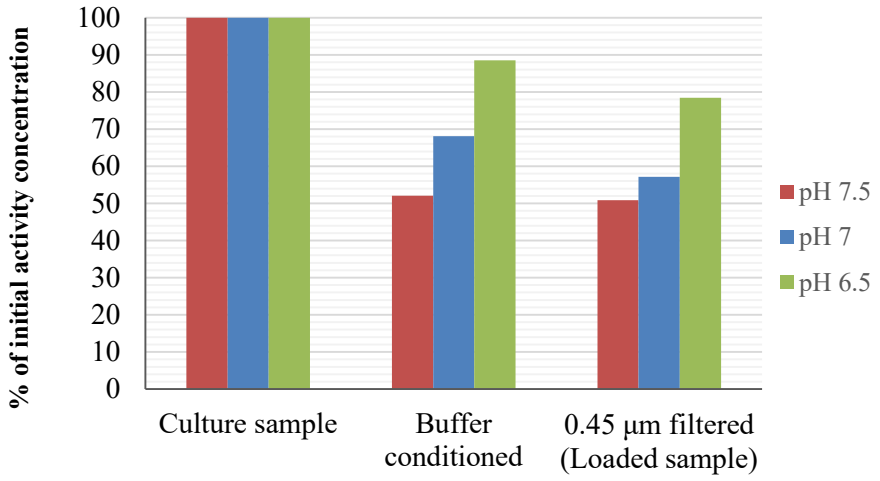


Fig. 4. 19. Recovery yield comparison between rhGLA$_{C\text{-}HIS}$ samples submitted to buffer conditioning (addition of concentrated binding buffer at pH 6.5, 7 or 7.5) and filtration (0.45 µm).

These results confirmed that rhGLA$_{C\text{-}HIS}$, as described in the literature for rhGLA (Ioannou et al., 2001), showed a marked decrease in the activity when adjusted to pH 7.5 (52% of initial activity concentration). This activity loss was reduced at pH 7 and 6.5 (68% and 89% of the initial concentration, respectively). Regarding the filtration step, an additional 11% was lost both for samples at pH 7 and 6.5, which could be due to some protein aggregation. The filtration loss in the sample at pH 7.5 was negligible. Even though the adjustment of the sample at pH 6.5 clearly showed the best overall conditioning yield (78%), the chromatography purification was carried out for the three samples, as binding at pH 7 and especially at pH 6.5 was uncertain. Figure 4.20 shows the results obtained for the affinity chromatography performance of rhGLA$_{C\text{-}HIS}$ samples at the different pH assayed. Interestingly, binding at pH 6.5 not only was efficiently achieved, but showed the best performance between the different conditions tested. Specifically, a 37% was recovered in the elution (purified) fraction, being pH 7 the next with a 23% of activity recovery in the same fraction. This would be likely because of the higher

stability environment provided by pH 6.5, which will keep more efficiently the enzymatic activity throughout the chromatography.

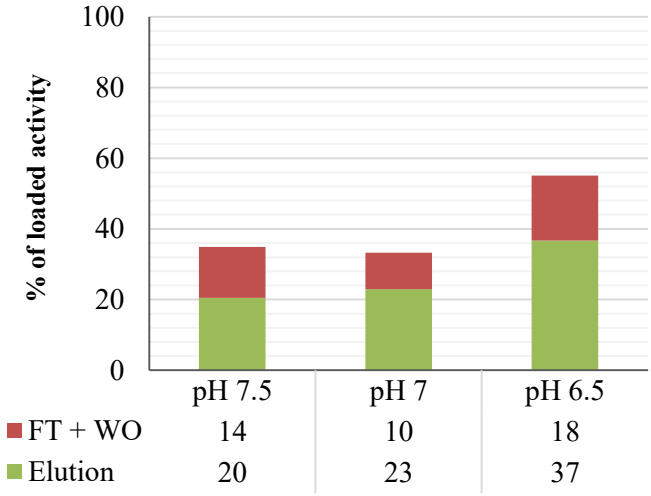


Fig. 4. 20. Ni2+ affinity chromatography performance of rhGLA$_{C\text{-}HIS}$ samples adjusted at different pH. Activity was quantified for the flow through (FT), wash-out (WO) and eluted fractions.

Unfortunately, in all cases the total activity recovery between the flow through, washout and elution fractions was below 60% of the total loaded activity. This could be due either to a progressive inactivation of the enzyme, to deficient elution conditions or to protein aggregation (the last leading to protein clogging within the column). Regarding the possibility of using an inefficient elution buffer, the imidazole concentration already used as a competing substance in the elution buffer (500 mM) was considered to be high enough and no change was done. Instead, variations in the buffer formulations were made to increase the protein stability and minimize the aggregation effect. Two additives were tested aiming to increase the chromatography recovery yield: 10% glycerol or 0.01% Tween 20. However, only pH 6.5 was selected for further process improvement (fig. 4.21).

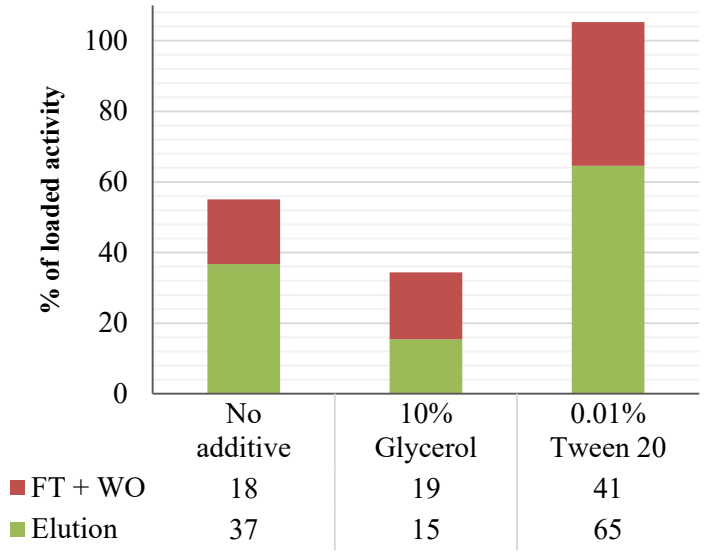


Fig. 4. 21. Ni2+ affinity chromatography performance of rhGLA$_{C\text{-}HIS}$ adjusted at pH 6.5 in absence or presence of additives (10% glycerol or 0.01% Tween). Activity was quantified for the flow through (FT), wash-out (WO) and eluted fractions.

Tween 20 increased the total recovery yield from 55% (with no additives) to a ~100%, meaning that no product was left inside the column. This result is consistent with the aggregation hypothesis, as Tween 20 (polysorbate 20) is described as being an efficient antiaggregant substance (Carpenter et al., 2002). Adding such substance, the enzyme would disaggregate and would be able to cross efficiently the whole chromatographic packed bed. On the other hand, addition of glycerol did not contribute to any improvement in the chromatographic process. In fact, and for unknown reasons, it even worsened the total recovery yield experienced with no additive (34% and 55%, respectively). Although the purification yield achieved with Tween 20 was not as high as desired (65%), it was considered good enough to obtain sufficient amounts of protein for glycan analysis. The fact that the rest of the activity was found in the flow through and washout fractions could mean that binding at pH 6.5 was, as expected from the resin characteristics, significantly low. An improvement could potentially be achieved by optimizing the pH between 6.5 and 7 to maximize the binding efficacy while

preserving the enzyme activity. Moreover, it would be worthwhile to lower the flow rate and/or to increase the column height in order to increase the residence time of the sample, allowing it to bind more efficiently to the solid phase. If so, this could lead to an increase in the activity recovered in the elution fraction and, consequently, a decrease in the amount of protein lost in the flow through.

The sample obtained from the chromatography process containing Tween 20 had a high purity degree, as seen by SDS-PAGE. The presence of $rhGLA_{C\text{-}HIS}$ was confirmed by western-blot (fig. 4.22). The total amount of protein obtained (1.2 mg, from a 100 mL culture in shake flask), was sufficient for protein characterization including the glycoprofile analysis.

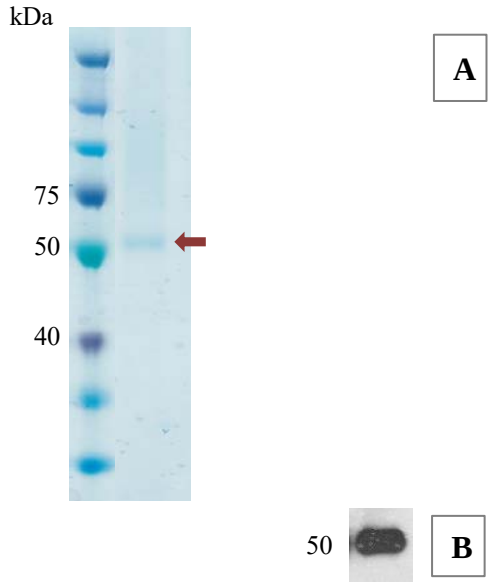


Fig. 4. 22. SDS-PAGE (PageBlue-stained) (A) and Western-blot (B) analyses of the $rhGLA_{C\text{-}HIS}$ variant purified by immobilized metal ion affinity chromatography (IMAC).

4.3.5. Protein characterization

The purification procedures studied in the previous sections allowed to obtain highly purified GLA samples that allowed its characterization by several methods.

4.3.5.1. Protein identification by peptide mass fingerprinting

A pure sample of rhGLA was submitted to an external proteomic analysis service to perform a peptide mass fingerprinting identification. The identified peptides are detailed in figure 4.23.

LDNGLARTPTMGWLHWERFMCNLDCQEEPDSCISEKLFMEMAELMVSEGWKDAGYEY
LCIDDCWMAPQRDSEGRLQADPQRFPHGIR↓QLANYVHSKGLKLGIYADVGNKTCAG
FPGSFGYYDIDAQTFADWGVDLLKFDGCYCDSLENLADGYKHMSLALNRTGRSIVYS
CEWPLYMWPFQKPNYTEIRQYCNHWR↓NFADIDDSWKSIKSILDWTSFNQERIVDVA
GPGGWNDPDMLVIGNFGLSWNQQVTQMALWAIMAAPLFMSNDLRHISPQAKALLQDK
DVIAINQDPLGKQGYQLRQGDNFEVWERPLSGLAWAVAMINRQEIGGPRSYTIAVAS
LGKGVACNPACFITQLLPVKR↓KLGFYEWTSRLRSHINPTGTVLLQLENTMQMSLKD
LL

Fig. 4. 23. Amino acid sequence of the mature α-galactosidase A human. Peptides identified by the peptide mass fingerprinting analyses are highlighted in green.

The outcome had a high score and identified enough peptides to obtain a 27% of protein sequence coverage. Therefore, the purified protein was re-confirmed as the human α-galactosidase A with high confidence.

4.3.5.2. Molecular weight by MALDI-TOF

Even though the sequence-inferred size for GLA is 45.4 kDa, the real molecular weight (Mw) is expected to be several units above the theoretical Mw due to the glycosylation sites of the protein. Taking into account that the Mw for each glycosylation is between 1.7 and 2.4 kDa assuming a mannose composition from 8 to 12 mannoses (calculated with GlycanMass, http://web.expasy.org/glycanmass/), and that 3 glycosylation sites are expected for rhGLA (Garman and Garboczi, 2004), the experimental Mw should be in the range of 50.5 – 52.6 kDa. Indeed, the analysis by MALDI-TOF of the rhGLA expressed by *Pichia pastoris* was 52.3 kDa (fig. 4.24).

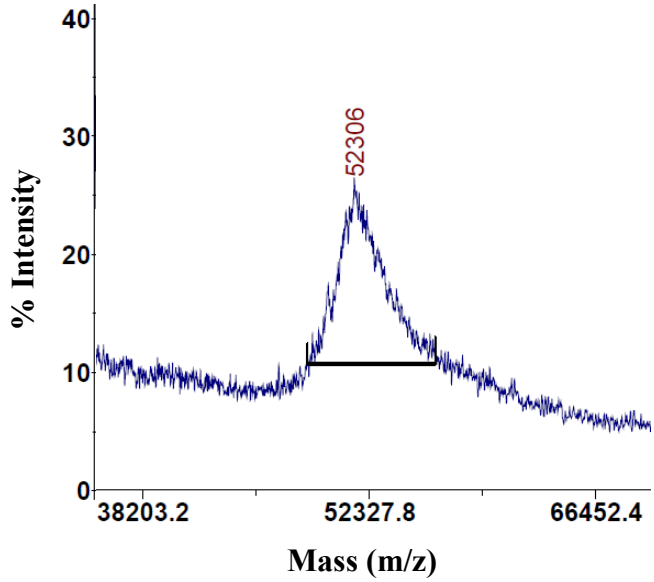


Fig. 4. 24. MS spectra of the purified rhGLA.

The result obtained is similar to those Mw reported for Fabrazyme® and Replagal®, the commercial versions of rhGLA (51.4 kDa, Lee et al., 2003), and also to the $rhGLA_{C\text{-}HIS}$ variant expressed in HEK293F (53.6 kDa, Corchero et al., 2011).

4.3.5.3. Protein size by dynamic light scattering (DLS)

Throughout some of the purification steps described in this chapter, aggregation was hypothesized to be one of the major issues responsible for decreased recovery yields. Several observations led to this assumption, such as partial enzymatic activity retention in the tangential microfiltration step, the protein loss in some chromatographic steps (mainly HIC and galactose-affinity), the "belly" profile of the rhGLA peak in SEC, and the improvement of IMAC performance when adding Tween 20 as anti-aggregating compound. To confirm this hypothesis, the proclivity of rhGLA to form aggregates was assessed by dynamic light scattering (DLS), a technique that allows to measure the metric size of particles in solution. Even though theoretical calculations of metric size from a certain Mw can be inaccurate, these estimations give a good reference to know the approximate value expected. In the case of GLA, the monomer was expected to have a size around 5 nm, meaning that the mature dimer form in solution should be around 10 nm.

The DLS results obtained for pure rhGLA showed not only a great variability between replicates of the same sample, but a wide range of sizes in each measure (mainly from 100 to 1000 nm, far bigger than the expected 10 nm size, fig 4.25).

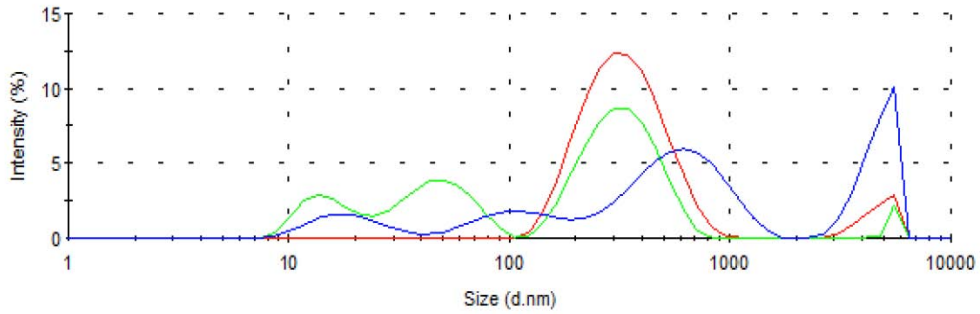


Fig. 4. 25. Analysis by DLS of pure rhGLA size. Red, green and blue profiles correspond to replicate measures of the same sample. rhGLA was in 100 acetate buffer, pH 4.5 (activity assay buffer).

Aiming to verify that the measured sizes corresponded to aggregates, different concentrations of Tween 20 were added as antiaggregating substance to aliquots of the previous sample, which were then re-analyzed (fig. 4.26).

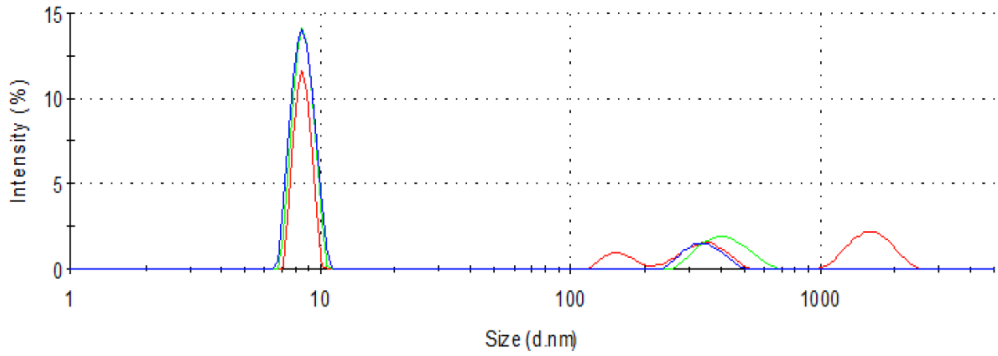


Fig. 4. 26. Analysis by DLS of pure rhGLA size. Samples analyzed were supplemented with 0.002% (red profile), 0.01% (green profile) or 0.02% (blue profile) of Tween 20.

This time, all measures shifted to the range of 7-10 nm (very close to the theoretically expected size of the GLA dimer), independently of the Tween 20 concentration tested. Even though some small peaks were detected in the range of 100 – 1100 nm, these could correspond either to the presence of small amounts of impurities in the sample or to irreversible aggregates. Noteworthy, DLS measures by %intensity are not size-independent: a given number of small particles show an intensity much lower than the same amount of big particles. These results proved that pure rhGLA was prone to form aggregates, as previously reported in the literature especially when the enzyme is overexpressed (Ioannou et al., 1992). Although samples had high levels of activity (maybe due to the dynamic nature of the aggregates generated) and the addition of an antiaggregating compound did not mean a relevant increase, the presence of such big particles seemed to have an impact in the yield of some chromatographic steps. As mentioned before in this

chapter, it would be worthwhile to test the addition of Tween 20 throughout the process looking forward to increasing the overall recovery yield of the purification processes.

4.3.5.4. Glycosylation profile

One of the critical parameters of therapeutic recombinant proteins is the glycosylation profile (Costa et al., 2014). Apart from being usually a key element for the proper folding, stabilization and biological activity, glycans play a major role in the bloodstream half-life of proteins and, consequently, in the prolongation of their therapeutic effect per dose (Huang et al., 2014). On the other hand, improper or non-native glycosylation patterns can easily lead to a fast bloodstream clearance, mainly due to the recognition by the reticuloendothelial system in the liver. This is the case of yeast high-mannose glycosylation (Bollok et al., 2009). However, and as explained in the general introduction, yeast glycan profile offer a good starting point to achieve the human M6P pattern that would not only increase the half-life of the recombinant lysosomal enzymes (as it is not efficiently recognized by the macrophages in the liver) (Beck, 2010), but also its cellular uptake driven by the M6P receptors.

In order to characterize the glycosylation profiles of modified strains generated in the future, it becomes essential to define a procedure to characterize such structures. With this purpose, and in collaboration with the Department of Analytic Chemistry of the University of Barcelona, the analysis of purified rhGLA was carried out. The results obtained confirmed that the peptides of the mature protein containing N108, N161 and N184 were indeed glycosylated, whereas no glycoform was detected for the peptide containing N377. This result is in accordance with the data reported in the literature for rhGLA produced in mammalian cell lines (Garman and Garboczi, 2004; Lee et al., 2003). Once confirmed the glycosylated peptides, a deeper study was performed to define the population of glycoforms for each one of them. After releasing the glycans,

samples were analyzed by HPLC (example corresponding to N184 shown in fig 4.27). According to the results obtained, all glycosylation sites showed similar distribution of glycoforms, although minor differences could be appreciated (fig 4.28).

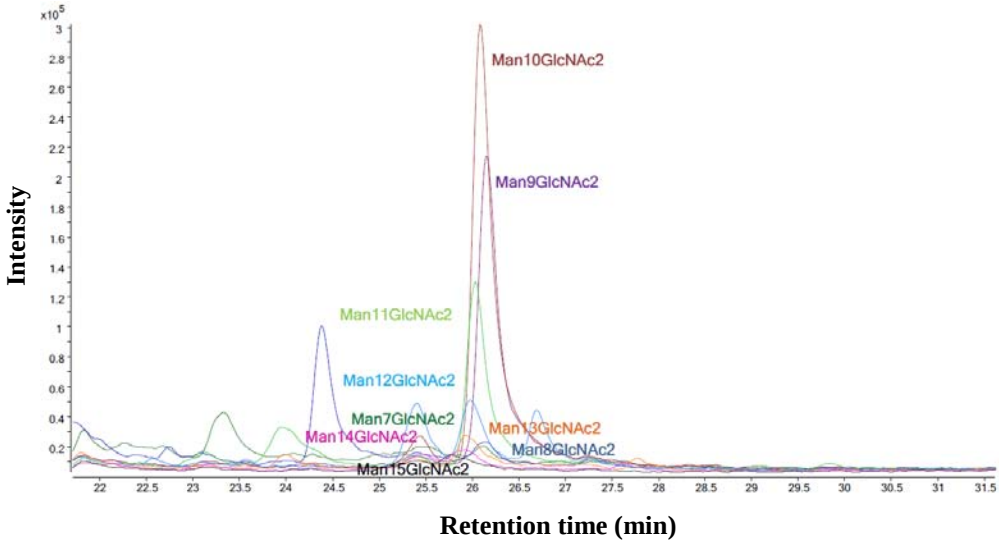


Fig. 4. 27. Chromatogram corresponding to the glycans released from the glycopeptide N184. The different glycoforms are represented in different colors.

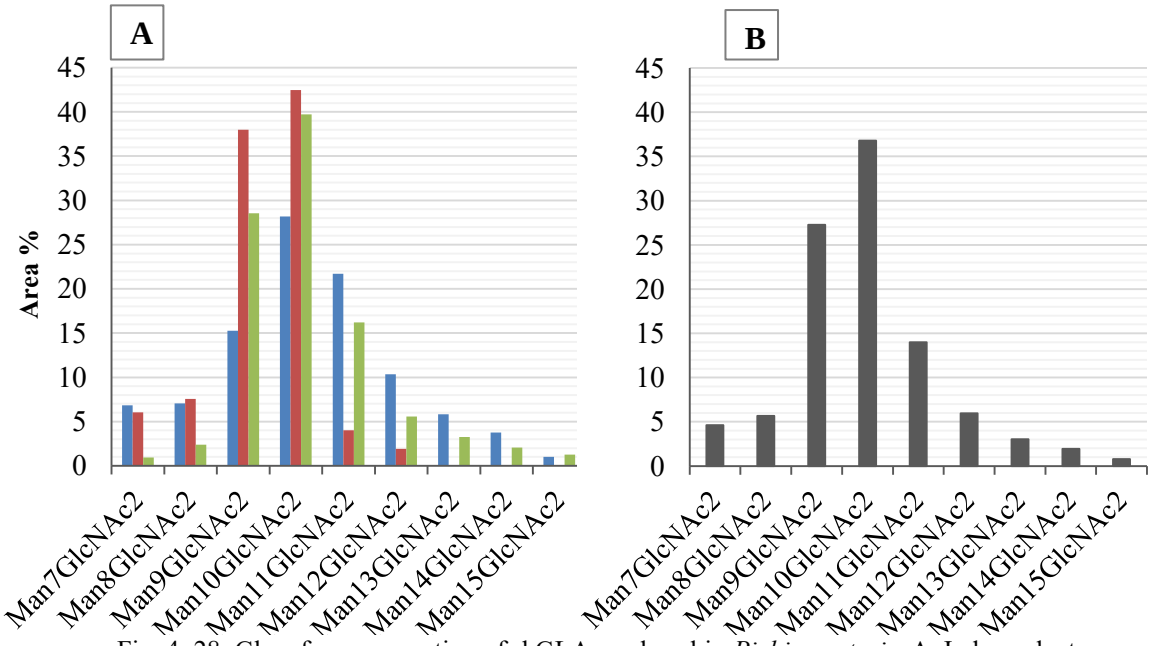


Fig. 4. 28. Glycoform occupation of rhGLA produced in *Pichia pastoris*. A. Independent distribution of glycoforms for each glycosylation site (Blue: glycopeptide N108. Red: glycopeptide N161. Green: glycopeptide N184). B. Overall glycoform distribution of rhGLA.

The major glycoforms detected contained between 9 and 11 mannose, being 10 the main structure for the 3 glycosylation sites. However, some relevant differences could be detected among them:

- N108: Man10 > Man11 > Man9
- N161: Man10 > Man9 > Man8
- N184: Man10 > Man9 > Man11

Even though a wide range of structures were detected, the absence of hypermannosylation could be confirmed. Considering that the mannose-6-phosphate glycosylation contain between 7 and 9 mannose residues , this would mean that the wild type *Pichia pastoris* glycan population contain a 37% of potential substrate structures for the addition phosphate groups. Although homogeneity of glycoproteins is a key feature for its use as therapeutic proteins, this could mean that the wild type strain would be useful to check the suitability of *Pichia pastoris* to achieve the M6P glycosylation profile. In case of detecting efficient mannose phosphorylation, such homogeneity could be achieved by further strain modification (e.g. disrupting the OCH1 gene responsible for the mannose extension).

4.4. Conclusions

- A purification process, based on a 3-step chromatography scheme, was defined for the purification of rhGLA. This scalable process is useful to obtain enough pure enzyme to eventually perform *in vitro* and *in vivo* assays.

- GLA variants (rhGLA, $rhGLA_{\Delta KDLL}$, $rhGLA_{C\text{-}HIS}$, $rhGLA_{C\text{-}RGD}$) were efficiently purified by galactose affinity chromatography. Despite the low yield obtained, the amount of enzyme obtained was enough for its catalytic characterization.

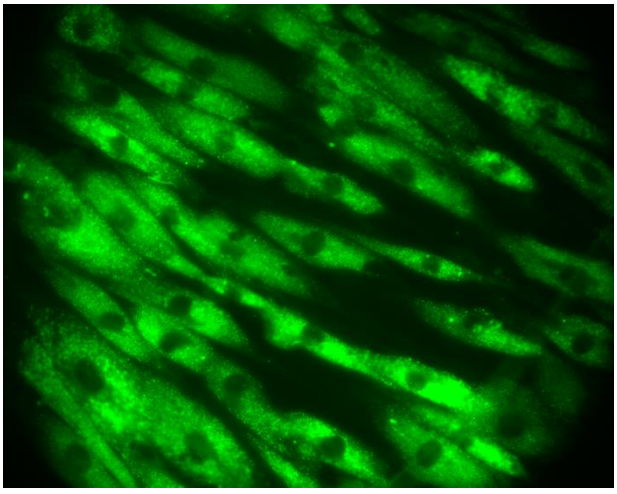
- The deletion of the 4 C-terminal amino acids of GLA (variant $rhGLA_{\Delta KDLL}$) leads to an increase of its specific activity. On the other hand, addition of C-ter sequences (variants $rhGLA_{C\text{-}HIS}$, $rhGLA_{C\text{-}RGD}$) is detrimental to the catalytic performance.

- A metal-affinity chromatography was set up for the purification of the His-tagged variant ($rhGLA_{C\text{-}HIS}$). This process will be useful for the purification of the enzyme to analyze the glycosylation pattern of future *P. pastoris* strains generated.

- A set of analytical procedures, including the glycosylation analysis, were implemented to identify and characterize rhGLA produced in *P. pastoris*.

Chapter 5

Development of *in* vitro assays for the determination of human recombinant α-galactosidase A uptake and intracellular activity



C h a p t e r 5

DEVELOPMENT OF *IN VITRO* ASSAYS FOR THE DETERMINATION OF HUMAN RECOMBINANT α-GALACTOSIDASE A UPTAKE AND INTRACELLULAR ACTIVITY

5.1. Introduction

In vitro assays are widely accepted as a useful tool to evaluate the bioactivity of promising compounds prior to undergo to *in vivo* experiments. The definition of a series of methods to assess the performance of the recombinant human alpha-galactosidase A is described in the present chapter.

By using relevant cell lines, intensive evaluations of a wide array of different compounds can be carried out focusing specifically on the expected bioactivity. Moreover, no special formulation is usually needed for the tested active principle apart from keeping it in a suitable stability buffer (especially in the case of proteins). In early development stages, *in vivo* assays would represent an unnecessary waste of resources. In first place, a high amount of animals would be used and later sacrificed, representing an undeniable bioethical issue. Managing animal facilities and maintaining animal colonies is costly and should be minimized to keep profitability of projects in early stages of development. Furthermore, animal are systems much more complex than cell cultures of isolated cell lines. This fact leads mainly to 3 complications: first, the amount of sample required is always higher in animal assays than in cell cultures. Second, active principles have to be formulated in order to yield the expected bioactivity, by ensuring a proper half-life and bioavailability. And third, sample analysis require usually more sophisticated methods. Even though *in vivo* assays are always a requisite in pre-clinical assays, especially with those aimed to be used as

biopharmaceuticals in humans, previous *in vitro* screenings with cell cultures help to lower costs notably and focus later in *in vivo* only on those promising leads.

In the case of human alpha-galactosidase A, these assays should be able to assess not only its intracellular catalytic activity, but also the internalization capability of the enzyme. In fact, the former is the critical feature why recombinant systems usually fail to produce efficient lysosomal enzymes, mainly because of their inability to yield mannose-6-phosphate glycans. The development of *Pichia pastoris* strains is expected to deliver different glycosylation variants and intermediates, and even though the presence of such modifications can be chemically analyzed, life systems will ultimately provide the real evidence of functionality.

This chapter describes the implementation of different methodologies to study the biological efficiency of rhGLA following different strategies. Specifically, 3 procedures were developed, based on (1) quantification of active internalized enzyme by extraction and activity assay, (2) quantification of degraded intracellular substrate by flow cytometry, or (3) qualitative evaluation of degraded intracellular substrate by fluorescence microcopy.

These methods were initially tested in 2 different cell lines in order to stablish, according to the consistency of the results and their respective ease of cultivation, which of them represents the best choice for the evaluation of rhGLA. Specifically, HUVEC (Human Umbilical Vein Endothelial Cells) and fibroblasts isolated from a Fabry disease patient (and therefore lacking endogenous GLA activity) were assessed.

Finally, a comparison between rhGLA and one of the available therapeutic products (Agalsidase alfa) was carried out with the established procedures to compare qualitatively and quantitatively the performance of both enzymes. This experiments helped to define the starting point of the unmodified rhGLA produced

in a wild type *P. pastoris* strain, as well as the expected *in vitro* efficiency of future variants.

5.2. Materials and methods

5.2.1. Proliferation, subculture and preparation of cell lines for *in vitro* assays

Fibroblasts from a clinically affected Fabry patient (Coriell Institute for Medical Research (US), #GM02775,) were cultured in M106 culture medium (ThermoFisher, #M-106-500) supplemented with Low Serum Growth Supplement (ThermoFisher, #S-003-10) and antibiotic-antimycotic solution, 100x (ThermoFisher, #15240-112) at 37 ºC and under 5% CO_2, humidified atmosphere. Human umbilical vein endothelial cells (ThermoFisher, #C-003-5C) were cultured in M200 medium (ThermoFisher, #M-200-500) supplemented also with Low Serum Growth Supplement and antibiotic-antimycotic solution at 37 ºC and under 5% CO_2, humidified atmosphere.

New cell cultures for in vitro assays were taken from a working cell bank stored in liquid nitrogen. 1 mL frozen vials were thawed at 37 ºC for 1-2 min and immediately mixed with 20 mL of the corresponding fresh supplemented culture medium pre-warmed at 37 ºC. Cell solutions were seed in 75 cm^2 Nunclon® Δ Surface flasks (Nunc, #156472). Subcultures were performed in order to reach significant amounts of cells for the in vitro assays. Once the cultures reached confluence, culture medium was removed and monolayers were washed with PBS (pH 7.4, Ca^{2+} and Mg^{2+} free) and treated with 5 mL of trypsin-EDTA 0.25% (ThermoFisher, #25200-056) for 7 min (for fibroblasts) or 3 min (for HUVEC) at 37 ºC. Trypsin was then neutralized by adding 5 mL of trypsin neutralizer (ThermoFisher, #R-002-100) and cells were gently resuspended by pipetting. Cell concentration was determined by cell counting in a Neubauer

hemocytometer while trypsin-neutralized cell solutions were centrifuged at 100 g, 5 min. The corresponding culture medium was used to resuspend each cell type and to adjust the concentration to $5 \cdot 10^4$ cells/mL. The cell suspension was then re-seeded in new 75 cm^2 culture flasks.

Culture media was renewed after 24 hours of seeding, and periodically after every 3 days of culture.

To perform the *in vitro* assays, and following the proliferation procedures described previously, both cell lines were subcultured either into 48-well TC-treated culture plates (Corning, #3548) for the quantification by activity measure of internalized enzyme and for the fluorometric quantification of internalized enzyme by cytometry, or into Lab-Tek® II 8-well Chamber Slide™ (Nunc, #154534) for the evaluation of internalized enzyme by fluorescence microscopy (fig. 5.1).

Fig. 5. 1. Lab-Tek® II 8-well Chamber Slide™. Treated, non-fluorescent microscopy glass allows cell attachment and growth for further fixation and microscopy examination.

When cultures reached confluence, 10 μg/mL of the fluorometric substrate N-Dodecanoyl-NBD-ceramide trihexoside (NBD-Gb3, Santa Cruz Biotechnology, #sc-360252,) were added to those culture wells aimed to be assayed fluorometrically by flow cytometry or microscopy. The cells were incubated for 24 h at 37 ºC, 5% CO_2 to allow the substrate to be internalized and accumulated within the cells. HUVEC cells were incubated simultaneously with 500 μM of

deoxygalactonojirimycin hydrochloride (DGJ, Sigma-Aldrich, #D9641) to inhibit endogenous activity of GLA. After the incubation with NBD-Gb3 (and DGJ in the case of HUVEC cells), all the cultures were washed once with PBS and different concentrations of rhGLA were applied. Two assays were sequentially conducted to set up the methodology for the *in vitro* analysis of the internalization capacity and intracellular activity of the recombinant GLA. In the first one, the purified recombinant enzyme produced in *Pichia pastoris* was diluted in supplemented M106 (for Fabry fibroblasts) or M200 (for HUVEC) media at increasing concentrations (0, 0.001, 0.005, 0.01, 0.05 and 0.1 AU/mL). Importantly, activity units used in this chapter refer to μmols MU · mL^{-1} · h^{-1} (unlike activity units used in other chapters, nmols MU · mL^{-1} · h^{-1}). Each concentration was tested by triplicate for the quantification by activity measure of internalized enzyme and for the fluorometric quantification by cytometry, and by duplicate for the evaluation by fluorescence microscopy. Cell monolayers were incubated for 24h at 37 ºC, 5% CO_2 with each rhGLA concentration. For the second assay, only Fabry fibroblasts were cultured to compare the performance of the rhGLA in comparison with the therapeutic enzyme Agalsidase alfa. Increasing concentrations of rhGLA (0, 0.01, 0.1 and 1 AU/mL) diluted in supplemented M106 media were applied by triplicate cultures for the quantification by activity measure of internalized enzyme and for the fluorometric quantification by cytometry, and duplicates for the evaluation by fluorescence microscopy. Cell monolayers were incubated for 24h or 48h at 37 ºC, 5% CO_2 with the different rhGLA concentrations.

5.2.2. Quantification of internalized enzyme by activity measurement

Culture monolayers assayed with increasing concentrations of rhGLA or Agalsidase alfa were washed 3 times with PBS to eliminate all the non-internalized GLA. 50 μL of lysis buffer (10 mM Tris, 100 mM NaCl, 5 mM EDTA, 2 mM protease inhibitor cocktail (Complete™ mini, EDTA-free,

11836153001, Roche), 1% NP-40 (NP40S, Sigma-Aldrich)) were added to each well and the monolayers were scratched mechanically to de-attach the cells. Cell suspensions were incubated in ice for 30 min. and then centrifuged at 2000 g, 10 min., 4 ºC. Cell debris (pellet fraction) were discarded and GLA enzymatic activity of supernatants was measured.

5.2.3. Fluorometric quantification of internalized enzyme by cytometry

Culture monolayers assayed with increasing concentrations of rhGLA or Agalsidase alfa were washed once with PBS and trypsinized (50 µL of trypsin-EDTA 0.25% per well) for 7 min (Fabry fibroblasts) or 3 min (HUVEC). Immediately after the trypsin incubation, 50 µL of trypsin inhibitor per well were added and cell suspensions were centrifuged at 100 g, 5 min, 4 ºC. Supernatant was discarded and cell pellets were fixed by resuspension in 50 µL of 4% formaldehyde diluted in PBS. Cell samples were analyzed with the Gallios Flux cytometer (Beckman Coulter), evaluating between 2000 and 3000 cells within 3 min per sample. The fluorescence measure of each sample was calculated by averaging the fluorescence signal of all the cells analyzed in each run.

5.2.4. Evaluation of internalized enzyme by fluorescence microscopy

As described above, cell cultures for the fluorometric evaluation by microscopy were seeded into LABTEK II 8-well Chamber Slide™, which are specially designed for its use in fluorescence microscopy. After the incubation with increasing concentrations of rhGLA or Agalsidase alfa, cultures were washed twice with PBS. Then, monolayers were fixed by adding 3.5% paraformaldehyde and incubation at 4ºC for 5 min, followed by 10 min at room temperature and 3 washing steps with HEPES (Sigma-Aldrich, #H4034). Finally, fixed cells were blocked by adding 100 mM Tris-HCl, pH 8. Fluorescence of cell cultures was

evaluated with a motorized inverted fluorescence microscope (Leica DMIRBE), equipped with camera (CoolSNAP fx). Images were taken using the Metamorph software. Image analysis was done with ImageJ software.

5.3. Results and discussion

5.3.1. Evaluation of 2 cell lines (human Fabry fibroblasts vs. HUVEC) and different techniques for the evaluation of the uptake capacity of recombinant human α-galactosidase A

Two different cell lines were acquired to be used in an initial study of different methods aimed to evaluate the internalization and intracellular activity of the rhGLA produced in *P. pastoris*. The idea was to assess, simultaneously, the suitability of the methodologies and proposed cellular types for the fast, reliable and affordable screening of multiple enzyme variants to be produced in the future by humanized yeast strains.

The first cell line selected were fibroblasts derived from a 16-year old Fabry disease patient. This culture was obtained from the Coriell Institute for Medical Research (Camdem, NJ, USA), whose renowned biobank provides cell lines from a vast number of human diseases (including a great number of lysosomal disorders). The selected cells lacked completely GLA activity, so any observed catalytic activity would correspond to the supplemented enzyme. Fibroblasts are the most common cell type in connective tissue in animals, and are derived from primitive mesenchyme.

Alternatively, HUVEC (Human Umbilical Vein Endothelial Cells) were tested. These cells are derived from the endothelium of the umbilical cord and are often used as a model system for pathogenesis studies of endothelial cells (Park et al., 2006). Unlike fibroblasts, they have a mesodermal origin. HUVECs are primary

cells whose performance and phenotypic characteristics are maintained only for a few population doublings, and possess endogenous GLA activity. In this case, cells from a pooled donor were supplied by Lonza.

A total of three methods were evaluated for both types of cells; the 1st and 2nd were based on quantitative measures (by intracellular extraction and activity assay or fluorescence flow cytometry), while the 3rd relied on qualitative evaluations of the internalized enzyme (by fluorescence microscopy). Of course, 2 common steps of cell growth and incubation with enzymatic solution were carried out for any of the different evaluation experiments. In the specific case of HUVECs, a treatment with a GLA inhibitor (DGJ) was carried out before the treatment with recombinant GLA, in order to inhibit and minimize the effect of the endogenous enzyme. In all cases, purified rhGLA (unmodified, with its native amino acid sequence) was used as the reference recombinant variant.

5.3.1.1. Quantification by extraction and activity measure of internalized enzyme

The first assay was based on the measurement of internalized activity after 24h of incubation in presence of defined concentrations of rhGLA (based on enzymatic activity measures). After the incubation, cell monolayers were washed to discard any trace of extracellular enzyme remains. Monolayers were then mechanically de-attached (by scrapping the surface) and submitted to a brief treatment of detergent (NP-40), aiming to solubilize cell membranes and free the intracellular proteins, eventually including the internalized enzyme. The results obtained from activity measures of extracted samples (fig. 5.2) showed that, for Fabry fibroflasts, the increase in the concentrations of rhGLA in the culture medium were clearly linked to the activity measures of cell extracts. Even though this behavior was roughly observed for HUVECs too, variability was notably higher between replicates,

leading to unexpected differences in measures between different enzyme concentrations in culture (i.e., between 0.01 and 0.05 AU/mL).

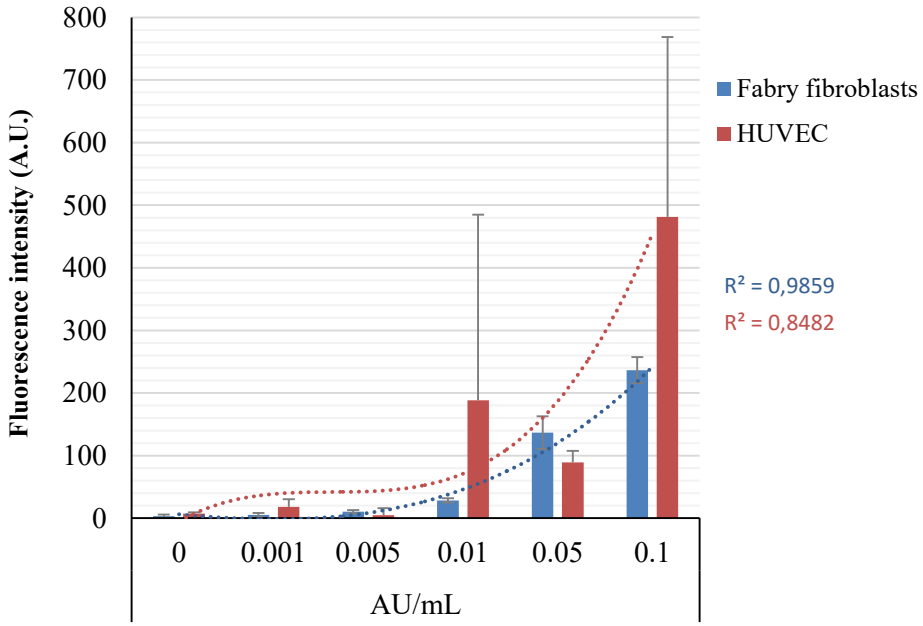


Fig. 5. 2. Activity measures of internalized GLA in Fabry fibroblasts (blue bars) and HUVEC (red bars) treated with different concentrations (expressed in AU/mL) of recombinant GLA produced in *P. pastoris*.

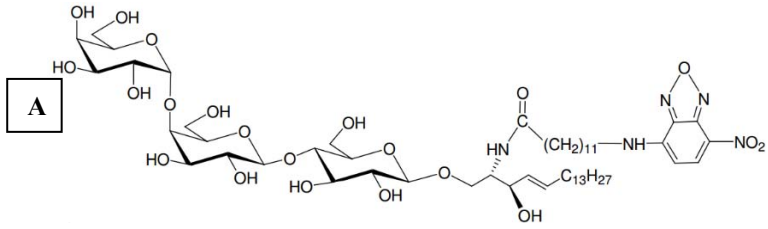
Although the origin of such differences is unclear, there are a number of presumable sources that could be responsible. The first one is the characteristic growth profile of HUVECs, which group themselves in clusters from where they start to spread throughout the surface (unlike what occurs with fibroblasts, which are able to grow individually and occupy the surface quite uniformly until they reach confluence). This profile can lead to significant differences in the amount of cells accumulated in each culture well at the end of the assay. To avoid this effect,

it would have been useful to normalize the activity measure of each culture well by the total amount of cells present on it. However, here resides one of the limitations of this method: cells are usually counted by means of a hemocytometer or by flow cytometry, but these methods require the previous de-attachment of cells from the culture surface (usually by a trypsin treatment) when working with adherent cells. The presence of trypsin remains could digest intracellular proteins released, including the internalized enzyme aimed to be quantified (in fact, this was the reason for using a scrapper to de-attach cells). One solution would be to use specific microscopy equipment where cells could be counted in the monolayer prior to de-attachment. Another approach would be the inhibition of trypsin after the treatment, and wash the cells in solution prior to cell lysis. A second reason for the observed variability is the presence of endogenous GLA activity in HUVECs. Even though these cells were submitted to a treatment with a GLA inhibitor (DGJ) before the treatment with different concentrations of the enzyme, cells could have recovered (at least partially) the endogenous activity during the rhGLA incubation, leading to a combination of their own and the externally supplied GLA activity. The only way to avoid this effect is to work with a *GLA* gene knockout endothelial cell line, but this option was not available by that time.

Fabry fibroblasts showed the best dose-response behavior for this assay, probably thanks to the culture homogeneity between replicates (in terms of proliferation and cellular concentration) and the total absence of endogenous GLA potentially interfering with the measure of recombinant enzyme internalized. Moreover, their growth profile and ease of culture was superior to that observed in HUVEC. Regarding the methodology studied, it showed a clear correlation between the rhGLA concentration administered and the amount of activity later recovered from cell extracts. Although this procedure allows to evaluate in a simple manner the uptake of the recombinant enzyme, its intracellular efficacy on the degradation of subcellular substrates cannot be inferred from the results obtained. Thus, the methods discussed below were developed to cover this point.

5.3.1.2. Fluorometric quantification of internalized enzyme by flow cytometry

In front of the need to determine the efficiency of the internalized GLA on the degradation of intracellular substrates, an alternative method based on the measure of a fluorescent substrate was studied. Specifically, before the incubation with rhGLA samples, cell cultures were treated with N-Dodecanoyl-NBD-ceramide trihexoside (NBD-Gb3, a fluorescent analog of natural ceramide trihexoside, a known substrate of GLA). This molecule, which is easily internalized by the cells, contains a 7-nitrobenzofurazan (NBD) group with maximum excitation and emission wavelengths of 465 nm a 535 nm, respectively (fig. 5.3). Once the ceramide trihexoside is catalytically cleaved by GLA, a loss of fluorescence can be seen and measured.



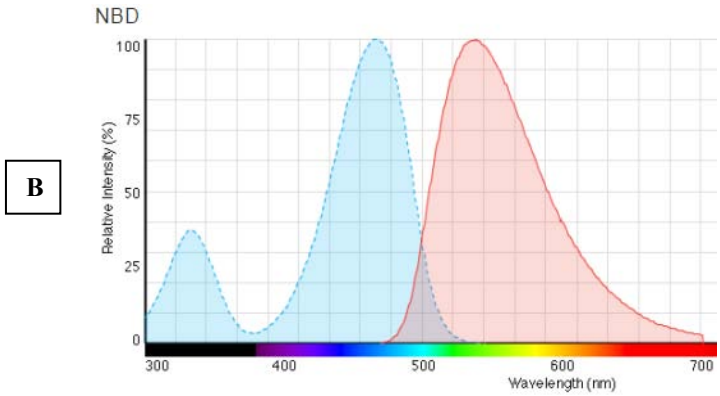


Fig. 5. 3. (A) Chemical structure of N-Dodecanoyl-NBD-ceramide trihexoside (NBD-Gb3). (Source: Matreya LLC, USA). (B) Excitation-emission fluorescence spectrum of NBD. (Source: ThermoFisher Scientific).

In the case of HUVEC, the treatment with NBD-Gb3 was carried out simultaneously with deoxygalactonojirimycin (DGJ). This galactose analog acts as a competitive inhibitor of GLA. Interestingly, at subinhibitory concentrations it can act as a molecular chaperone for unstable enzyme mutants, allowing its trafficking to the lysosomes (Benjamin et al., 2009). Under the name of Migalastat, the molecule is currently being tested as a therapeutic agent for the treatment of Fabry patients, alone or in combination with current ERTs. However, in the interest of the present method, HUVEC cells were treated with inhibitory concentrations of DGJ aiming to mask the effects of endogenous GLA.

After the incubation with different rhGLA, cultures were treated with trypsin to de-attach cells. Samples were analyzed by flow cytometry enabling cell counting as well as quantification of cell fluorescence (fig. 5.4).

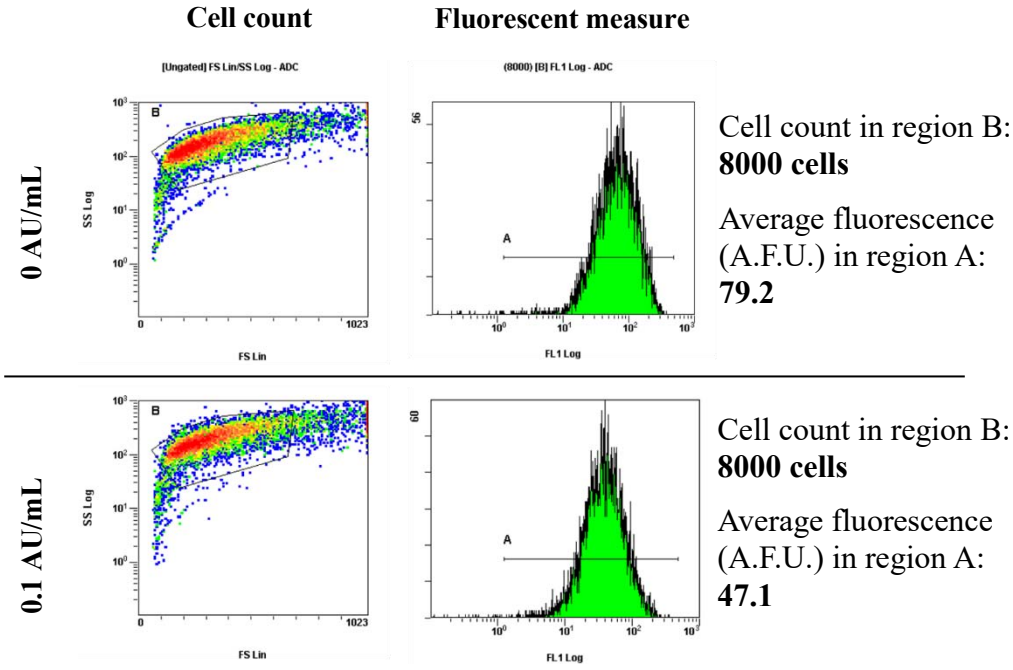


Fig. 5. 4. Example of the results obtained from fluorescent cell counting by flow cytometry. Dot plots of fluorescent cells (left) and their corresponding fluorescence histogram (right), for an untreated sample (up) and a sample treated with 0.1 AU/mL of rhGLA (down) are shown. The data correspond Fabry fibroblasts.

Measures of Fabry fibroblasts showed a good correlation between cellular fluorescence and concentration of rhGLA supplemented in culture, reaching a minimum remaining fluorescence of 55% with respect to the fluorescence measured for untreated cells (fig 5.5). On the other hand, this correlation was not observed for HUVECs, which reached a similar minimum fluorescence value of 61% with the lowest rhGLA concentration tested (0.001 AU/mL). Even though a lower value was obtained with 0.005 AU/mL (42%), it was considered an outlier because the results obtained for higher concentrations of rhGLA showed again remaining fluorescence of 56%, equivalent to the value observed for fibroblasts treated with the same rhGLA concentration. It was assumed that the remaining fluorescence could be due either to inaccessible substrate to the internalized enzyme, or to undegraded substrate still present on early endosomes, where the internal pH would not be low enough to yield significant GLA activity, since the

enzyme is expected to be fully active in the acid environment provided inside mature lysosomes.

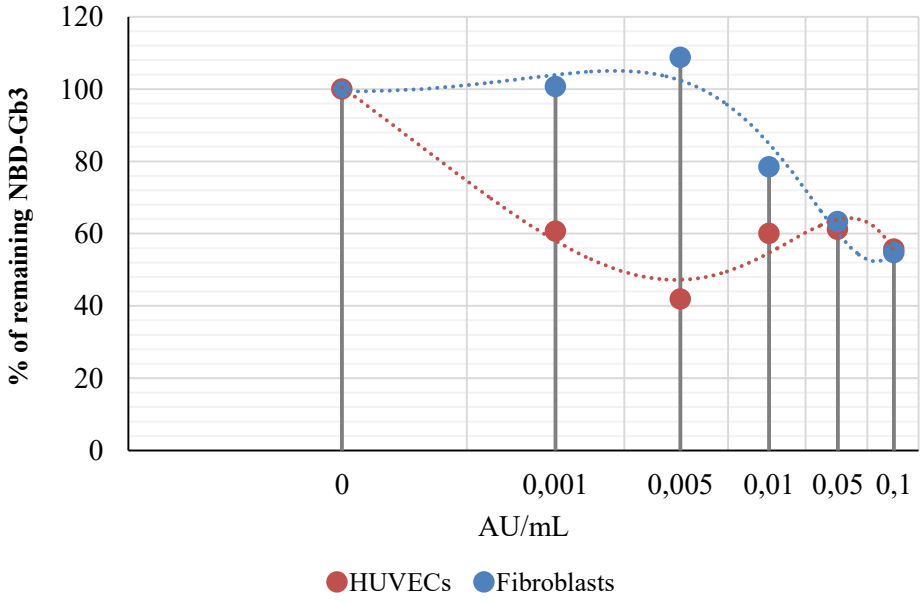


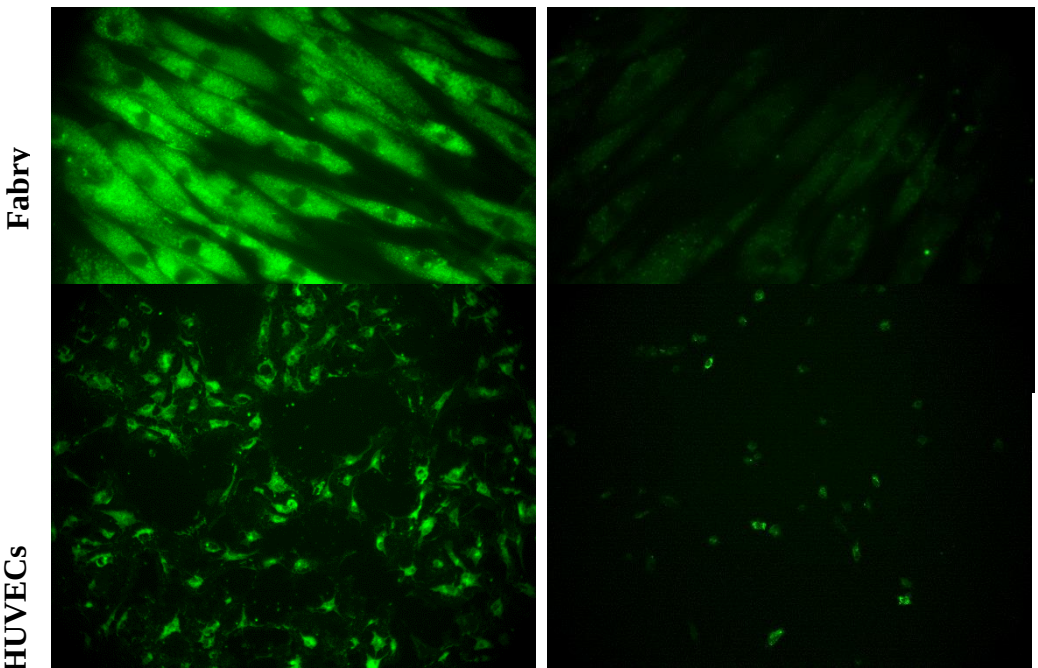
Fig. 5. 5. Degradation profile of the fluorescent substrate NBD-Gb3 accumulated in Fabry fibroblasts (blue dots) or HUVEC (red dots) treated with increasing concentrations of rhGLA.

The poor correlation between substrate degradation and the concentration of rhGLA supplemented to the cultures in HUVECs remains unclear. Likely, and as hypothesized in the previous method, it could be due to the inefficient inhibition of endogenous GLA, which would degrade the fluorescent substrate synergistically with the externally supplied rhGLA. In any case, Fabry fibroblasts showed again the most reliable correlation between the increasing rhGLA doses and the expected response. The flow cytometry-based method reflects, in contrast with the extraction and quantification of internalized enzyme previously described, the capability of rhGLA to degrade intracellular substrates.

5.3.1.3. Evaluation of internalized enzyme by fluorescence microscopy

The 3rd procedure tested was the qualitative evaluation of NBD-Gb3 degradation by fluorescence microscopy. In this assay, cultures were treated identically to those used in fluorescent cytometry. However, this time cells were grown in culture chambers made with non-fluorescent glass microscope slides at the bottom, which is specially treated to support growth of adherent cells. Moreover, a procedure of cell fixation was carried out instead of the trypsin treatment, in order to check the fluorescence of the cultures by microscopy directly in the chambers.

In order to simplify the assay, only the maximum concentration of rhGLA (0.1 AU/mL) was tested together with the negative controls lacking the enzyme, for both cell types. A green filter was used to expose the green fluorescence expected inside the cells. Control cultures untreated with rhGLA helped to differentiate clearly the internalization of the fluorescent substrate (NBD-Gb3) in both cell types (fig 5.6). As expected, it was distributed throughout the whole cells except for the nucleus, and some small highly fluorescent spots could be appreciated maybe corresponding to insoluble cumulus of substrate (since some of them were still visible in cultures treated with rhGLA).



In both cell types (Fabry fibroblasts and HUVECs), a decrease in cell fluorescence was absolutely evident. Yet, the remaining fluorescence seemed to be homogenous in most of the cells observed, while HUVECs showed higher heterogeneity with some cells still retaining significant amounts of fluorescent substrate and others becoming almost invisible due to its absence.

In general, observation of fibroblasts was more consistent mainly thanks to its better attachment to the available surface. In contrast to what was observed for HUVECs, grown fibroblasts cultures in the microscopy chambers were able to bear the different washing steps required during the enzyme treatment protocol and the later cell fixation procedure. This feature helped to easily find fields full of cells during microscopy inspection, in contrast to HUVEC cultures where big extensions of monolayer were lost before the final observation.

Figure 5. Image obtained by fluorescence microscopy can be used as a qualitative technique even at the cellular level. However, such calculations would require an extensive imaging analysis, and it was decided to use the method as a qualitative assessment of the rhGLA intracellular activity instead. In combination with cytometry, the efficacy of the internalized enzyme could be determined both quantitatively and by visual inspection.

Fig 5. Image obtained by fluorescence microscopy can be used as a qualitative technique even at the cellular level. However, such calculations would require an extensive imaging analysis, and it was decided to use the method as a qualitative assessment of the rhGLA intracellular activity instead. In combination with cytometry, the efficacy of the internalized enzyme could be determined both quantitatively and by visual inspection.

Fig 5. Fabry fibroblasts or HUVEC cultures, untreated or treated with 0.1 AU/mL of rhGLA.

5.3.1.4. Overview of cell types and methods evaluated

Clearly, rhGLA showed to be functional *in vitro* for all the 3 methods tested and with both cell types. Referring to the latter, significant advantages and better quality of results were observed for Fabry fibroblasts to the detriment of HUVEC

cells (table 5.1). The excellent growth profile of the fibroblasts, together with the lack of endogenous alpha-galactosidase potentially interfering in the assays, were the main properties responsible for choosing this cell line as the reference to test future rhGLA variants and discarding HUVEC. In addition, Fabry fibroblasts consist a relevant cell line for the evaluation of an enzyme intended to be used, indeed, for enzyme replacement therapy for the same disease.

Table 5. 1. Summary of cell types tested and their characteristics.

Cell type	Growth profile	Attachment strength	Endogenous GLA activity
Fabry fibroblasts	Homogeneous monolayer - Consistent cell count between replicas	Very high - Resistant to multiple washing steps	NO
HUVEC	Extension of monolayer from cell clusters - Cell count variability between replicas	Medium – Low - Easily de-attached, source of variability	YES - Previous inhibition is required (e.g. with DGJ) - Potential source of odd results

Regarding the different methods evaluated, no one was considered to stand out from the others. Even though the information obtained individually from each one could be considered redundant, the procedures provided different type of data that can help to obtain a solid outcome about the performance of the enzyme (table 4.2).

Table 6. 1. Summary of methods implemented and evaluated and their main characteristics.

Method	Studied parameter	Quantitative / Qualitative	Complexity

Extraction and quantification of internalized enzyme	Efficiency of internalization	Quantitative	LOW - No specific reagents or additional equipment required
Fluorescence cytometry	Efficiency of internalization + Intracellular activity	Quantitative	MEDIUM - Fluorescent substrate required - Use of fluorescence flow cytometer
Fluorescence microscopy	Efficiency of internalization + Intracellular activity	Qualitative	MEDIUM – HIGH - Fluorescent substrate required - Culture in specific culture chambers - Use of fluorescence microscope

The extraction and quantification of internalized enzyme resulted to be the simplest method, and no specific equipment is needed. Thus, it represents a good strategy to screen rhGLA variants eventually produced in humanized yeast strains, because the assay provides information about the internalization capacity of the enzyme. Then, promising candidates could be validated by both fluorescence-based methods that would confirm the intracellular activity over a relevant substrate. If visual documentation is not strictly necessary, the use of fluorescence microscopy could be missed, since it is a qualitative assay (unless quantification is performed by computer analysis of the images) that requires specific culture material. Besides, the information obtained about rhGLA would be similar to that acquired by fluorescence flow cytometry.

5.3.2. Cellular uptake of rhGLA produced in *Pichia pastoris* versus the therapeutic product Agalsidase alfa

Once stablished the procedures and the appropriate cell type for the evaluation of rhGLA produced in *P. pastoris*, a comparative analysis was made with the therapeutic product Agalsidase alfa (marketed by Shire Plc as Replagal®, produced by a gene activated established human fibroblast cell line).

According to the results discussed in the previous paragraphs, Fabry fibroblasts were considered to be the best choice for the evaluation of recombinant alpha-Galactosidase A. In all the procedures tested, this cell type showed the clearest correlation between enzyme doses and expected response (increase of intracellular GLA activity in the first assay, and degradation of fluorescent substrate in the second and third assays). The comparison between rhGLA and Agalsidase alfa was carried out with the 3 methods implemented, since they were considered to be complementary. This time, the maximum concentration of rhGLA was increased from 0.1 to 1 AU/mL in all the assays, and a single concentration of Agalsidase alfa was added (0.01 AU/mL).

5.3.2.1. Quantification by extraction and activity measure of internalized enzyme

After treating the monolayers with different GLA solutions, with concentrations ranging from 0.01 to 1 AU/mL, the quantification of enzymatic activity from cell extracts evidenced clear differences in the internalization efficiency. In cultures treated with the therapeutic product Agalsidase alfa, a low dose of 0.01 AU/mL was able to yield similar activity values than those observed for 1 AU/mL of rhGLA, the highest concentration tested (fig. 5.7). Thus, this result suggests that a 100 times concentrated rhGLA sample would be necessary to achieve similar uptake levels to that observed for Agalsidase alfa.

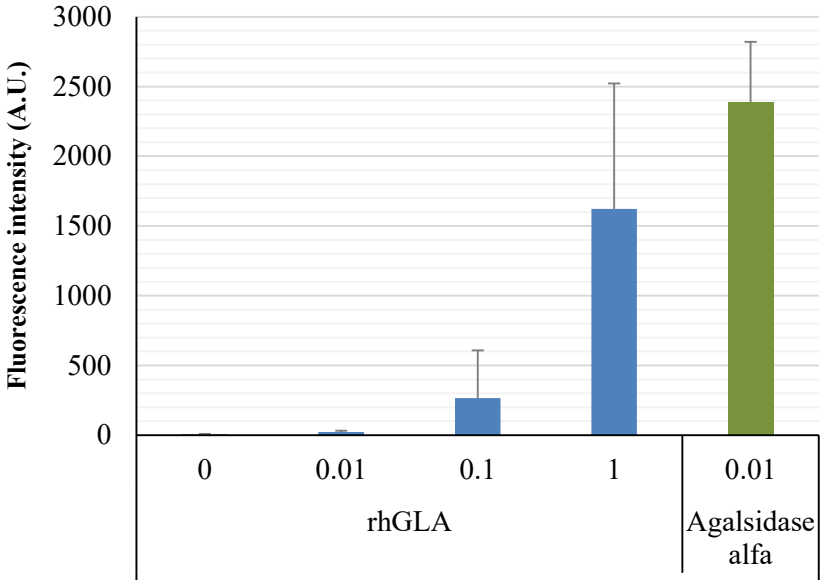


Fig. 5. 7. Activity measures of internalized GLA in Fabry fibroblasts treated with different concentrations (expressed in AU/mL) of recombinant GLA produced in *P. pastoris* (blue bars) and 0.01 UA/mL of Agalsidase alfa (green bar).

The observed difference would be mainly due to the lack of mannose-6-phosphate glycosylation structures in the enzyme produced in *P. pastoris*. As discussed in previous chapters, glycans are known to be the key signal by which lysosomal enzymes are internalized through recognition by M6P receptors present on the cell membrane. Although the correlation between dose and internalized rhGLA was evident, we suggest that the enzyme would be inefficiently internalized by means of alternative, non-specific mechanisms, such as micropinocytosis (Petros and DeSimone, 2010; Falcone et al., 2006).

5.3.2.2. Fluorometric quantification of internalized enzyme by flow cytometry

The analysis of intracellular GLA activity by fluorescence cytometry showed the expected dose-response relationship for rhGLA (fig. 5.8). However, and for unknown reasons, the observed NBD-Gb3 degradation levels were substantially lower than in equivalent experiments discussed in the previous paragraph. When compared with the cultures treated with Agalsidase alfa, rhGLA manifested again a limited efficacy. Specifically, a 100-fold increase in rhGLA concentration (1 AU/mL) was necessary to reduce the intracellular fluorescence to similar levels to those induced by treatment with Agalsidase alfa (0.01 AU/mL).

These results are in accordance with the data observed in the activity measures of internalized enzyme, confirming the enzyme deficiency in its internalization capacity.

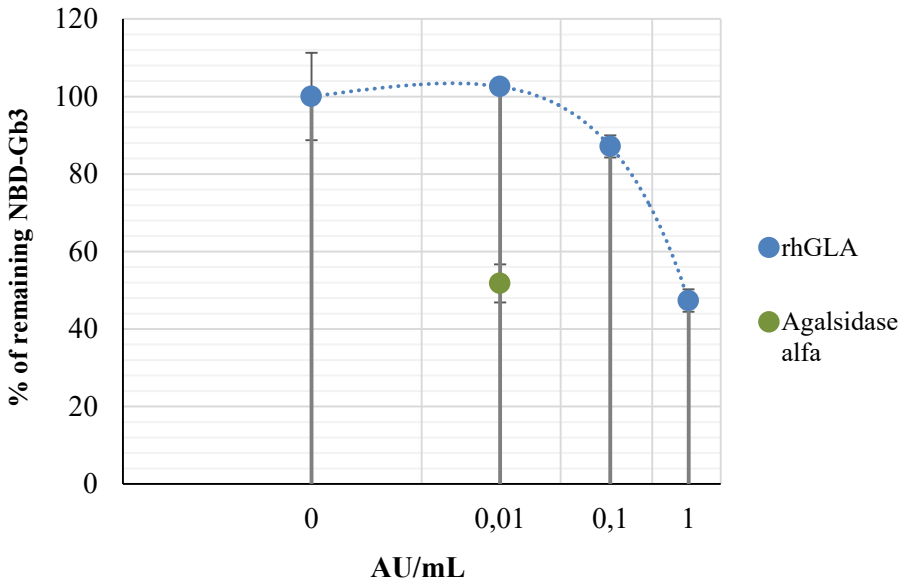


Fig. 5. 8. Degradation profile of the fluorescent substrate NBD-Gb3 accumulated in Fabry fibroblasts treated with rhGLA produced in *P. pastoris* (blue dots) and 0.01 AU/mL of Agalsidase alfa (green dot).

5.3.2.3. Evaluation of internalized enzyme by fluorescence microscopy

The cellular uptake of increasing concentrations of rhGLA, compared to a single dose of Agalsidase alfa, was also evaluated by fluorescence microscopy. As expected from the results presented in the previous paragraphs, the correlation between the dose administered to the cultures and the reduction of fluorescent substrate (NBD-Gb3) was evident (fig 5.9). Again, 0.01 AU/mL of Agalsidase alfa achieved similar substrate degradation levels to those obtained with 1 AU/mL of rhGLA. These results confirm the higher uptake efficiency of the current therapeutic enzyme produced in a human cell line, likely thanks to its high mannose-6-phosphate content.

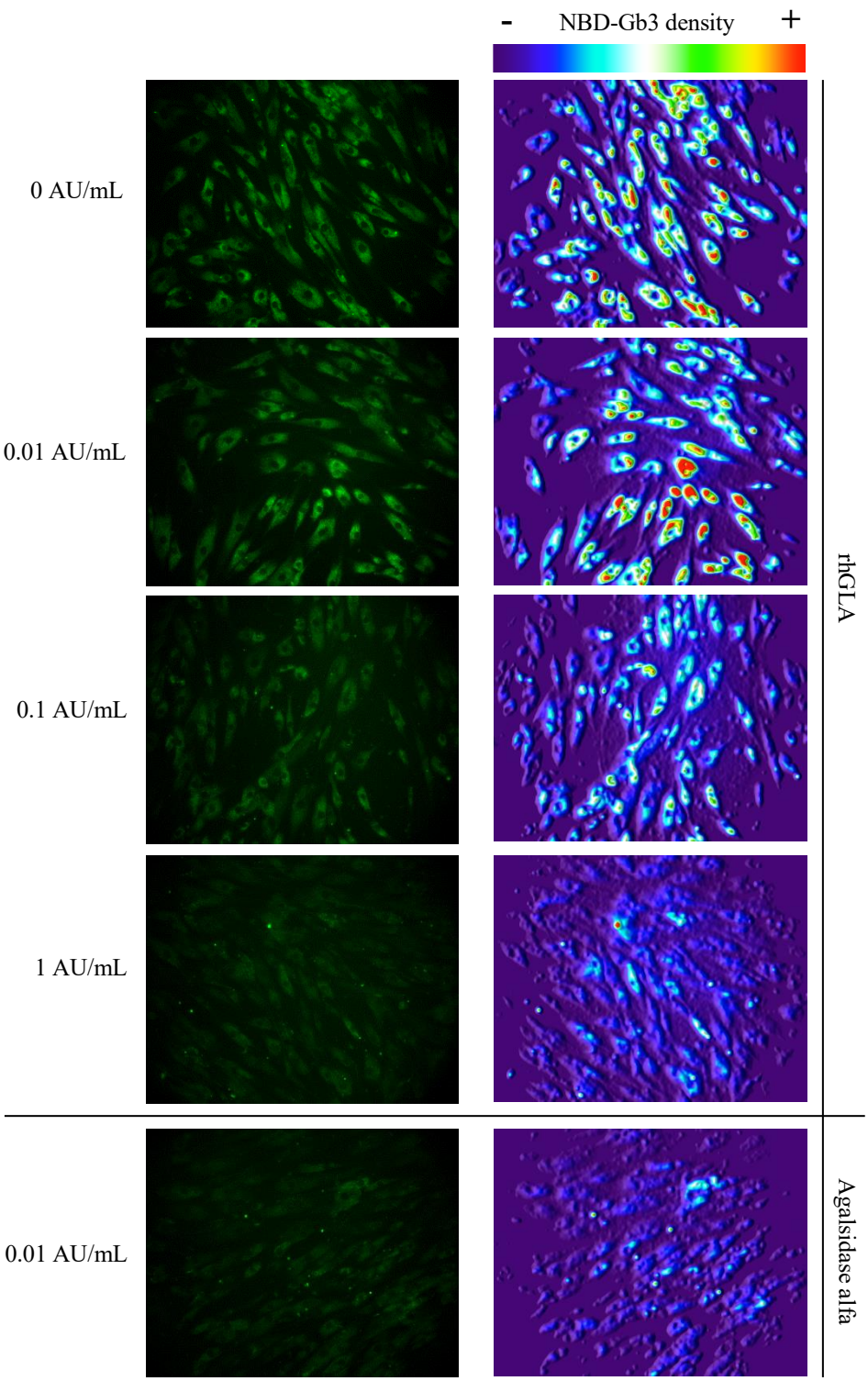


Fig. 5. 9. Images obtained by fluorescence microscopy from cultures treated with increasing concentrations of rhGLA or 0.01 AU/mL of Agalsidase alfa. A thermal filter was applied to the pictures to highlight the differential density of the fluorescent substrate.

Globally, the low internalization capacity of alpha-Galactosidase A with the native high-mannose yeast glycosylation profile was confirmed. However, these results were expected, since lysosomal enzymes require specific recognition signals to be incorporated into lysosomes. As rhGLA lacks mannose-6-phosphate glycans recognized by cation-independent M6P receptors (CI-M6PR) present in the cell surface, its intracellular activity would rely in alternative, unspecific uptake pathways which would be much less efficient. For this reason, the use of an alternative internalization signal based on a peptide signal (RGD-motif) was initially proposed as an alternative strategy for the proper enzyme uptake. Unfortunately, the results obtained suggested that modifications in the N- or C-terminus of the protein sequence (especially in the N-terminus) promoted a marked decrease in the enzyme potency. Therefore, the efforts are currently focused toward the generation of humanized yeast strains able to produce heterologous proteins with a mannose-6-phosphate glycosylation profile.

The primary objective of the work presented in this chapter was to stablish methods for the evaluation of future enzyme variants, and the results helped to define the basal *in vitro* performance of rhGLA produced in a *P. pastoris* strain (and having native high-mannose glycan structures). Modified strains to yield human glycan structures could be easily evaluated by any of the methods implemented. Moreover, the presence M6P glycans in future enzyme variants could be specifically assessed by a simple modification of the protocols: the addition of free M6P to the cultures, in order to saturate CI-M6P receptors of the cell surface. In that case, the enzyme uptake by the M6P pathway would be impaired and a marked decrease in rhGLA uptake should be seen, proving indirectly the presence of the expected glycan profile in the recombinant enzyme, and ultimately the success in the generation of a humanized *P. pastoris* strain.

5.4. Conclusions

- Fibroblasts isolated from a Fabry disease patient were stablished as the suitable cell type for the analysis of recombinant alpha-Galactosidase A. In contrast with HUVEC, Fabry fibroblasts are easier to handle and provide results of higher quality.

- The 3 methods implemented and tested showed to be complementary tools to study the cellular uptake of rhGLA. The fluorescence-based assays allowed to evidence, in addition, the intracellular activity of a relevant substrate. Therefore, the methods are of great value to analyze future enzyme variants expressed by humanized *P. pastoris* strains.

- Alpha-Galactosidase A produced in a wild-type *P. pastoris* strain, and thus having a high-mannose glycosylation pattern, evidenced a poor internalization capability in comparison with the therapeutic product Agalsidase alfa (containing mannose-6-phosphate glycans) produced in a human cell line.

GENERAL CONCLUSIONS

The global conclusions extracted from this study are highlighted below:

- Human α-galactosidase A is better expressed in a Mut^+ (normal methanol utilization phenotype) *P. pastoris* strain. Moreover, the overexpression of HAC1p together with the insertion of multiple copies of the rhGLA expression cassette (2 – 3 copies) improves significantly the productivity of the enzyme (7 fold higher than a single copy in a Mut^+ strain). These results prove that significant expression levels of rhGLA are achievable in *P. pastoris* by strain engineering.

- O_2-limited fed-batch is a good strategy for the production of rhGLA in high cell density bioreactor cultures. The process described in this work proved to be productive, reliable and cost-efficient. Presumably, it would be also easily scalable, although pilot-scale cultures should be performed to confirm this hypothesis.

- Non-tagged rhGLA can be purified following the 3-step chromatography method described in this work. The process has been developed using scalable chromatographic media and columns, and allows to obtain highly pure enzyme with an acceptable overall recovery yield. Moreover, affinity-chromatography methods, have been successfully used to obtain small quantities of pure rhGLA, providing a useful tool to get easily in a single-step process small quantities of product for its characterization.

- The fusion of small peptides to the native GLA protein sequence has an impact on its specific enzymatic activity. However, the addition of a 6xHistidine tag at the C-terminus is feasible (1/4 of catalytic activity is retained according to this study), which is of great interest for the use of this variant ($rhGLA_{C\text{-}HIS}$) as a model lysosomal enzyme in strain glycoengineering

research. On the other hand, the addition of sequences to the N-terminus of the enzyme impairs its activity and/or its expression profile.

- The deletion of the last 4 amino acids (KDLL) form the C-terminus of rhGLA induce a 1.6-fold increase in the specific activity of the enzyme. Although additional studies must be done to elucidate the effect of the modification in critical features such as protein stability or *in vitro* and *in vivo* efficacy, such improvement could potentially allow and increased physiological effect while maintaining the administered doses.

- A complete set of analytical procedures have been implemented for the efficient characterization of rhGLA. Between these methods, glycan analysis and *in vitro* assays described in this work are of special interest. They allow to define the glycosylation structures of the enzyme and assess its therapeutic effect in a relevant cell line (fibroblasts from Fabry patient). Although the result was anticipated, rhGLA with native yeast glycosylation profile showed a poor cellular uptake profile. This result highlights the importance of yeast humanization, and the methods developed here will be useful to evaluate the effect of future *P. pastoris* strain modifications aimed to adapt its glycosylation profile.

In summary, the results obtained allowed to stablish a platform for the expression of rhGLA in *P. pastoris* and enable the use of the enzyme as a relevant example of lysosomal hydrolase. This work allowed to maintain and expand the research line focused on strain glycoengineering, generating additional tools and intellectual property.

In addition, the work leading to the results presented in this work allowed the development of new tools (strains, vectors, etc.) and the integration of novel procedures and techniques at Bioingenium SL, some of which are nowadays used on a regular basis in a wide range of projects developed by the company.

Abad, S., Kitz, K., Hörmann, A., Schreiner, U., Hartner, F.S., Glieder, A., 2010. Real-time PCR-based determination of gene copy numbers in *Pichia pastoris*. Biotechnol. J. 5, 413–20.

Abian, O., Alfonso, P., Velazquez-Campoy, A., Giraldo, P., Pocovi, M., Sancho, J., 2011. Therapeutic strategies for Gaucher disease: miglustat (NB-DNJ) as a pharmacological chaperone for glucocerebrosidase and the different thermostability of velaglucerase alfa and imiglucerase. Mol. Pharm. 8, 2390–7.

Ahmad, M., Hirz, M., Pichler, H., Schwab, H., 2014. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. Appl. Microbiol. Biotechnol. 98, 5301–5317.

Arnau, C., Ramon, R., Casas, C., Valero, F., 2010. Optimization of the heterologous production of a *Rhizopus oryzae* lipase in *Pichia pastoris* system using mixed substrates on controlled fed-batch bioprocess. Enzyme Microb. Technol. 46, 494–500.

Aw, R., Polizzi, K.M., 2013. Can too many copies spoil the broth? Microb. Cell Fact. 12, 128.

Baumann, K., Maurer, M., Dragosits, M., Cos, O., Ferrer, P., Mattanovich, D., 2008. Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. Biotechnol. Bioeng. 100, 177–83.

Beck, M., 2010. Therapy for lysosomal storage disorders. IUBMB Life 62, 33–40.

Benjamin, E.R., Flanagan, J.J., Schilling, A., Chang, H.H., Agarwal, L., Katz, E., Wu, X., Pine, C., Wustman, B., Desnick, R.J., Lockhart, D.J., Valenzano, K.J., 2009. The pharmacological chaperone 1-deoxygalactonojirimycin increases alpha-galactosidase A levels in Fabry patient cell lines. J. Inherit. Metab. Dis. 32, 424–40.

Berdichevsky, M., D'Anjou, M., Mallem, M.R., Shaikh, S.S., Potgieter, T.I., 2011. Improved production of monoclonal antibodies through oxygen-limited cultivation of glycoengineered yeast. J. Biotechnol. 155, 217–24.

Berek, D., 2010. Size exclusion chromatography - A blessing and a curse of science and technology of synthetic polymers. J. Sep. Sci. 33, 315–335.

Bethencourt, V., 2009. Virus stalls Genzyme plant. Nat. Biotechnol. 27, 681.

Bill, R.M., 2014. Playing catch-up with *Escherichia coli*: using yeast to increase success rates in recombinant protein production experiments. Front. Microbiol. 5, 1–5.

Bollok, M., Resina, D., Valero, F., Ferrer, P., 2009. Recent patents on the *Pichia pastoris* expression system: expanding the toolbox for recombinant protein production. Recent Pat. Biotechnol. 3, 192–201.

Bornhorst, J.A., Falke, J.J., 2000. Purification of proteins using polyhistidine affinity tags. Methods Enzymol. 326, 245–54.

Boze, H., Céline, L., Patrick, C., Fabien, R., Christine, V., Yves, C., Guy, M., 2001. High-level secretory production of recombinant porcine follicle-stimulating hormone by *Pichia pastoris*. Process Biochem. 36, 907–913.

Çalık, P., Ata, Ö., Güneş, H., Massahi, A., Boy, E., Keskin, A., Öztürk, S., Zerze, G.H., Özdamar, T.H., 2015. Recombinant protein production in *Pichia pastoris* under glyceraldehyde-3-phosphate dehydrogenase promoter: From carbon source metabolism to bioreactor operation parameters. Biochem. Eng. J. 95, 20–36.

Callewaert, N., Laroy, W., Cadirgi, H., Geysens, S., Saelens, X., Min Jou, W., Contreras, R., 2001. Use of HDEL-tagged Trichoderma reesei mannosyl oligosaccharide 1,2-alpha-D-mannosidase for N-glycan engineering in *Pichia pastoris*. FEBS Lett. 503, 173–8.

Carpenter, J.F., Chang, B.S., Garzon-Rodriguez, W., Randolph, T.W., 2002. Rational design of stable lyophilized protein formulations: theory and practice. Pharm. Biotechnol. 13, 109–33.

Cereghino, G.P.L., Cereghino, J.L., Ilgen, C., Cregg, J.M., 2002. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Protein Eng. 329–332.

Cereghino, J.L., Cregg, J.M., 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol. Rev. 24, 45–66.

Chakma, J., Masum, H., Perampaladas, K., Heys, J., Singer, P. a, 2011. Indian vaccine innovation: the case of Shantha Biotechnics. Global. Health 7, 9.

Chan, J.M.E.M.I. of T., 2005. A central composite design to investigate antibody fragment production by *Pichia pastoris*. Thesis (M. Eng.) - Massachusetts Institute

of Technology, Biological Engineering Division, 2005. [http://hdl.handle.net/1721.1/33872]. Accessed 1 December 2015.

Charoenrat, T., Ketudat-Cairns, M., Stendahl-Andersen, H., Jahic, M., Enfors, S.-O., 2005. Oxygen-limited fed-batch process: an alternative control for *Pichia pastoris* recombinant protein processes. Bioprocess Biosyst. Eng. 27, 399–406.

Chen, Y., Jin, M., Goodrich, L., Smith, G., Coppola, G., Calhoun, D.H., 2000b. Purification and characterization of human alpha-galactosidase A expressed in insect cells using a baculovirus vector. Protein Expr. Purif. 20, 228–36.

Chen, Y., Jin, M., Egborge, T., Coppola, G., Andre, J., Calhoun, D.H., 2000a. Expression and characterization of glycosylated and catalytically active recombinant human alpha-galactosidase A produced in *Pichia pastoris*. Protein Expr. Purif. 20, 472–84.

Chiba, Y., 2009. New Era of Glycoscience: Intrinsic and Extrinsic Functions Performed by Glycans Glycan Engineering and Production of "Humanized" Glycoprotein in Yeast Cells 32, 786–795.

Chiba, Y., Sakuraba, H., Kotani, M., Kase, R., Kobayashi, K., Takeuchi, M., Ogasawara, S., Maruyama, Y., Nakajima, T., Takaoka, Y., Jigami, Y., 2002. Production in yeast of alpha-galactosidase A, a lysosomal enzyme applicable to enzyme replacement therapy for Fabry disease. Glycobiology 12, 821–8.

Choi, B.-K., Bobrowicz, P., Davidson, R.C., Hamilton, S.R., Kung, D.H., Li, H., Miele, R.G., Nett, J.H., Wildt, S., Gerngross, T.U., 2003. Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. Proc. Natl. Acad. Sci. U. S. A. 100, 5022–7.

Ciofalo, V., Barton, N., Kreps, J., Coats, I., Shanahan, D., 2006. Safety evaluation of a lipase enzyme preparation, expressed in *Pichia pastoris*, intended for use in the degumming of edible vegetable oil. Regul. Toxicol. Pharmacol. 45, 1–8.

Corchero, L., Mendoza, R., Lorenzo, J., Domı, C., Rodrı, V., Ferrer-miralles, N., Villaverde, A., 2011. Integrated Approach to Produce a Recombinant, His-Tagged Human a-Galactosidase A in Mammalian Cells.

Costa, A.R., Rodrigues, M.E., Henriques, M., Oliveira, R., Azeredo, J., 2014. Glycosylation: impact, control and improvement during therapeutic protein production. Crit. Rev. Biotechnol.

Courtois, F., Schneider, C.P., Agrawal, N.J., Trout, B.L., 2015. Rational Design of Biobetters with Enhanced Stability. J. Pharm. Sci. 104, 2433–40.

Coutinho, M.F., Alves, S., 2015. From rare to common and back again: 60years of lysosomal dysfunction. Mol. Genet. Metab. In Press, Corrected Proof.

Coutinho, M.F., Prata, M.J., Alves, S., 2011. Mannose-6-phosphate pathway: A review on its role in lysosomal function and dysfunction. Mol. Genet. Metab.

Cox, T.M., Aerts, J.M.F.G., Andria, G., Beck, M., Belmatoug, N., Bembi, B., Chertkoff, R., Vom Dahl, S., Elstein, D., Erikson, A., Giralt, M., Heitner, R., Hollak, C., Hrebicek, M., Lewis, S., Mehta, A., Pastores, G.M., Rolfs, A., Miranda, M.C.S., Zimran, A., 2003. The role of the iminosugar N-butyldeoxynojirimycin (miglustat) in the management of type I (non-neuronopathic) Gaucher disease: a position statement. J. Inherit. Metab. Dis. 26, 513–26.

Cregg, J.M., Barringer, K.J., Hessler, A.Y., Madden, K.R., 1985. *Pichia pastoris* as a host system for transformations. Mol. Cell. Biol. 5, 3376–85.

Cregg, J.M., Madden, K.R., Barringer, K.J., Thill, G.P., Stillman, C.A., 1989. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Mol. Cell. Biol. 9, 1316–23.

Cregg, J.M., Tolstorukov, I., Kusari, A., 2009. Expression in the Yeast *Pichia pastoris*, 1st ed, Methods in Enzymology. Elsevier Inc.

Cregg, J.M., Tolstorukov, I., Kusari, A., Jay, A., Madden, K., Chappell, T., 2010. Expression of Recombinant Genes in the Yeast *Pichia pastoris* 1–14.

Dahms, N.M., Olson, L.J., Kim, J.-J.P., 2008. Strategies for carbohydrate recognition by the mannose 6-phosphate receptors. Glycobiology 18, 664–78.

Darby, R.A.J., Cartwright, S.P., Dilworth, M. V, Bill, R.M., 2012. Which yeast species shall I choose? *Saccharomyces cerevisiae* versus *Pichia pastoris* (review). Methods Mol. Biol. 866, 11–23.

De Schutter, K., Lin, Y.-C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouzé, P., Van de Peer, Y., Callewaert, N., 2009. Genome sequence of the recombinant protein production host *Pichia pastoris*. Nat. Biotechnol. 27, 561–6.

Deegan, P.B., Baehner, A.F., Barba Romero, M.-A., Hughes, D.A., Kampmann, C., Beck, M., 2006. Natural history of Fabry disease in females in the Fabry Outcome Survey. J. Med. Genet. 43, 347–52.

Demain, A.L., Vaishnav, P., 2009. Production of recombinant proteins by microbes and higher organisms. Biotechnol. Adv. 27, 297–306.

Desnick, R.J., Schuchman, E.H., 2002. Enzyme replacement and enhancement therapies: lessons from lysosomal disorders. Nat Rev Genet 3, 954–966.

Dodick, D.W., Goadsby, P.J., Silberstein, S.D., Lipton, R.B., Olesen, J., Ashina, M., Wilks, K., Kudrow, D., Kroll, R., Kohrman, B., Bargar, R., Hirman, J., Smith, J., 2014. Safety and efficacy of ALD403, an antibody to calcitonin gene-related peptide, for the prevention of frequent episodic migraine: a randomised, double-blind, placebo-controlled, exploratory phase 2 trial. Lancet. Neurol. 13, 1100–7.

Falcone, S., Cocucci, E., Podini, P., Kirchhausen, T., Clementi, E., Meldolesi, J., 2006. Macropinocytosis: regulated coordination of endocytic and exocytic membrane traffic events. J. Cell Sci. 119, 4758–69.

Felber, M., Pichler, H., Ruth, C., 2014. Strains and molecular tools for recombinant protein production in *Pichia pastoris*. Methods Mol. Biol. 1152, 87–111.

Fickers, P., 2014. *Pichia pastoris*: a workhorse for recombinant protein production. Curr. Res. Microbiol. Biotechnol. 2, 354–363.

Filocamo, M., Morrone, A., 2011. Lysosomal storage disorders: molecular basis and laboratory testing. Hum. Genomics 5, 156–169.

Fuller M, Meikle PJ, H.J., n.d. Epidemiology of lysosomal storage diseases: an overview. In: Mehta A, Beck M, Sunder-Plassmann G, editors. Fabry Disease: Perspectives from 5 Years of FOS. Oxford Oxford PharmaGenesis; 2006. Chapter 2.

Garcia-Ortega, X., Ferrer, P., Montesinos, J.L., Valero, F., 2013. Fed-batch operational strategies for recombinant Fab production with *Pichia pastoris* using the constitutive GAP promoter. Biochem. Eng. J. 79, 172–181.

Garman, S.C., Garboczi, D.N., 2004. The molecular defect leading to Fabry disease: structure of human alpha-galactosidase. J. Mol. Biol. 337, 319–35.

Gasser, B., Prielhofer, R., Marx, H., Maurer, M., Nocon, J., Steiger, M., Puxbaum, V., Sauer, M., Mattanovich, D., 2013. *Pichia pastoris*: protein production host and model organism for biomedical research. Future Microbiol. 8, 191–208.

General Electric Life Sciences. Strategies for Protein Purification Handbook (2010). [http://www.gelifesciences.com/webapp/wcs/stores/servlet/catalog/en/GELifeSciences-es/service-and-support/handbooks/] Accessed 7 December 2015.

Germain, D.P., 2010. Fabry disease. Orphanet J. Rare Dis. 5, 30.

Grubb, J.H., Vogler, C., Sly, W.S., 2010. New strategies for enzyme replacement therapy for lysosomal storage diseases. Rejuvenation Res. 13, 229–36.

Guce, A.I., Clark, N.E., Salgado, E.N., Ivanen, D.R., Kulminskaya, A. a, Brumer, H., Garman, S.C., 2010. Catalytic mechanism of human alpha-galactosidase. J. Biol. Chem. 285, 3625–32.

Guerfal, M., Ryckaert, S., Jacobs, P.P., Ameloot, P., Van Craenenbroeck, K., Derycke, R., Callewaert, N., 2010. The HAC1 gene from *Pichia pastoris*: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. Microb. Cell Fact. 9, 49.

Hadjivasiliou, A. 2014. Evaluate Pharma. Orphan Drug Report 2014. [http://info.evaluategroup.com/rs/evaluatepharmaltd/images/2014OD.pdf] Accessed 20 November 2015.

Hailong Liu, Yufeng Qin, Yuankai Huang, Yaosheng Chen, Peiqing Cong, and Z.H., 2014. Direct Evaluation of the Effect of Gene Dosage on Secretion of Protein from Yeast *Pichia pastoris* by Expressing EGFP. J.Microbiol. Biotechnol. 24, 144–151.

Hall, A.K., Carlson, M.R., 2014. The current status of orphan drug development in Europe and the US. Intractable rare Dis. Res. 3, 1–7.

Hamilton, S.R., Cook, W.J., Gomathinayagam, S., Burnina, I., Bukowski, J., Hopkins, D., Schwartz, S., Du, M., Sharkey, N.J., Bobrowicz, P., Wildt, S., Li, H., Stadheim, T.A., Nett, J.H., 2013. Production of sialylated O-linked glycans in *Pichia pastoris*. Glycobiology 23, 1192–203.

Hamilton, S.R., Davidson, R.C., Sethuraman, N., Nett, J.H., Jiang, Y., Rios, S., Bobrowicz, P., Stadheim, T.A., Li, H., Choi, B.-K., Hopkins, D., Wischnewski, H., Roser, J., Mitchell, T., Strawbridge, R.R., Hoopes, J., Wildt, S., Gerngross, T.U., 2006. Humanization of yeast to produce complex terminally sialylated glycoproteins. Science 313, 1441–3.

Hamilton, S.R., Gerngross, T.U., 2007. Glycosylation engineering in yeast: the advent of fully humanized yeast. Curr. Opin. Biotechnol. 18, 387–392.

Hantzopoulos, P.A., Calhoun, D.H., 1987. Expression of the human α-galactosidase A in *Escherichia coli* K-12. Gene 57, 159–169.

Hartner, F.S., Ruth, C., Langenegger, D., Johnson, S.N., Hyka, P., Lin-Cereghino, G.P., Lin-Cereghino, J., Kovar, K., Cregg, J.M., Glieder, A., 2008. Promoter

library designed for fine-tuned gene expression in *Pichia pastoris*. Nucleic Acids Res. 36, e76.

Higgins, D.R., Cregg, J.M., 1998. *Pichia Protocols*, Methods in Molecular Biology.

Hohenblum, H., Gasser, B., Maurer, M., Borth, N., Mattanovich, D., 2004. Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. Biotechnol. Bioeng. 85, 367–75.

Hopkins, D., Gomathinayagam, S., Hamilton, S.R., 2015. A practical approach for O-linked mannose removal: the use of recombinant lysosomal mannosidase. Appl. Microbiol. Biotechnol. 99, 3913–27.

Hu, C.-X., Huang, H., Zhang, L., Huang, Y., Shen, Z.-F., Cheng, K.-D., Du, G.-H., Zhu, P., 2009. A new screening method based on yeast-expressed human dipeptidyl peptidase IV and discovery of novel inhibitors. Biotechnol. Lett. 31, 979–84.

Huang, L., Muyldermans, S., Saerens, D., 2010. Nanobodies®: proficient tools in diagnostics. Expert Rev. Mol. Diagn. 10(6):777-85.

Inan, M., Aryasomayajula, D., Sinha, J., Meagher, M.M., 2006. Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase. Biotechnol. Bioeng. 93, 771–8.

Inan, M., Meagher, M.M., 2001. Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of *Pichia pastoris*. J. Biosci. Bioeng. 92, 585–589.

Invitrogen. (2000) Pichia fermentation process guidelines. [https://tools.thermofisher.com/content/sfs/manuals/pichiaferm_prot.pdf] Accessed 7 December 2015.

Ioannou, Y. a, Bishop, D.F., Desnick, R.J., 1992. Overexpression of human alpha-galactosidase A results in its intracellular aggregation, crystallization in lysosomes, and selective secretion. J. Cell Biol. 119, 1137–50.

Ioannou, Y.A., Zeidner, K.M., Gordon, R.E., Desnick, R.J., 2001. Fabry disease: preclinical studies demonstrate the effectiveness of alpha-galactosidase A replacement in enzyme-deficient mice. Am. J. Hum. Genet. 68, 14–25.

Ishii, S., Suzuki, Y., Fan, J.Q., 2000. Role of Ser-65 in the activity of alpha-galactosidase A: characterization of a point mutation (S65T) detected in a patient with Fabry disease. Arch. Biochem. Biophys. 377, 228–33.

Jacobs, P.P., Inan, M., Festjens, N., Haustraete, J., Van Hecke, A., Contreras, R., Meagher, M.M., Callewaert, N., 2010. Fed-batch fermentation of GM-CSF-producing glycoengineered *Pichia pastoris* under controlled specific growth rate. Microb. Cell Fact. 9, 93.

Jahic, M., Veide, A., Charoenrat, T., Teeri, T., Enfors, S.-O., 2006. Process technology for production and recovery of heterologous proteins with *Pichia pastoris*. Biotechnol. Prog. 22, 1465–73.

Jahic, M., Wallberg, F., Bollok, M., Garcia, P., Enfors, S.O., 2003. Temperature limited fed-batch technique for control of proteolysis in *Pichia pastoris* bioreactor cultures. Microb. Cell Fact. 18, 2(1):6.

Jahnke, B.U. 2014. Lysosomal Storage Disease: an Example of a Lucrative Ultra-Orphan Market. Thomson Reuters Report. [http://lsconnect.thomsonreuters.com/lysosomal-storage-disease-example-lucrative-ultra-orphan-market/] Accessed 7 November 2015.

Jamshad, M., Darby, R.A.J., 2012. Yeast Transformation to Generate High-Yielding Clones (Recombinant Protein Production in Yeast). Methods in Molecular Biology 866, 57–63.

Jiang, B., Argyros, R., Bukowski, J., Nelson, S., Sharkey, N., Kim, S., Copeland, V., Davidson, R.C., Chen, R., Zhuang, J., Sethuraman, N., Stadheim, T.A., 2015. Inactivation of a GAL4-like transcription factor improves cell fitness and product yield in glycoengineered *Pichia pastoris strains*. Appl. Environ. Microbiol. 81, 260–71.

Julien, C., 2006. Production of humanlike recombinant proteins in *Pichia pastoris*. Production. Bioprocess International. January 2006, 22-31.

Jungo, C., Schenk, J., Pasquier, M., Marison, I.W., von Stockar, U., 2007. A quantitative analysis of the benefits of mixed feeds of sorbitol and methanol for the production of recombinant avidin with *Pichia pastoris*. J. Biotechnol. 131, 57–66.

Kang, T.S., Stevens, R.C., 2009. Structural aspects of therapeutic enzymes to treat metabolic disorders. Hum. Mutat. 30, 1591–610.

Karaoglan, M., Yildiz, H., Inan, M., 2014. Screening of signal sequences for extracellular production of Aspergillus niger xylanase in *Pichia pastoris*. Biochem. Eng. J. 92, 16–21.

Kawahara, T., Yanagi, H., Yura, T., Mori, K., 1997. Endoplasmic Reticulum Stress-induced mRNA Splicing Permits Synthesis of Transcription Factor

Hac1p/Ern4p That Activates the Unfolded Protein Response. Mol. Biol. Cell 8, 1845–1862.

Kiraga, J., Mackiewicz, P., Mackiewicz, D., Kowalczuk, M., Biecek, P., Polak, N., Smolarczyk, K., Dudek, M.R., Cebrat, S., 2007. The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. BMC Genomics 8, 163.

Kizhner, T., Azulay, Y., Hainrichson, M., Tekoah, Y., Arvatz, G., Shulman, A., Ruderfer, I., Aviezer, D., Shaaltiel, Y., 2015. Characterization of a chemically modified plant cell culture expressed human α-Galactosidase-A enzyme for treatment of Fabry disease. Mol. Genet. Metab. 114, 259–67.

Kornfeld, S., 1990. Lysosomal enzyme targeting. Biochem. Soc. Trans. 18, 367–374.

Krainer, F.W., Dietzsch, C., Hajek, T., Herwig, C., Spadiut, O., Glieder, A., 2012. Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway. Microb. Cell Fact. 11, 22.

Küberl, A., Schneider, J., Thallinger, G.G., Anderl, I., Wibberg, D., Hajek, T., Jaenicke, S., Brinkrolf, K., Goesmann, A., Szczepanowski, R., Pühler, A., Schwab, H., Glieder, A., Pichler, H., 2011. High-quality genome sequence of *Pichia pastoris* CBS7435. J. Biotechnol. 154, 312–20.

Kurtzman, C.P., 2009. Biotechnological strains of *Komagataella* (*Pichia*) *pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. J. Ind. Microbiol. Biotechnol. 36, 1435–8.

Kurtzman, C.P., 2005. Description of *Komagataella phaffii* sp. nov. and the transfer of *Pichia pseudopastoris* to the methylotrophic yeast genus *Komagataella*. Int. J. Syst. Evol. Microbiol. 55, 973–6.

Kuzmanov, U., Jiang, N., Smith, C.R., Soosaipillai, A., Diamandis, E.P., 2009. Differential N-glycosylation of kallikrein 6 derived from ovarian cancer cells or the central nervous system. Mol. Cell. Proteomics 8, 791–8.

Laemmli, U.K., 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature 227, 680–685.

LeBowitz, J.H., Grubb, J.H., Maga, J.A., Schmiel, D.H., Vogler, C., Sly, W.S., 2004. Glycosylation-independent targeting enhances enzyme delivery to lysosomes

and decreases storage in mucopolysaccharidosis type VII mice. Proc. Natl. Acad. Sci. U. S. A. 101, 3083–8.

Lee, C.C., Williams, T.G., Wong, D.W.S., Robertson, G.H., 2005. An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*. Plasmid 54, 80–85.

Lee, K., Jin, X., Zhang, K., Copertino, L., Andrews, L., Baker-Malcolm, J., Geagan, L., Qiu, H., Seiger, K., Barngrover, D., McPherson, J.M., Edmunds, T., 2003. A biochemical and pharmacological comparison of enzyme replacement therapies for the glycolipid storage disorder Fabry disease. Glycobiology 13, 305–13.

Li, H., Sethuraman, N., Stadheim, T.A., Zha, D., Prinz, B., Ballew, N., Bobrowicz, P., Choi, B.-K., Cook, W.J., Cukan, M., Houston-Cummings, N.R., Davidson, R., Gong, B., Hamilton, S.R., Hoopes, J.P., Jiang, Y., Kim, N., Mansfield, R., Nett, J.H., Rios, S., Strawbridge, R., Wildt, S., Gerngross, T.U., 2006. Optimization of humanized IgGs in glycoengineered *Pichia pastoris*. Nat. Biotechnol. 24, 210–215.

Liachko, I., Dunham, M.J., 2014. An autonomously replicating sequence for use in a wide range of budding yeasts. FEMS Yeast Res. 14, 364–7.

Liachko, I., Youngblood, R.A., Tsui, K., Bubb, K.L., Queitsch, C., Raghuraman, M.K., Nislow, C., Brewer, B.J., Dunham, M.J., 2014. GC-rich DNA elements enable replication origin activity in the methylotrophic yeast *Pichia pastoris*. PLoS Genet. 10(3):e1004169.

Lin-Cereghino, G.P., Stark, C.M., Kim, D., Chang, J., Shaheen, N., Poerwanto, H., Agari, K., Moua, P., Low, L.K., Tran, N., Huang, A.D., Nattestad, M., Oshiro, K.T., Chang, J.W., Chavan, A., Tsai, J.W., Lin-Cereghino, J., 2013. The effect of α-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. Gene 519, 311–7.

Lin-Cereghino, J., Lin-Cereghino, G.P., 2007. Vectors and strains for expression. Methods Mol. Biol. 389, 11–26.

Liu, B., Gong, X., Chang, S., Yang, Y., Song, M., Duan, D., Wang, L., Ma, Q., Wu, J., 2009. Disruption of the OCH1 and MNN1 genes decrease N-glycosylation on glycoprotein expressed in *Kluyveromyces lactis*. J. Biotechnol. 143, 95–102.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., Darnell, J., 2012. Molecular Cell Biology. 7th edition. New York: W. H. Freeman. ISBN-10: 1-4292-3413-X

Looser, V., Bruhlmann, B., Bumbak, F., Stenger, C., Costa, M., Camattari, A., Fotiadis, D., Kovar, K., 2015. Cultivation strategies to enhance productivity of *Pichia pastoris*: A review. Biotechnol. Adv. 33, 1177–93.

Macauley-Patrick, S., Fazenda, M.L., McNeil, B., Harvey, L.M., 2005. Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22, 249–70.

Maga, J.A., Zhou, J., Kambampati, R., Peng, S., Wang, X., Bohnsack, R.N., Thomm, A., Golata, S., Tom, P., Dahms, N.M., Byrne, B.J., LeBowitz, J.H., 2013. Glycosylation-independent Lysosomal Targeting of Acid -Glucosidase Enhances Muscle Glycogen Clearance in Pompe Mice. J. Biol. Chem. 288, 1428–1438.

Marshall, J., Ashe, K.M., Bangari, D., McEachern, K., Chuang, W.-L., Pacheco, J., Copeland, D.P., Desnick, R.J., Shayman, J.A., Scheule, R.K., Cheng, S.H., 2010. Substrate reduction augments the efficacy of enzyme therapy in a mouse model of Fabry disease. PLoS One 5, e15033.

Marx, H., Mecklenbräuker, A., Gasser, B., Sauer, M., Mattanovich, D., 2009. Directed gene copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA locus. FEMS Yeast Res. 9, 1260–70.

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., Porro, D., 2012. Recombinant protein production in yeasts. Methods Mol. Biol. 824, 329–58.

Mattanovich, D., Graf, A., Stadlmann, J., Dragosits, M., Redl, A., Maurer, M., Kleinheinz, M., Sauer, M., Altmann, F., Gasser, B., 2009. Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris. Microb. Cell Fact. 8, 29.

Mehta A, Hughes DA. Fabry Disease. 2002 Aug 5 [Updated 2013 Oct 17]. In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2015. [http://www.ncbi.nlm.nih.gov/books/NBK1292/] Accessed 20 October 2015.

Mellitzer, A., Ruth, C., Gustafsson, C., Welch, M., Birner-Grünberger, R., Weis, R., Purkarthofer, T., Glieder, A., 2014. Synergistic modular promoter and gene optimization to push cellulase secretion by *Pichia pastoris* beyond existing benchmarks. J. Biotechnol. 191, 187–95.

Meyer, H. P., Brass, J., Jungo, C., Klein, J., Wenger, J., Mommers, M.R.. 2008. An Emerging Star for Therapeutic and Catalytic Protein Production. BioProcess International, June 2008. [http://www.bioprocessintl.com/analytical/upstream-

development/an-emerging-star-for-therapeutic-and-catalytic-protein-production-183993/] Accessed 28 November 2015.

Mindell, J.A., 2012. Lysosomal Acidification Mechanisms. Annu. Rev. Physiol. 74, 69–86.

Miyamura, N., Araki, E., Matsuda, K., Yoshimura, R., Furukawa, N., Tsuruzoe, K., Shirotani, T., Kishikawa, H., Yamaguchi, K., Shichiri, M., 1996. A carboxy-terminal truncation of human alpha-galactosidase A in a heterozygous female with Fabry disease and modification of the enzymatic activity by the carboxy-terminal domain. Increased, reduced, or absent enzyme activity depending on number of amino acid residues delected. J. Clin. Invest. 98, 1809–17.

Morel, C.F., Clarke, J.T.R., 2009. The use of agalsidase alfa enzyme replacement therapy in the treatment of Fabry disease. Expert Opin. Biol. Ther. 9, 631–9.

Nett, J.H., Stadheim, T.A., Li, H., Bobrowicz, P., Hamilton, S.R., Davidson, R.C., Choi, B., Mitchell, T., Bobrowicz, B., Rittenhour, A., Wildt, S., Gerngross, T.U., 2011. A combinatorial genetic library approach to target heterologous glycosylation enzymes to the endoplasmic reticulum or the Golgi apparatus of *Pichia pastoris*. Yeast. 2011 Mar; 28(3):237-52.

Nordén, K., Agemark, M., Danielson, J.Å.H., Alexandersson, E., Kjellbom, P., Johanson, U., 2011. Increasing gene dosage greatly enhances recombinant expression of aquaporins in *Pichia pastoris*. BMC Biotechnol. 11, 47.

Oh, D.B., 2015. Glyco-engineering strategies for development of therapeutic enzymes with improved efficacy for the treatment of lysosomal storage diseases. BMB Rep. 48, 438–444.

Owens, D.R., Landgraf, W., Schmidt, A., Bretzel, R.G., Kuhlmann, M.K., 2012. The emergence of biosimilar insulin preparations--a cause for concern? Diabetes Technol. Ther. 14, 989–96.

Parenti, G., Andria, G., Ballabio, A., 2015. Lysosomal Storage Diseases: From Pathophysiology to Therapy. Annu. Rev. Med. 66, 471–486.

Park, H.-J., Zhang, Y., Georgescu, S.P., Johnson, K.L., Kong, D., Galper, J.B., 2006. Human umbilical vein endothelial cells and human dermal microvascular endothelial cells offer new insights into the relationship between lipid metabolism and angiogenesis. Stem Cell Rev. 2, 93–102.

Pedro, A.Q., Maia, C., Sousa, F., Queiroz, J.A., Passarinha, L.A., 2013. *Pichia pastoris*: A Recombinant Microfactory for Antibodies and Human Membrane Proteins 23, 587–601.

Petros, R.A., DeSimone, J.M., 2010. Strategies in the design of nanoparticles for therapeutic applications. Nat. Rev. Drug Discov. 9, 615–27.

Pisani, A., Daniele, A., Di Domenico, C., Nigro, E., Salvatore, F., Riccio, E., 2015. Late diagnosis of Fabry disease caused by a de novo mutation in a patient with end stage renal disease. BMC Res. Notes 8, 711.

Pla, I.A., Damasceno, L.M., Vannelli, T., Ritter, G., Batt, C.A., Shuler, M.L.,. Evaluation of Mut^+ and Mut^S *Pichia pastoris* phenotypes for high level extracellular scFv expression under feedback control of the methanol concentration. Biotechnol. Prog. 22, 881–8.

Potgieter, T.I., Cukan, M., Drummond, J.E., Houston-Cummings, N.R., Jiang, Y., Li, F., Lynaugh, H., Mallem, M., McKelvey, T.W., Mitchell, T., Nylen, A., Rittenhour, A., Stadheim, T. a, Zha, D., d'Anjou, M., 2009. Production of monoclonal antibodies by glycoengineered *Pichia pastoris*. J. Biotechnol. 139, 318–25.

Puxbaum, V., Mattanovich, D., Gasser, B., 2015. Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. Appl. Microbiol. Biotechnol. 99, 2925–38.

Qin, X., Qian, J., Yao, G., Zhuang, Y., Zhang, S., Chu, J., 2011. GAP Promoter Library for Fine-Tuning of Gene Expression in *Pichia pastoris*. Appl. Environ. Microbiol. 77, 3600–3608.

Reski, R., Parsons, J., Decker, E.L., 2015. Moss-made pharmaceuticals: from bench to bedside. Plant Biotechnol. J. 13, 1191–8.

Rich, J.R., Withers, S.G., 2009. Emerging methods for the production of homogeneous human glycoproteins. Nat. Chem. Biol. 5, 206–15.

Saftig P. Physiology of the lysosome. In: Mehta A, Beck M, Sunder-Plassmann G, editors. Fabry Disease: Perspectives from 5 Years of FOS. Oxford: Oxford PharmaGenesis; 2006. Chapter 3.
[http://www.ncbi.nlm.nih.gov/books/NBK11604/] Accessed 18 November 2015.

Sakuraba, H., Chiba, Y., Kotani, M., Kawashima, I., Ohsawa, M., Tajima, Y., Takaoka, Y., Jigami, Y., Takahashi, H., Hirai, Y., Shimada, T., Hashimoto, Y., Ishii, K., Kobayashi, T., Watabe, K., Fukushige, T., Kanzaki, T., 2006. Corrective

effect on Fabry mice of yeast recombinant human alpha-galactosidase with N-linked sugar chains suitable for lysosomal delivery. J. Hum. Genet. 51, 341–52.

Schiffmann, R., 2009. Fabry disease. Pharmacol. Ther. 122, 65–77.

Shemesh, E., Deroma, L., Bembi, B., Deegan, P., Hollak, C., Weinreb, N.J., Cox, T.M., 2015. Enzyme replacement and substrate reduction therapy for Gaucher disease. Cochrane database Syst. Rev. 3, CD010324.

Shirley, M., 2015. Sebelipase Alfa: First Global Approval. Drugs 75, 1935–40.

Solá, R.J., Griebenow, K., 2010. Glycosylation of therapeutic proteins: an effective strategy to optimize efficacy. BioDrugs 24, 9–21.

Sreenivas, S., Krishnaiah, S.M., Govindappa, N., Basavaraju, Y., Kanojia, K., Mallikarjun, N., Natarajan, J., Chatterjee, A., Sastry, K.N., 2015. Enhancement in production of recombinant two-chain Insulin Glargine by over-expression of Kex2 protease in *Pichia pastoris*. Appl. Microbiol. Biotechnol. 99, 327–36.

Stadlmayr, G., Mecklenbräuker, A., Rothmüller, M., Maurer, M., Sauer, M., Mattanovich, D., Gasser, B., 2010. Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. J. Biotechnol. 150, 519–29.

Sunga, A.J., Tolstorukov, I., Cregg, J.M., 2008. Posttransformational vector amplification in the yeast *Pichia pastoris*. FEMS Yeast Res. 8, 870–6.

Surribas, A., Stahn, R., Montesinos, J.L., Enfors, S.-O., Valero, F., Jahic, M., 2007. Production of a Rhizopus oryzae lipase from *Pichia pastoris* using alternative operational strategies. J. Biotechnol. 130, 291–9.

Syed, Y.Y., Dhillon, S., 2013. Ocriplasmin: a review of its use in patients with symptomatic vitreomacular adhesion. Drugs 73, 1617–25.

Tambuyzer, E., 2010. Rare diseases, orphan drugs and their regulation: questions and misconceptions. Nat. Rev. Drug Discov. 9, 921–9.

Tekoah, Y., Tzaban, S., Kizhner, T., Hainrichson, M., Gantman, A., Golembo, M., Aviezer, D., Shaaltiel, Y., 2013. Glycosylation and functionality of recombinant β-glucocerebrosidase from various production systems. Biosci. Rep. 33.

Temming, K., Schiffelers, R.M., Molema, G., Kok, R.J., 2005. RGD-based strategies for selective delivery of therapeutics and imaging agents to the tumour vasculature. Drug Resist. Updat. 8, 381–402.

Thompson, C.A., 2010. FDA approves kallikrein inhibitor to treat hereditary angioedema. Am. J. Health. Syst. Pharm. 67, 93.

Tiels, P., Baranova, E., Piens, K., De Visscher, C., Pynaert, G., Nerinckx, W., Stout, J., Fudalej, F., Hulpiau, P., Tännler, S., Geysens, S., Van Hecke, A., Valevska, A., Vervecken, W., Remaut, H., Callewaert, N., 2012. A bacterial glycosidase enables mannose-6-phosphate modification and improved cellular uptake of yeast-produced recombinant human lysosomal enzymes. Nat. Biotechnol. 30, 1225–31.

Tolner, B., Smith, L., Begent, R.H.J., Chester, K. a, 2006. Production of recombinant protein in *Pichia pastoris* by fermentation. Nat. Protoc. 1, 1006–21.

Trentmann, O., Khatri, N.K., Hoffmann, F., 2004. Reduced oxygen supply increases process stability and product yield with recombinant *Pichia pastoris*. Biotechnol. Prog. 20, 1766–75.

Tsukimura, T., Kawashima, I., Togawa, T., Kodama, T., Suzuki, T., Watanabe, T., Chiba, Y., Jigami, Y., Fukushige, T., Kanekura, T., Sakuraba, H., 2012. Efficient Uptake of Recombinant α-Galactosidase A Produced with a Gene-Manipulated Yeast by Fabry Mice Kidneys. Mol. Med. 18, 76–82.

Unzueta, U., Vázquez, F., Accardi, G., Mendoza, R., Toledo-Rubio, V., Giuliani, M., Sannino, F., Parrilli, E., Abasolo, I., Schwartz, S., Tutino, M.L., Villaverde, A., Corchero, J.L., Ferrer-Miralles, N., 2015. Strategies for the production of difficult-to-express full-length eukaryotic proteins using microbial cell factories: production of human alpha-galactosidase A. Appl. Microbiol. Biotechnol. 99, 5863–74.

Vadhana, A.K.P., Samuel, P., Berin, R.M., Krishna, J., Kamatchi, K., Meenakshisundaram, S., 2013. Improved secretion of Candida antarctica lipase B with its native signal peptide in Pichia pastoris. Enzyme Microb. Technol. 52, 177–83.

Valero, F., 2012. Lipases and Phospholipases. Methods in Molecular Biology 861, 161–178.

Valero, F. (2013). Bioprocess Engineering of Pichia pastoris, an Exciting Host Eukaryotic Cell Expression System, Protein Engineering. Technology and Application, Dr. Tomohisa Ogawa (Ed.), ISBN: 978-953-51-1138-2

van der Klei, I.J., Yurimoto, H., Sakai, Y., Veenhuis, M., 2006. The significance of peroxisomes in methanol metabolism in methylotrophic yeast. Biochim. Biophys. Acta 1763, 1453–62.

Van Roy, M., Ververken, C., Beirnaert, E., Hoefman, S., Kolkman, J., Vierboom, M., Breedveld, E., Hart, B., Poelmans, S., Bontinck, L., Hemeryck, A., Jacobs, S., Baumeister, J., Ulrichts, H., 2015. The preclinical pharmacology of the high affinity anti-IL-6R Nanobody® ALX-0061 supports its clinical development in rheumatoid arthritis. Arthritis Res. Ther. 17, 135.

Vanz, A.L., Lünsdorf, H., Adnan, A., Nimtz, M., Gurramkonda, C., Khanna, N., Rinas, U., 2012. Physiological response of *Pichia pastoris* GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes. Microb. Cell Fact. 11, 103.

Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. [http://www.ncbi.nlm.nih.gov/books/NBK1908/] Accessed 13 October 2015.

Vassileva, A., Chugh, D.A., Swaminathan, S., Khanna, N., 2001. Effect of copy number on the expression levels of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris*. Protein Expr. Purif. 21, 71–80.

Vedder, A.C., Breunig, F., Donker-Koopman, W.E., Mills, K., Young, E., Winchester, B., Ten Berge, I.J.M., Groener, J.E.M., Aerts, J.M.F.G., Wanner, C., Hollak, C.E.M., 2008. Treatment of Fabry disease with different dosing regimens of agalsidase: effects on antibody formation and GL-3. Mol. Genet. Metab. 94, 319–25.

Vervecken, W., Kaigorodov, V., Callewaert, N., Geysens, S., De Vusser, K., Contreras, R., 2004. In vivo synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. Appl. Environ. Microbiol. 70, 2639–46.

Vogl, T., Hartner, F.S., Glieder, A., 2013. New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. Curr. Opin. Biotechnol. 24, 1094–101.

Walsh, G., 2014. Biopharmaceutical benchmarks 2014. Nat. Biotechnol. 32, 992–1000.

Walsh, G., Jefferis, R., 2006. Post-translational modifications in the context of therapeutic proteins. Nat. Biotechnol. 24, 1241–52.

Wang, J., Nguyen, V., Glen, J., Henderson, B., Saul, A., Miller, L.H., 2005. Improved yield of recombinant merozoite Surface protein 3 (MSP3) from *Pichia pastoris* using chemically defined media. Biotechnol. Bioeng. 90, 838–47.

Warnock, D.G., Bichet, D.G., Holida, M., Goker-Alpan, O., Nicholls, K., Thomas, M., Eyskens, F., Shankar, S., Adera, M., Sitaraman, S., Khanna, R., Flanagan, J.J., Wustman, B.A., Barth, J., Barlow, C., Valenzano, K.J., Lockhart, D.J., Boudes, P., Johnson, F.K., 2015. Oral Migalastat HCl Leads to Greater Systemic Exposure and Tissue Levels of Active α-Galactosidase A in Fabry Patients when Co-Administered with Infused Agalsidase. PLoS One 10, e0134341.

Werber, Y., 2004. Lysosomal storage diseases market. Nat. Rev. Drug Discov. 3, 9–10.

Wildt, S., Gerngross, T.U., 2005. The humanization of N-glycosylation pathways in yeast. Nat. Rev. Microbiol. 3, 119–28.

Xiong, A.-S., Yao, Q.-H., Peng, R.-H., Zhang, Z., Xu, F., Liu, J.-G., Han, P.-L., Chen, J.-M., 2006. High level expression of a synthetic gene encoding Peniophora lycii phytase in methylotrophic yeast *Pichia pastoris*. Appl. Microbiol. Biotechnol. 72, 1039–47.

Yang, S., Kuang, Y., Li, H., Liu, Y., Hui, X., Li, P., Jiang, Z., Zhou, Y., Wang, Y., Xu, A., Li, S., Liu, P., Wu, D., 2013. Enhanced production of recombinant secretory proteins in *Pichia pastoris* by optimizing Kex2 P1' site. PLoS One 8, e75347.

Yasuda, K., Chang, H.-H., Wu, H.-L., Ishii, S., Fan, J.-Q., 2004. Efficient and rapid purification of recombinant human alpha-galactosidase A by affinity column chromatography. Protein Expr. Purif. 37, 499–506.

Zarate, Y., Hopkin, R., 2008. Fabry's disease. Lancet 372, 1427–1435.

Zhang, Z., Lutz, B., 2002. Cre recombinase-mediated inversion using lox66 and lox71: method to introduce conditional point mutations into the CREB-binding protein. Nucleic Acids Res. 30, e90.

Zhu, T., Guo, M., Tang, Z., Zhang, M., Zhuang, Y., Chu, J., Zhang, S., 2009. Efficient generation of multi-copy strains for optimizing secretory expression of porcine insulin precursor in yeast *Pichia pastoris*. J. Appl. Microbiol. 107, 954–63.