Universidade do Minho
Escola de Engenharia

Marlene Isabel Cerqueira Montenegro Baptista

GLYCANZYME - CHARACTERISATION OF NOVEL ENZYMES FOR PROTEIN GLYCOENGINEERING

October 2019

GLYCANZYME - CHARACTERISATION OF NOVEL ENZYMES FOR PROTEIN GLYCOENGINEERING
Marlene Isabel Cerqueira Montenegro Baptista
UMinho|2019

Universidade do Minho
Escola de Engenharia

Marlene Isabel Cerqueira Montenegro Baptista

GLYCANZYME - CHARACTERISATION OF NOVEL ENZYMES FOR PROTEIN GLYCOENGINEERING

Master Thesis
Master Program in Biotechnology

Work supervised by:
Doutora Tatiana Aguiar
Doutora Carla Oliveira

October 2019

ACKNOWLEDGMENTS

I spent the last two years studying Biotechnology and I strongly believe, from the knowledge I acquired, that Biotechnology will revolutionize our entire world.

First and foremost, I want to thank my supervisors. Over these last nine months, Dr Tatiana Aguiar provided me with the most appropriate and thoughtful guidance in every step of my research project. I had the privilege to work with a fantastic person whose valuable knowledge contributed significantly to improve my skills and guided me to look at molecular biotechnology in a more comprehensive way. I esteem her curiosity and commitment to Science and I would like to thank her for all the commitment she also dedicated to me. I also would like to thank her for introducing me the wonderful world of glycosylation. I must add a special thanks to Dr Carla Oliveira, my co-supervisor, for all the guidance, all the knowledge she transmitted me about recombinant protein production and for teaching me that organization is a very important skill.

I am very grateful to Professor Lucília Domingues for receiving me with open arms at the Molecular Biotechnology Laboratory and for introducing me to the molecular biotechnology world. I am very thankful for the great advices when they were necessary. I must add a special word of thanks to Professor Peter Philippsen. I cannot leave unmentioned the suggestions he gave me in my work. My special thanks to everyone at the Molecular Biotechnology Laboratory for such a pleasant working atmosphere and help. I am particularly grateful to Ana, Sara Baptista, Joana, Carlos, Pedro and Rui.

I also want to thank my friends Bruna Basto, Rafael Marques, Eduarda Gomes, Mário Azevedo, Gonçalo Carvalho, Marta Teixeira and Tiago Gião for the love and patience everyday throughout these 5 years in the university. For all the good moments, I am deeply grateful. I must thank my Father, my Mother, my Brother and my Husband for their unconditional love and support. I would just like to thank them for always valuing me, for all they have sacrificed for me and for giving me the opportunity to pursue my dreams. I must ultimately thank my parents for teaching me that when you dedicate all your efforts to something, one day someone will value you for what you have done.

Lastly, I would like to thank the support given by FCT and by the European Regional Development Fund under the scope of Compete2020 and Norte2020: strategic funding of UID/BIO/04469/2019, Project EcoBioInks4SmartTextiles (PTDC/CTM-TEX/30298/2017–POCI-01-0145-FEDER-030298) and BioTecNorte operation (NORTE-01-0145-FEDER-000004).

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

***GLYCANZYME* - Characterisation of novel enzymes for protein glycoengineering**

ABSTRACT

One of the most common protein post-translational modifications occurring in eukaryotes is glycosylation, which is critical for several biological processes. There are different enzymes involved in the processing of *N*-glycans in eukaryotes, among them endo-β-*N*-acetylglucosaminidases (ENGases), whose physiological roles are still poorly understood. ENGases of the glycoside hydrolase (GH) family 85 are a class of enzymes (EC 3.2.1.96) that, in addition to hydrolytic activity against the diacetylchitobiose core of *N*-glycans, can also display transglycosylation activity. Given their usefulness for the analysis and glycoengineering of glycopeptides and glycoproteins, these enzymes have become a focus of biotechnological interest.

Envisioning the identification of novel GH85 ENGases with useful action, this thesis focused on the characterization of two new putative ENGases of this family, one from the filamentous fungus *Ashbya gossypii* (AgENGase) and another from the yeast *Zygosaccharomyces rouxii* (ZrENGase). Previous experimental results hinted at the existence of ENGase activity in these organisms and in their genome one homologue gene exists that should code for a putative GH85 ENGase (*AFR597W* in *A. gossypii* and *ZYRO0B07216g* in *Z. rouxii*). Bioinformatic approaches were initially used to assess the potential activity of these putative ENGases and important residues for the hydrolytic and transglycosylation activity were found to be conserved in AgENGase. Only the putative 3D model of ZrENGase presented the typical (β_8/α_8)-TIM-barrel structure of ENGases. To characterise these proteins, their coding genes were cloned into an *Escherichia coli* expression plasmid and their recombinant production was optimized by testing different *E. coli* expression strains, production media and induction conditions. However, the recombinant proteins produced remained mainly in the insoluble fraction after cell lysis and further improvements in their production and purification will be required to generate enough protein for activity studies and eventual application. Nevertheless, different protocols were already optimized in the scope of this thesis for the future evaluation of the deglycosylation activity of these recombinant ENGases.

Considering the bioinformatic analysis and the results from enzymatic assays performed in crude cell extracts of wild *A. gossypii*, the AgENGase action could be located in the intracellular space. Envisioning the understanding of the physiological role of the putative AgENGase, a gene with no homolog in *Saccharomyces cerevisiae* and with unknown function, the disruption of the *A. gossypii AFR597W* was also performed. Thereby, AgENGase was found to be important for sporulation, as the generated *afr597w* mutants lost the ability to sporulate. Nevertheless, further studies will be needed to fully understand the physiological and enzymatic function of this enzyme.

Palavras-chave: *Ashbya gossypii*; endo-β-*N*-acetylglucosaminidases; recombinante protein production

RESUMO

Uma das modificações pós-tradução de proteínas mais comuns que ocorrem em eucariotas é a glicosilação, que é crítica para vários processos biológicos. Existem diferentes enzimas envolvidas no processamento de *N*-glicanos em eucariotas, entre elas as endo-β-*N*-acetilglucosaminidases (ENGases), cujos papéis fisiológicos ainda são pouco compreendidos. ENGases da família glicosil hidrolase (GH) 85 são uma classe de enzimas (EC 3.2.1.96) que, além da atividade hidrolítica no núcleo de diacetil-quitobiose de *N*-glicanos, também podem exibir atividade de transglicosilação. Dada a sua utilidade para a análise e a glico-engenharia de glicopéptidos e glicoproteínas, essas enzimas apresentam-se com forte interesse biotecnológico.

Tendo em vista a identificação de novas ENGases da família GH85 de interesse, esta tese focou a caracterização de duas novas supostas ENGases dessa família, uma do fungo filamentoso *Ashbya gossypii* (AgENGase) e outra da levedura *Zygosaccharomyces rouxii* (ZrENGase). Resultados experimentais anteriores sugeriram a existência de atividade ENGase nesses organismos e, no seu genoma, encontrou-se um gene homólogo que poderá codificar para putativas ENGases da família GH85 (*AFR597W* em *A. gossypii* e *ZYRO0B07216g* em *Z. rouxii*). Abordagens bioinformáticas foram usadas inicialmente para avaliar a atividade potencial dessas ENGases e foram identificados resíduos importantes para a atividade hidrolítica e de transglicosilação conservados na putativa AgENGase. Por outro lado, apenas o suposto modelo 3D previsto para a putativa ZrENGase apresentou a estrutura típica de um barril TIM (β_8/α_8) das ENGases. Para caracterizar essas proteínas, os seus genes foram clonados num plasmídeo de expressão de *Escherichia coli* e a sua produção recombinante foi otimizada testando diferentes estirpes de expressão de *E. coli*, meios de produção e condições de indução. No entanto, as proteínas recombinantes produzidas permaneceram principalmente na fração insolúvel após a lise celular e serão necessárias melhorias adicionais na sua produção e purificação para gerar quantidade de proteína suficiente para estudos de atividade e eventuais aplicações. No entanto, diferentes protocolos já foram otimizados no âmbito desta tese para avaliação futura da atividade de desglicosilação dessas ENGases recombinantes.

Considerando a análise bioinformática e os resultados de ensaios enzimáticos realizados em extratos celulares brutos de *A. gossypii wild*, foi possível determinar a localização da AgENGase como sendo intracelular. Visando compreender o papel fisiológico da suposta AgENGase foi também realizada a interrupção do gene *AFR597W* de *A. gossypii*. Dessa forma, a AgENGase foi considerada importante para a esporulação, pois os mutantes *afr597w* gerados perderam a capacidade de esporular. No entanto, serão necessários mais estudos para entender completamente a função fisiológica e enzimática dessa enzima.

Palavras-chave: *Ashbya gossypii*; endo-β-*N*-acetilglucosaminidases; produção de proteínas recombinantes

TABLE OF CONTENTS

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS ii

ACKNOWLEDGMENTS iii

STATEMENT OF INTEGRITY iv

ABSTRACT v

RESUMO vi

INTRODUCTION 2

1. Overview on glycosylation 2

1.1. Glycosylation and its biological significance 2

1.2. Diversity of glycans 3

1.3. *N*-glycan biosynthetic pathway 6

1.4. Synthetic methods for glycoprotein engineering 8

2. ENGases 11

2.1. Mechanism of action 13

2.2. ENGases applications 14

3. *A. gossypii* 16

4. *Z. rouxii* 17

5. Recombinant protein production in *Escherichia coli*: advantages and challenges 18

OBJECTIVES 20

MATERIALS AND METHODS 22

1. Bioinformatics 22

2. Strains and media 22

3. Molecular Biology 24

3.1. DNA constructs 24

3.1.1. Cloning and expression plasmid construction of putative AgENGase and ZrENGase-encoding genes 24

3.1.2. Disruption of *AFR597W* gene in *A. gossypii* 28

3.2. *A. gossypii* transformation and screening of transformants 28

4. Production and purification of recombinant ENGases 29

4.1. SDS- and Native-PAGE Analysis 30

5. ENGase/chitinase activity detection 31

5.1. Analysis of ENGase deglycosylation activity 31

5.2. Analysis of β-*N*-acetylglucosaminidase and chitobiosidase 32

activity in *A. gossypii*.....32
RESULTS AND DISCUSSION.....35
1. Bioinformatic characterization.....35
2. Production and purification of recombinant AgENGase and ZrENGase.....41
2.1. Expression plasmid construction.....42
2.2. Optimization of production and purification of the recombinant proteins.....43
3. Optimization of electrophoresis- and HPLC-based methodologies to analyse ENGase deglycosylation activity.....53
4. Analysis of intra and extracellular ENGase/chitinase activity in *A. gossypii*.....56
5. Disruption of the *AFR597W* gene in *A. gossypii* and characterization of the generated mutant.....58
CONCLUSION.....62
REFERENCES.....64
APPENDICES.....76

LIST OF FIGURES

Figure 1. Common classes of eukaryotic glycans. The common structures of a proteoglycan, a glycosylphosphatidylinositol (GPI)-anchored glycoprotein, a glycoprotein and a glycosphingolipid are represented. Figure adapted from 1. 2

Figure 2. Most common glycosidic bonds. The following glycosidic bonds are represented: A) (β-*N*-acetylglucosamine)-*N*-asparagine; B) (α-*N*-acetylgalactosamine)-*O*-serine/threonine; and C) (β-*N*-acetylglucosamine)-*O*-serine/threonine. The monosaccharides through which the glycans are bound are shown in gray. Figure taken from 9. 3

Figure 3. Types of *N*-glycans found in eukaryotes. *N*-glycans, linked to the Asn-X-Ser/Thr consensus sequence on eukaryotic glycoproteins, can be classified into three fundamental types: oligomannose, complex and hybrid. Each *N*-glycan contains a common core, $Man_3GlcNAc_2Asn$. Figure taken out from [14]. 5

Figure 4. Processing and maturation of an *N*-glycan. The mature Dol-P-P-GlcNAc is transferred to the Asn-X-Ser/Thr consensus sequence of a protein, during protein synthesis, as the proteins are translocated to the ER. After transferring fourteen sugars to the protein, α-glycosidases I and II, present in ER, remove three glucose residues and mannosidase removes a mannose residue. These reactions are directly linked to lectin-assisted glycoprotein folding, which determines whether the glycoprotein continues to the Golgi or is degraded. If folding is correct, the glycoproteins continue maturation in this organelle. Figure adapted from 11. 7

Figure 5. *N*-glycan processing in the Golgi of different eukaryotic taxa. Figure adapted from 1. 8

Figure 6. Scheme of transglycosylation reaction catalysed by some ENGase of GH85 family. Figure adapted from [58]. 12

Figure 7. Hydrolytic mechanism of action of ENGases of the GH85 family. The mechanism involves an acid-base catalysis reaction supported by the side chain of a residue of glutamate and asparagine. The intermediate formed is an oxazoline. Figure adapted from[65]. 13

Figure 8. Life cycle of *A. gossypii*. The main phases of the filamentous fungus *A. gossypii* life cycle are illustrated: a) and b) represent the isotropic growth phase; c), d) and e) represent the polar growth phase; and f) represents the sporulation phase. Figure adapted from [93]. 16

Figure 9. Life cycle of *Z. rouxii*. In most of the life cycle, *Z. rouxii* cells remain in a haploid state and then reproduce by budding, like most conventional yeasts, leading to the formation of a transient heterokaryotic zygote. Afterwards, the diploid zygote undergoes meiosis and sporulation and the haploid state is finally restored. Figure adapted from[108]. 18

Figure 10. pETM20 vector sequence features. 25

Figure 11. Schematic representation of the steps followed for the construction of pETM20-*ZYR00B07216g* (pETM20-ZrENGase) and pETM20-*AFR597W* (pETM20-AgENGase). PlasmaDNA software was used for *in silico* plasmid construction and representation. 26

Figure 12. Multiple alignment of several ENGases of the GH85 family, namely the putative ENGases of *A. gossypii* and *Z. rouxii,* and corresponding dendogram. Multiple alignment was performed using the Clustal Omega tool and the dendogram was built based on the multiple alignment performed and representend using the FigTree tool. Residues conserved between the homologues are shaded in black. The putative or experimentally determined residues involved in the hydrolytic (red box) and transglycosylation activity (green box) are represented. Accession numbers for the represented ENGases are described in Table 7. Data included in the poster presented at the 4th Meeting of Medicinal Biotechnology (Appendix E). 37

Figure 13. Multiple alignment of several ENGases of the GH85 family, namely the putative ENGases from *A. gossypii* and *Z. rouxii.* Multiple alignment was performed using Clustal Omega tool. Residues conserved between the homologues are shaded in black. Important residues involved in the catalysis besides the catalytic

residues that may be conserved in Endo-M, Endo-Om and the putative ZrENGase and AgENGase are represented in a red box. Accession numbers for the represented ENGases are described in Table 7. 38
Figure 14. Endo-A structure used as template for the homology-based models of AgENGase and ZrENGase. Endo-A mutant (PDB: 2VTF.1.A) structure was used as a template for the construction of the homology-based models of AgENGase and ZrENGase. This structure was visualized in The Pymol Molecular Graphics System, Version 2.1.0 Schrödinger, LLC. 39
Figure 15. Prediction of the 3D structure of the putative AgENGase and ZrENGase from homology-based models. The models (A and D) were built using the SWISS-MODEL server and were visualized using The Pymol Molecular Graphics System, Version 2.1.0 Schrödinger, LLC. A- Predicted model for the possible AgENGase 3D structure, B- Representation of the solvent accessible surface of the putative AgENGase; C- Representation of residues that may establish hydrophobic interactions around the catalytic pocket (Tyr-341 Phe-132, Tyr-134, Phe-135, Phe-207 and Phe-139), D- Predicted model for the possible 3D structure ZrENGase, E- Representation of the solvent accessible surface of the putative ZrENGase, F- Representation of residues that may establish hydrophobic interactions around the catalytic pocket (Phe-111, Tyr-317, Tyr-113, Phe-114 and Phe-183). Data included in the poster presented in the 4th Meeting of Medicinal Biotechnology (Appendix E). 39
Figure 16. Overlapping of the generated models for the putative AgENGase and ZrENGase. A- Model resulting from superposition of the predicted model for the possible AgENGase with the predicted model for the possible ZrENGase; B- Overlapping of the secondary structure of the possible AgENGase (in yellow) with that of the possible ZrENGase (in red). 40
Figure 17. Confirmation of *E. coli* NZY5α clones containing pETM20-AgENGase and pETM20-ZrENGase. A- Colony PCR products were assessed by electrophoresis on a 1 % TAE agarose gel and the image was acquired with an UV transilluminator. A1-A4 correspond to the clones tested for pETM20-AgENGase transformants (one expected amplicon with 2045 bp) and Z1-Z4 correspond to the clones tested for pETM20-ZrENGase transformants (one expected amplicon with 1931 bp). B- After plasmid extraction, digestion with restriction enzymes was performed to verify if the insert was in the correct orientation. DNA fragments were separated by electrophoresis on a 1 % TAE agarose gel and the image was acquired with an UV transilluminator. Lane 1 corresponds to pETM20-AgENGase digested with *Pst*I (two expected fragments with 6445 and 1362 bp) and lane 2 corresponds to pETM20-ZrENGase digested with *Xba*I and *Xho*I (two expected fragments with 5329 and 2364 bp). 42
Figure 18. Recombinant production of AgENGase and ZrENGase in *E. coli* Origami 2 (DE3) and SHuffle T7 Express in LB and AI media at 37 and 20 ℃. Insoluble (IF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes. Note that insoluble fraction at 37 ℃ was diluted 1:5 in order to facilitate band visualization. 44
Figure 19. Recombinant production of AgENGase and ZrENGase in *E. coli* Origami 2 (DE3) and SHuffle T7 Express in LB and TB media at 20 ℃. Insoluble (IF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes. 45
Figure 20. Recombinant production of AgENGase and ZrENGase in *E. coli* BL21 (DE3) in LB, TB and AI media at 20 ℃. Total (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes. 46
Figure 21. Recombinant production of AgENGase and ZrENGase in *E. coli* Origami 2 (DE3) in LB medium at 18 ℃. Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes. 47

Figure 22. Recombinant production of AgENGase and ZrENGase in *E. coli* Origami 2 (DE3) in LB medium at 16 ℃ with and without additives (1 M sorbitol, 0.8 M NaCl, 0.05 M sodium phosphate buffer pH 7.4). Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes. 48
Figure 23. Lysis of a pellet collected from the production of the putative ZrENGase in *E. coli* Origami 2 (DE3) at 4 ℃ in LB medium with 1 M sucrose or a combination of 1 M sucrose and 0.2 M arginine. Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusion TrxA-His-tag-TEV-ZrENGase (88 kDa) are in the red boxes. 49
Figure 24. Lysis of pellets collected from the production of the putative ZrENGase in *E. coli* Origami 2 (DE3) at 16 ℃ in LB medium and of the putative AgENGase in *E. coli* Origami 2 (DE3) at 4 ℃ in LB medium in the presence of 2 M urea. Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) in the insoluble fraction are in the red boxes and in the soluble fraction are in green boxes. 50
Figure 25. Recombinant production of ZrENGase in *E. coli* Origami 2 (DE3), BL21 (DE3) and SHuffle T7 Express in LB medium supplemented or not with 0.2 M NaCl, 0.5 % (w/v) glycerol and 0.2 M arginine at 16 ℃. Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes. 51
Figure 26. Column purification of the soluble protein fractions from productions in *E. coli* Origami 2 (DE3) at 20, 18, 16 and 4 ℃ of the putative ENGases. Fraction that did not bind to the column (FT), fraction from the last washing step (LW) and fractions were proteins were eluted (E1 and E2) were resolved in a 15 % acrylamide gel and stained with. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes. Black arrow is pointing to the band that may correspond to the putative ZrENGase. 52
Figure 27. Gel electrophoresis detection of Endo-H ENGase deglycosylation activity using different substrates. Band shifts resulting from the reaction of the referred substrates with a model ENGase, Endo-H, were detected in a 15 % acrylamide gel. 1- RNase B from bovine pancreas (negative control), 2- RNase B deglycosylated with Endo-H, 3- β-galactosidase from *A. niger* (negative control), 4- β-galactosidase deglycosylated with Endo-H, 5- invertase from *S. cerevisiae* (negative control), 6- invertase deglycosylated with Endo-H. The black arrows indicate the deglycosylation shift of each substrate, the red arrow is pointing to possible contaminants of the β-galactosidase substrate solution and the asterisk represents the band corresponding to Endo-H. 53
Figure 28. Chromatogram obtained from the analysis of the reaction of β-galactosidase with Endo-H (A) as well as the native control (B) by HPLC. Deglycosylation of glycoproteins was monitored by HPLC using a Jasco chromatograph equipped with an ELSD detector (SEDEX 85, SEDERE) and a Prevail Carbohydrate ES column (5 μm, 250 × 4.6 mm, Altech). A mixture of acetonitrile-water (60:40 % (v/v)), pumped at 0.9 mL/min, was used as mobile phase. The injection volume was defined as 20 μL. 54
Figure 29. Chromatogram obtained from the analysis of the reactions of invertase with Endo-H (A) as well as the negative control (B) by HPLC. Deglycosylation of glycoproteins was monitored by HPLC using a Jasco chromatograph equipped with an ELSD detector (SEDEX 85, SEDERE) and a Prevail Carbohydrate ES column (5 μm, 250 × 4.6 mm, Altech). A mixture of acetonitrile-water (60:40 % (v/v)), pumped at 0.9 mL/min, was used as mobile phase. The injection volume was defined as 20 μL. 55
Figure 30. Substrates provided in the Chitinase Assay Kit used to detect β-*N*-acetylglucosaminidase/chitobiosidase activity. A- 4-Methylumbelliferyl *N*-acetyl-β-D-glucosaminide, B- 4-Methylumbelliferyl β-D-*N,N'*-diacetylchitobioside hydrate. Figure adapted from Pubchem. 57

Figure 31. Specific ENGase/chitinase activity in *A. gosspii* crude cell extracts (intracellular) and concentrated culture supernatants (extracellular). Specific enzymatic activities were normalized per amount of total protein in the reaction mix. Error bars represent the standard deviation from two biological replicates. 57
Figure 32. Schematic representation of the strategy used for the disruption of *AFR597W* in *A. gossypii*, and verification PCR of the generated mutants. (A) Construction of the *A. gossypii* mutant strain. The *AFR597W* gene was replaced by the *lox*P-*GEN3-lox*P cassette. (B) Verification of the integration of the *lox*P-*GEN3-lox*P cassette in the *AFR597W* locus and confirmation of its excision by colony PCR with primers F3.F/F3.R and V3-Kan_FW/F3.R. 59
Figure 33. Colony radial growth measured in *A. gossypii afr597w* mutant strain and another *A. gossypii* mutant strain that also has resistance to G418 (control) after 4 days of growth on selective agar-solidified AFM. The values represent the mean of two biological replicas. Error bars represent the standard deviation of the biological replicas. 60

LIST OF TABLES

Table 1. Summary of the predicted catalytic residues involved in the hydrolytic and transglycosylation activity of some representative ENGases of the GH85 family. Residues that were experimentally determined to be essential for hydrolytic or transglycosylation activity are represented in bold.... 14
Table 2. Bacterial strains used in this work as cloning and expression hosts. 23
Table 3. Media used in this work for bacterial growth. 23
Table 4. pETM20 vector features. 25
Table 5. Primers used in this work. Overhang sequences are underlined. 27
Table 6. ENGases of the GH85 family. The UniProt Accession Number and the origin of the ENGases used to perform a multiple alignment with the putative ENGases of *A. gossypii* and *Z. roxii* are represented. 36
Table 7. Summary of several computed features for the putative ENGases. The determination of several physical and chemical parameters (theorical pI, predicted MW and extinction coefficient) of the putative ENGases was done using ProtParam tool. Dissulfide bonds were predicted using DiANNA web server. 41
Table 8. Total protein concentration determined for intracellular content and concentrated culture supernatants. 56

LIST OF ABBREVIATIONS AND ACRONYMS

AFM	*Ashbya* full medium
AI	Auto-Induction
BSA	Bovine Serum Albumin
CAZy	Carbohydrate-Active enZymes
CV	Column volume
Dol-P	Dolicol phosphate
Dol-P-Glu	Dolicol phosphate glucose
Dol-P-Man	Dolicol phosphate mannose
Dol-PP-GlcNAc	*N*-acetylglucosamine dolicol pyrophosphate
EDTA	Ethylenediamine tetra acetic acid
ELSD	Evaporative Light Scattering Detector
ENGases	Endo-β-*N*-acetylglucosaminidases
ER	Endoplasmic Reticulum
ERAD	Endoplasmic Reticulum Associated Degradation
G418	Geneticin
GAG	Glycosaminoglycan
Gal	Galactose
GDP	Guanosine diphosphate
GH	Glycoside hydrolases
GlcNAc	*N*-acetylglucosamine
GlcNAc-P	*N*-acetylglucosamine 1-phosphate
GPI	Glycosylphosphatidylinositol
GST	Glutathione S-transferase
HPLC	High-Perfomance Liquid Chromatography
IgG	Immunoglobulin G
IMAC	Imobilized Metal Afinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria-Bertani
MALDI-TOF	Matrix-assisted laser desorption/ionization-time-of-flight
Man	Mannose
MBP	Maltose binding protein

MCP-3	Monocyte Chemotactic Protein-3
MS	Mass spectrometry
MW	Molecular weight
NCL	Native Chemical Ligation
NGT	*N*-glycosyltransferase
NMR	Nuclear magnetic resonance
NOHBY	NO Homolog in Baker's Yeast
NusA	N-utilizing substance A
OD	Optical density
ORF	Open reading frame
OST	Oligosaccharide transferase
PDB	Protein Data Bank
PMSF	Phenylmethylsulphonyl fluoride
PNGase	Peptide:*N*-glycosidase
QC	Quality Control
QMEAN	Qualitative Model Energy ANalysis
RMSD	Root-mean-square deviation
RNase B	Ribonuclease B
SCO	Single Cell Oil
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SNFG	Symbol Nomenclature for Glycans
SOB	Super Optimal Broth
SPPS	Solid-Phase Peptide Synthesis
SUMO	Small ubiquitin-like modifier
TAE	Tris-Acetate-EDTA
TB	Terrific broth
TIM	Triosephosphate isomerase
TrxA	Thioredoxin A
UDP-GlcNAc	Uridine diphosphate-*N*-acetylglucosamine

To my parents, brother and husband

INTRODUCTION

INTRODUCTION

1. Overview on glycosylation

1.1. Glycosylation and its biological significance

Glycobiology is the science that deals with the study of the structure, biosynthesis, biology and evolution of carbohydrates (glycans) distributed in Nature in the context of the biological supports to which they are attached. Glycans are commonly covalently attached to proteins or lipids to form glycoconjugates (glycoproteins, glycolipids and proteoglycans). Considering the structural variability of the glycans, glycoconjugates are hybrid molecules that may present new functionalities and specificities[1]. The most common classes of eukaryotic glycans are represented in **Figure 1**.

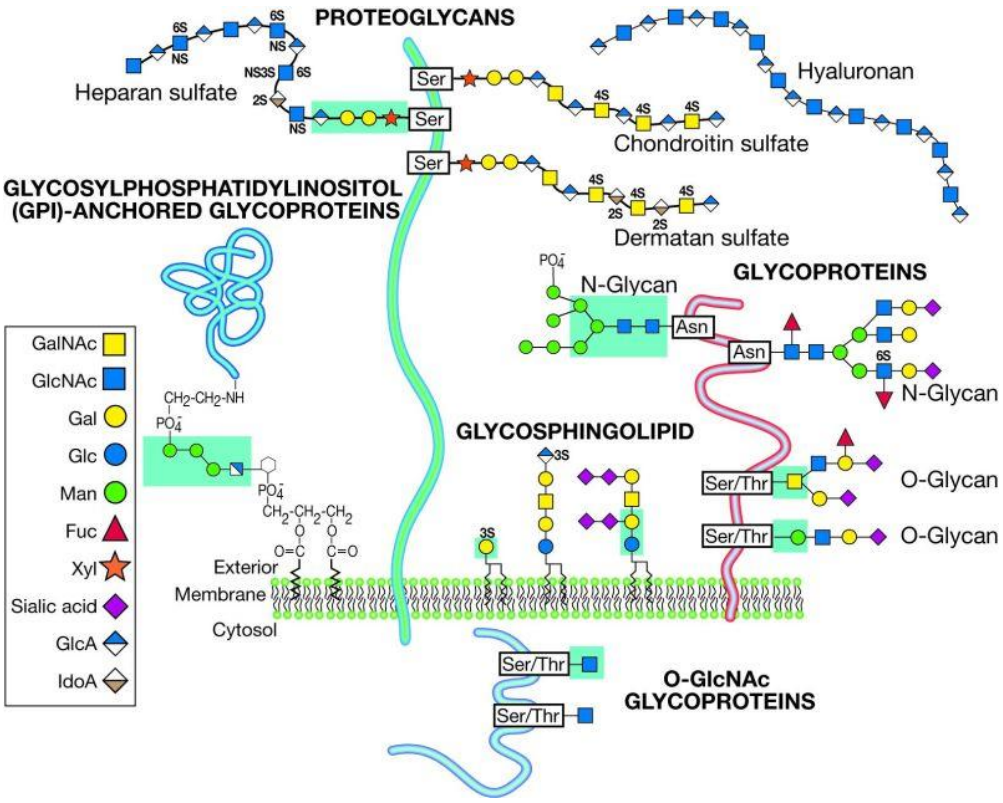


Figure 1. Common classes of eukaryotic glycans. The common structures of a proteoglycan, a glycosylphosphatidylinositol (GPI)-anchored glycoprotein, a glycoprotein and a glycosphingolipid are represented. Figure adapted from 1.

A proteoglycan has one or more glycosaminoglycan (GAG) chains attached to a protein through a typical core region ending in a xylose residue that is linked to the hydroxyl group of a serine residue. On

the other hand, a glycophosphatidylinositol anchor is a glycan bridge between phosphatidylinositol and a phosphoethanolamine (amide linkage to the carboxyl terminus of a protein). A glycosphingolipid (often called a glycolipid) consists of a glycan usually attached via glucose or galactose to the terminal primary hydroxyl group of the lipid moiety ceramide[2].

Glycans are involved in several important cellular communication processes, such as cell adhesion[3], pathogen-host interactions[4] and immune response[5]. Protein glycosylation, a common and essential processing strategy in eukaryotes, differs from other post-translational covalent modifications in relation to the size and complexity of the added group and the magnitude of the cellular machinery that is required for their synthesis and modulation[6]. This post-translational modification can affect the intrinsic properties of proteins, such as folding[7], stability[8], solubility[8] and intracellular traffic[9]. Depending on the type of glycosidic linkage (**Figure 2**), the protein glycosylation can be divided into *N*- or *O*-glycosylation.

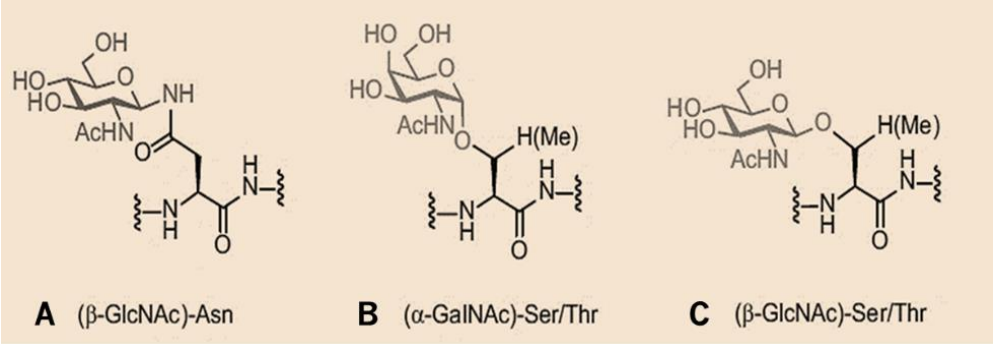


Figure 2. Most common glycosidic bonds. The following glycosidic bonds are represented: **A)** (β-*N*-acetylglucosamine)-*N*-asparagine; **B)** (α-*N*-acetylgalactosamine)-*O*-serine/threonine; and **C)** (β-*N*-acetylglucosamine)-*O*-serine/threonine. The monosaccharides through which the glycans are bound are shown in gray. Figure taken from 9.

In *N*-glycosylation, the binding occurs between the *N*-acetylglucosamine (GlcNAc) of the *N*-glycan and the nitrogen atom of the side chain of an asparagine residue located in the Asn-X-Ser/Thr consensus region (where X denotes any amino acid except proline). In *O*-glycosylation, the first monosaccharide of the glycan (e.g., α-*N*-acetylgalactosamine, GlcNAc or fucose in mammals and mannose in fungi) binds to the protein through the oxygen atom of the side chain of a serine or threonine residue[4,10].

1.2. Diversity of glycans

In general, glycans present a great variety of monosaccharide units connected by different types of glycosidic bonds, structured in a linear or branched way, and can be modified artificially by chemical

or enzymatic methods[11]. There is a symbolic nomenclature for glycans accepted by the scientific community, the SNFG (Symbol Nomenclature for Glycans), which will be used in this dissertation (**Appendix A**)[12].

Most of the reactions involved in the glycoprotein biosynthetic pathway are not complete, which leads to an incomplete addition of glycans to the proteins and subsequent incomplete processing of glycans already bound to the proteins, thereby creating structural diversity. The diversity generated from the sub-stoichiometric transfer of a glycan into a protein is called macro-heterogeneity. On the other hand, the diversity generated from the sub-stoichiometric glycan processing is called micro-heterogeneity. In general, macro-heterogeneity of glycoproteins is caused by the presence or absence of a glycan at a given site in a protein, while the micro-heterogeneity of glycoproteins is caused by the presence of different glycan structures in the same site in a protein[13].

At the organism level, the structural diversity of glycans can vary between cells and between organelles due to differential expression of the processing enzymes. The structural diversity of glycans can still be caused by reactions of phosphorylation, methylation, among others[13]. Based on the type of glycosidic linkage involved in their attachment to proteins, glycans can be categorized into *N*-glycans or *O*-glycans. All the *N*-glycans present in eukaryotes present a common core, Manα1-3 (Manα1-6) Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn-X-Ser/Thr, and are classified into three types (**Figure 3**), according to the modifications made to this base sequence: i) oligomannose, in which there is only extension of the glycan structure with mannose residues (Man); ii) complex, in which the extension of the core is made from ramifications initiated with the addition of GlcNAc residues; and iii) hybrid, in which the core branch extends from Manα1-6 with Man residues and from Manα1-3 with the initial addition of one or two GlcNAc residues. Complex *N*-glycans may have more than six GlcNAc-initiated branches (called antennas) and each one may be elongated with repeats of Galβ1-4GlcNAc (where Gal represents a residue of galactose).

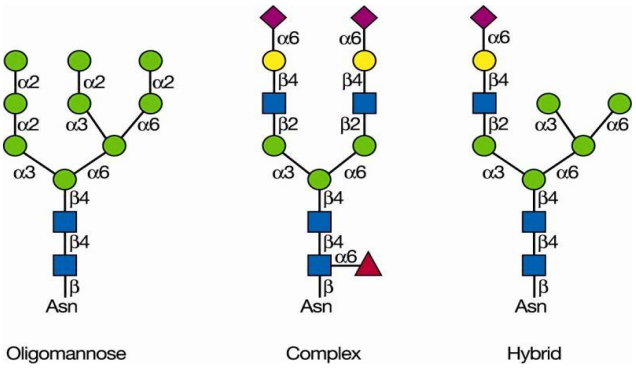


Figure 3. Types of *N*-glycans found in eukaryotes. *N*-glycans, linked to the Asn-X-Ser/Thr consensus sequence on eukaryotic glycoproteins, can be classified into three fundamental types: oligomannose, complex and hybrid. Each *N*-glycan contains a common core, $Man_3GlcNAc_2Asn$. Figure taken out from [14].

Rapid advances in high-throughput mass spectrometry (MS) techniques have allowed an increase in the number of comprehensive glycomics studies based on glycoconjugate fragmentation and its analysis. This type of approach allows evaluating: i) the degree of heterogeneity; ii) the type of glycosylation; iii) the glycosylation sites; iv) the identification of the glycan-linked protein; v) the branching of glycans; vi) the number and length of antennas in complex glycans, their composition and substitution by groups of fucose, sialic acid or other groups; and vii) the determination of complete sequences of individual glycans. Based on the information obtained from these studies it was possible to verify the evolutionary trends in *N*-glycan processing in eukaryotes[1].

In fungi, *N*-glycan processing is restricted by the glycoprotein folding quality control (QC) process and is limited by the diversity of sugar residues that can be used in its extension[1]. The *N*-glycans present in fungi are mainly composed of mannosyl residues (high-mannose type), although some fungi may also exhibit glucose, xylose, fucose, pyruvate or phosphate residues[15,16]. Generally, filamentous fungi produce shorter high-mannose type *N*-glycans, while yeasts produce more extensive high-mannose type *N*-glycans, often branched to generate hypermannose outer chains[17]. Interestingly, the diversity of *N*-glycans produced by the filamentous fungus *Ashbya gossypii* (syn. *Eremothecium gossypii*), which was characterized by MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization-time-of-flight) MS, and nuclear magnetic resonance (NMR)[18], resembles that of yeasts more than that of others filamentous fungi with respect to composition and extent (*N*-glycans of high-mannose type in the range of $Man_{4\text{-}18}GlcNAc_2$, sometimes phosphorylated)[18]. Complex *N*-glycans such as those present in mammals have not yet been detected in fungi[16]. *N*-glycans present in plants have no branching in the core, but have a xylose residue attached to a mannose residue through β-1,2 linkage[19].

There are several classes of enzymes involved in *N*-glycosylation, including glycosidases (EC 3.2.1.-) and glycosyltransferases (EC 2.4.x.y). Glycosyltransferases catalyse the transfer of sugar residues (monosaccharides) from activated donor substrates to specific acceptor molecules, forming glycosidic bonds, while glycosidases catalyse the hydrolysis of the glycosidic bond between two or more sugar residues, or between a carbohydrate and a non-carbohydrate moiety in a regio and stereospecific manner[3].

The major glycosidases involved in the deglycosylation of *N*-glycans occurring in Nature are the peptide:*N*-glycosidase (PNGase; EC 3.5.1.52), which hydrolyses the glycosylamine linkage between two N-acetylglucosamine residues and an asparagine (generating a free peptide and an intact oligosaccharide with the *N*-*N'*-diacetylchitobiose unit at the reducing end), and endo-β-*N*-acetylglucosaminidase (ENGase; EC 3.2.1.96), which cleaves the oligosaccharide linkage between the *N*-*N'*-diacetylchitobiose core of asparagine-linked glycans[20]. Given its relevance to the aim of this project, in section 2 a brief bibliographic review on ENGases is provided. These enzymes are also quite useful for the artificial synthesis of glycoproteins, as discussed in section 1.4.

1.3. *N*-glycan biosynthetic pathway

In eukaryotes, the glycoproteins produced in greater amounts and of greater relevance at the cellular level are those that present *N*-glycans. As such, this topic will discuss its biosynthesis.

N-glycans biosynthesis (**Figure 4**) in eukaryotes initiates on the cytoplasmic face of the endoplasmic reticulum (ER) membrane, with the transfer of an *N*-acetylglucosamine 1-phosphate (GlcNAc-P), from uridine diphosphate-*N*-acetylglucosamine (UDP-GlcNAc), to a lipid precursor, dolicol phosphate (Dol-P), to generate *N*-acetylglucosamine dolicol pyrophosphate (Dol-PP-GlcNAc). This reaction is catalysed by the enzyme *N*-acetylglucosamine-1-phosphotransferase. Subsequently, still in the cytoplasmic side of the ER, $Man_5GlcNAc_2$-P-P-Dol is generated from the transfer of sugar residues from UDP-GlcNAc and GDP-Man, catalysed by specific glycosyltransferases. The $Man_5GlcNAc_2$-P-P-Dol precursor is further translocated to the lumen of the ER, where the formation of the final sugar precursor, Dol-P-P-glycan, is performed by the addition of four Man residues from Dol-P-Man and of three glucose residues from Dol-P-Glc. Afterwards, the glycan constituted by fourteen sugars is transferred *en bloc* into an Asn-X-Ser/Thr consensus sequence of a protein that is being synthesized and translocated through the ER membrane. This transfer reaction is catalysed by a complex of multiple protein subunits called oligosaccharide transferase (OST). Following this transfer, the glycans bonded to proteins are further

remodelled in the ER and Golgi complex through a series of reactions catalysed by enzymes with membrane domains in these organelles, such as α-glycosidases I and II and glycosyltransferases. These enzymes are quite sensitive to the physiological and biochemical conditions of the cells where the glycoproteins are being synthesized. Therefore, the type and mode of branching of the sugars in a mature glycoprotein will depend on the type of cell in which the glycoprotein is produced and the physiological state of the cell at that time[14].

Trimming of the $Glc_3Man_9GlcNAc_2Asn$ begins in the ER lumen with the removal of the glucose residues by the action of α-glycosidase I and II. At this point, correctly folded glycoproteins follow to the Golgi complex, whereas improperly folded glycoproteins are retained in the ER by the protein folding QC machinery or directed to degradation by the ER-associated degradation pathway machinery, ERAD (Endoplasmic Reticulum Associated Degradation)[21,22]. The QC system is composed of lectins and chaperones that maintain glycoproteins in the ER until they are correctly folded or, otherwise, targeted for degradation[23]. These systems help preventing cellular toxicity associated with secretion of misfolded proteins and their potential negative effect on cellular homeostasis. Before exiting the ER, α-mannosidase I may act on some glycoproteins, removing the terminal Man residue and, consequently, generating a $Man_8GlcNAc_2$ isomer[1].

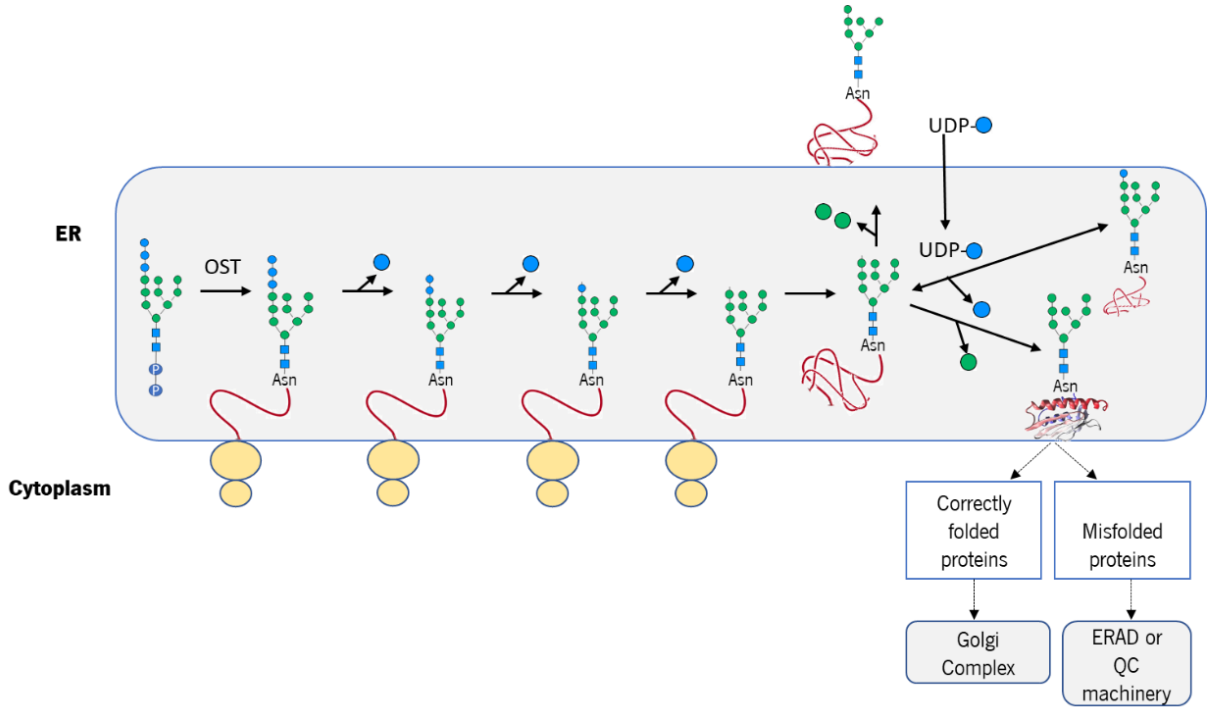


Figure 4. Processing and maturation of an *N*-glycan. The mature Dol-P-P-GlcNAc is transferred to the Asn-X-Ser/Thr consensus sequence of a protein, during protein synthesis, as the proteins are translocated to the ER. After transferring fourteen sugars to the protein, α-glycosidases I and II, present in ER, remove three glucose residues and mannosidase removes a mannose residue. These reactions are directly linked to lectin-assisted glycoprotein folding, which determines whether the glycoprotein continues to the Golgi or is degraded. If folding is correct, the glycoproteins continue maturation in this organelle. Figure adapted from 11.

While the early stages of *N*-glycan processing within the ER are highly conserved across species (from yeasts to mammals), great variability occurs at the Golgi complex level, where glycosidases and glycosyltransferases vary greatly from species to species, leading to a large variety of glycans produced by different organisms or even within different cell types (**Figure 5**)[1,24,25].

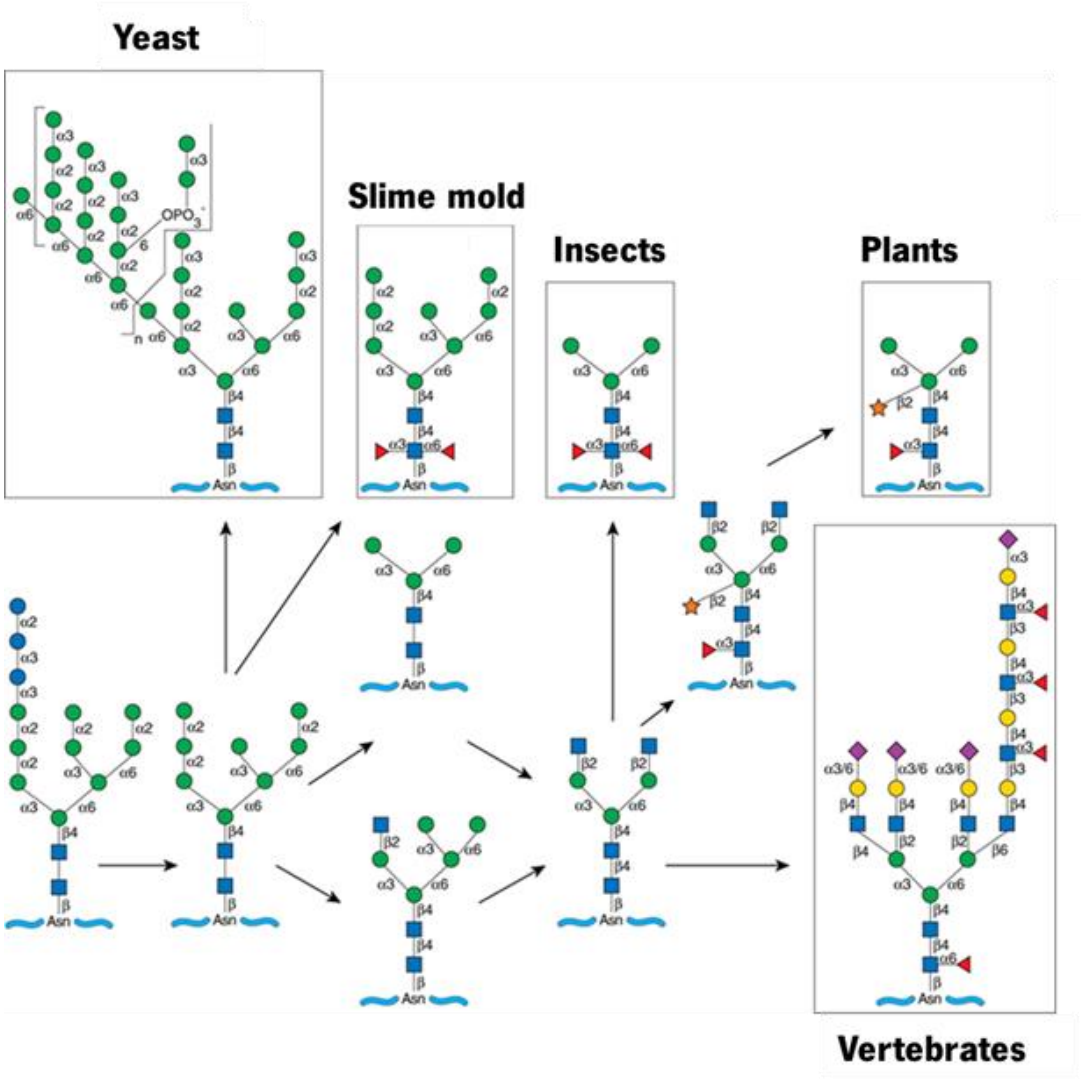


Figure 5. *N*-glycan processing in the Golgi of different eukaryotic taxa. Figure adapted from 1.

1.4. Synthetic methods for glycoprotein engineering

Glycoproteins are normally produced in the cells as a heterogeneous mixture of glycoforms, i.e., with the core polypeptide chain decorated with different glycan structures, eventually attached to different sites. The isolation of pure glycoforms is a complicated process and, as such, to combat the recurrent need to produce homogeneous materials for basic structure-activity studies and for biomedical

applications, several methods have been developed for the synthesis of glycopeptides and glycoproteins with well-defined oligosaccharide structures[26].

The most common strategy used for the synthesis of homogeneous glycoproteins and glycopeptides involves the incorporation of pre-formed glycosyl amino acids in conventional Solid-Phase Peptide Synthesis (SPPS) protocols. This strategy has proved useful for the synthesis of glycopeptides containing small oligosaccharides, but its extension to larger glycoproteins faces several challenges, such as low coupling efficiency and low solubility of the protected polypeptides[27]. Therefore, other synthetic strategies have been developed to tackle these challenges, such as the Native Chemical Ligation (NCL) method, the chemo-selective method, also known as the "tag and modify" method, the chemo-enzymatic method and the direct enzymatic glycosylation method[28]. A schematic summary of these strategies is shown in **Appendix B**.

Regarding the NCL method, it was initially developed for total protein synthesis[29]. This method consists of a chemo-selective transthioesterification reaction between a peptide attached to a thioester group and a cysteine residue at the N-terminus of another peptide fragment that form a native peptide bond without need for protection of the side chains of other amino acids[30]. In 2008, Yamamoto *et al.*[31] performed the first complete chemical synthesis of a glycoprotein containing a complex type *N*-glycan, a glycoform of the MCP-3 protein (Monocyte Chemotactic Protein-3).

The "tag and modify" method has also been used as an alternative to the SPPS method. In this strategy, the glycosylation of recombinant proteins at specific sites can be achieved by chemically selective binding of glycans and proteins via bio-orthogonal tags. The implementation of this strategy consists in the introduction of specific tags in the sites selected for glycosylation through directed mutagenesis. Subsequently, the tags react with a modified glycan via bio-orthogonal chemo-selective binding. The most commonly used tags include the natural amino acid cysteine and other unnatural amino acids with linked azido, aldehyde, alkyne or alkene groups. The major drawback of this strategy is that unnatural (chemo-selective) linkages introduced may sometimes not mimic perfectly the functions of their respective natural linkages[32,33].

An alternative strategy combining the enzymatic synthesis of sugar chains with the chemical synthesis of polypeptides is the chemoenzymatic approach. It is a preferred method for synthesizing sialic acid-containing structures[34]. This strategy requires the preparation of polypeptides with only a linked monosaccharide and the enzymatic extension of the sugar chain with free oligosaccharides in aqueous solutions without the need for protecting groups. The extension of the sugar chain can be achieved with

the assistance of glycosyltransferases and endoglycosidases[35,36]. The main advantage of this strategy is that it is convergent, i.e. allows the synthesis of a high number of complex glycopeptides. In addition, it allows avoiding problems associated with incompatibility of handling protective groups for glycosylation and overall deprotection at the end. The major disadvantages of the chemoenzymatic approach based on endoglycosidases are that the efficiency of the transglycosylation is low and product hydrolysis may occur, since these enzymes also have hydrolytic activity[37].

In 2013, Lomino *et al.*[38] developed a direct enzymatic glycosylation method, which allows the specific binding of a monosaccharide moiety to synthetic or natural polypeptides/proteins. This method consists in the direct glycosylation of free polypeptides with a complex oligosaccharide through the combination of two enzymes, a bacterial *N*-glycosyltransferase (NGT), which catalyses the initial glycosylation of the polypeptide, and an ENGase, which transfers an *N*-glycan to the monosaccharide moiety already bound to the polypeptide. The same authors have demonstrated that the attachment of the monosaccharide to an asparagine residue of the polypeptide is resistant to PNGase-catalysed hydrolysis and has reduced susceptibility to ENGase-catalysed hydrolysis[38].

Alternatively, there are two strategies that allow the synthesis of glycoproteins *in vivo*: the post-translational modification of glycoproteins, through the reengineering of metabolic pathways in yeast[39] and fungi[40], and the co-translating incorporation at a specific site of a synthetic amino acid through suppressor tRNA technology[41]. The re-engineering of metabolic pathways has been increasingly used because of the growing need for recombinant glycoproteins for therapeutic use. Although yeasts and fungi are promising candidates as recombinant protein expression systems, since they can perform typical post-translational modifications of eukaryotes, these organisms modify their glycoproteins with glycans containing many mannoses. Thus, the glycoproteins produced by these organisms generally cannot be used at the pharmaceutical level, as this type of glycans normally affects the pharmacokinetics and pharmacodynamics of recombinant proteins of animal origin, which natively have more complex glycosylation patterns. However, to overcome these difficulties, scientists developed a tool combining the deletion of genes involved in hypermannosylation with the overexpression of specific glycosidases. In most cases, the gene encoding for the Och1p mannosyltransferase is deleted together with the overexpression of an α-1,2-mannosidase[42–44].

All the techniques described above are complementary and can be combined to increase the efficiency of glycopeptide and glycoprotein synthesis. Nevertheless, for the advancement of this field it is necessary to respond to some shortcomings, such as: i) lack of mechanistic and genetic studies related

to glycosylation pathways, which are necessary for designing glycoengineering strategies in hosts other than mammals and for better controlling the final product; (ii) lack of enzymes (native or mutant) that are more efficient and more flexible in the reactions they catalyse and of the glycans they use as substrate; and iii) lack of new methods that allow directed chemical glycosylation of proteins through native sugar-amino acid linkages[33].

2. ENGases

ENGases are glycoside hydrolases (GH) that act on the *N*,*N*'-diacetylchitobiose core of glycoproteins constituted by *N*-glycans. After the hydrolysis of the β-1,4 glycosidic linkage between the two GlcNAc residues, one residue remains attached to the protein and the other becomes the reducing terminal of the released *N*-glycan[45]. According to the Carbohydrate-Active enZymes Database (CAZy), these enzymes are classified into two large families: GH18 and GH85. This classification was developed according to significant similarities in the tertiary structure, catalytic residues and conserved mechanism of action[46]. GH18 family also includes chitinases, which are the representative enzymes of this family. ENGases from GH18 and GH85 families share a homology between 20 to 25 %. ENGases from both families present a TIM (triose phosphate isomerase) barrel in its tertiary structure, consisting of 8 α-helices and 8 parallel β-sheets alternating along the structure.

Regarding their cellular function, ENGases are kwown to be involved in the cytosolic catabolism of free oligosaccharides derived from the ER lumen of eukaryotes, likely to maximize the reuse of sugars in the cells[47]. In bacteria, ENGases have been reported to be involved in nutrient acquisition for bacterial growth[49] and in defence responses[48]. In filamentous fungi, ENGases have been found important for hyphal growth, sporulation and other physiological purposes[47,48]. Since in myxobacteria ENGases have been found to be produced in higher concentration during the sporulation phase[49], ENGases seem to be also important for the sporulation process in bacteria. Nevertheless, the physiological roles of ENGases are still poorly understood.

Considering the scope of this dissertation, the focus of the next topics will be the GH85 family, which is composed of ENGases of prokaryotic and eukaryotic origin[49,50]. According to the Pfam database, the ENGases of this family have only one conserved domain, which corresponds to the catalytic domain that has the designation of PF03644[51].

Murakami *et al.*[52] revealed that the ENGase of the GH85 family found in the methylotrophic yeast *Ogatea minuta* (Endo-Om) is not detected in the culture medium where the cells grow, confirming the theory that yeast ENGases are likely to be located in the cytosol. This observation had previously been verified in other eukaryotic ENGases of the GH85 family, such as the human ENGase, the Endo-CE present in *Caenorhabditis elegans*, and the *Arabidopsis thaliana* ENGase[53–55]. Murakami *et al.*[52] also found that the Endo-Om presented hydrolytic activity in relation to complex *N*-glycans, unlike other ENGases found in bacteria that cannot hydrolyse complex *N*-glycans (such as, Endo-A from *Arthobacter protophormiae*[49], Endo-D from *Streptococcus pneumoniae*[56] and Endo-BH from *Bacillus halodurans*[57]). Relative to ENGases found in fungi, they act at pH between 5.5 and 7.5 and hydrolyse mainly trimannosylated glycans and complex biantennary glycans[20].

In addition to hydrolytic activity, some ENGases of the GH85 family exhibit transglycosylation activity (**Figure 6**), i.e. they can transfer a free oligosaccharide to another acceptor molecule other than water[58]. This activity was detected in Endo-A, which is able to transfer high mannose type *N*-glycans to monosaccharides such as GlcNAc and glucose to form a novel oligosaccharide[59]; in Endo-M from *Mucor hiemalis*, which shows significant transglycosylation activity in relation to complex *N*-glycans[60]; and in Endo-Om, whose transglycosylation activity is most significant in oligosaccharides of the high-mannose type and does not occur in substrates with fucose or bifurcated structures in the GlcNAc[52].

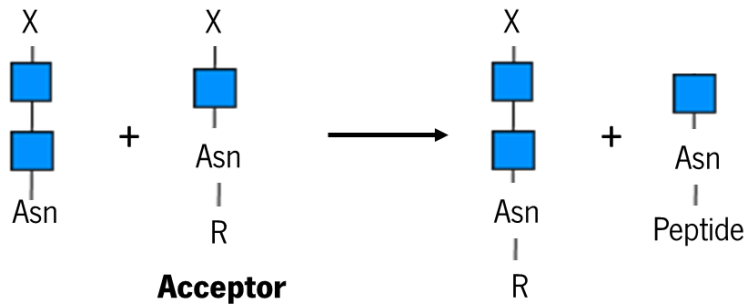


Figure 6. Scheme of the transglycosylation reaction catalysed by some ENGase of the GH85 family. Figure adapted from [58].

2.1. Mechanism of action

The mechanism of action predicted for ENGases of the GH85 family is based on enzyme mutation analyses[61–63] and X-ray crystallography studies of Endo-A[64] and Endo-D[48]. These are the only characterized ENGases of the GH85 family that present an X-ray structure deposited in Protein Data Bank (PDB). All ENGases of the GH85 family retain the configuration of the anomeric carbon and hydrolyse the substrate in a two-step double-displacement mechanism, involving a covalent glycosyl-enzyme intermediate. The reaction involves acid-base catalysis supported by the side chain of two amino acids, typically a glutamate and an asparagine, located at about 5.5 Å[64]. This mechanism of action involves the participation of the neighbouring 2-acetamino group of the second GlcNAc residue of the substrate as an intramolecular nucleophile. An oxazoline is responsible for mediating this mechanism (**Figure 7**).

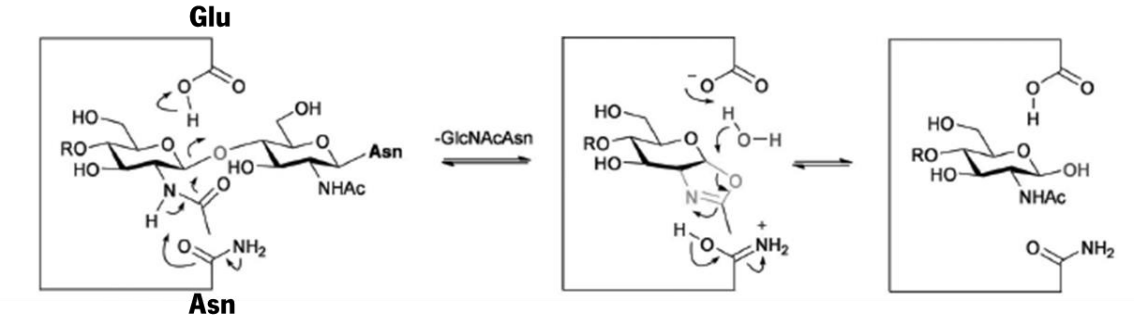


Figure 7. Hydrolytic mechanism of action of ENGases of the GH85 family. The mechanism involves an acid-base catalysis reaction supported by the side chain of a residue of glutamate and asparagine. The intermediate formed is an oxazoline. Figure adapted from[65].

In the first step, the carboxylic acid of the glutamate residue, which forms a hydrogen bond with the glycosidic oxygen, acts as a general acid/base catalyst, facilitating the release of the leaving group. This release is also aided by the concomitant attack of the carbonyl oxygen of the 2-acetamido group of the substrate. Abbott *et al.*[48] proposed that a second conserved asparagine residue acts as the general base, involved in the formation of the oxazoline intermediate, which pKa is modulated by the glutamate residue. The resulting intermediate is cleaved by a process wherein the general acid catalyst facilitates the water attack and the second asparagine residue facilitates the release of the 2-acetamido group to form the product, a hemiacetal sugar[69].

Several ENGases of the GH85 family have already been discovered and yet a few have been characterized. **Table 1** shows the predicted catalytic residues involved in the hydrolytic and

transglycosylation activity of some representative ENGases. In the case of some ENGases, the residues have been experimentally confirmed, but in other cases, they were predicted based on multiple sequence alignments[67]. In addition to the residues described in **Table 1**, there are others at the catalytic centre that are not directly involved in the hydrolytic or transglycosylation activity but are important for other molecular/biochemical functions related to catalysis. For example, the Trp-251 residue present in Endo-M is important for determining the specificity of the substrate. This residue is conserved in Endo-A (Trp-244), Endo-BH (Trp-243), Endo-CC1 (Trp-253) and Endo-Om (Trp-266)[68]. The Tyr-217 residue from Endo-M is also highly conserved in Endo-A (Tyr-205) and Endo-Om (Tyr-231), and is thought to be involved in the formation of the product resulting from the hydrolysis of the intermediate oxazolinium ion[62,69,70].

Table 1. Summary of the predicted catalytic residues involved in the hydrolytic and transglycosylation activity of some representative ENGases of the GH85 family. Residues that were experimentally determined to be essential for hydrolytic or transglycosylation activity are represented in bold.

ENGase (origin)	Residues involved in hydrolytic activity	Residues involved in transglycosylation activity	Reference
Endo-A (*Arthrobacter protophormiae*)	**Asn-171, Glu-173**	**Trp-216**	64,71
Endo-BH (*Bacillus halodurans* C-125)	Asn-170, **Glu-172**	**Trp-215**	57
Endo-BL (*Bifidobacterium longum*)	Asn-230, Glu-232	Trp-279	67
Endo-D (*Streptococcus pneumoniae*)	**Asn-322 e Glu-324**	**His-371**	72
AtENGase 2 (*Arabidopsis thaliana*)	Asn-179, Glu-181	Trp-226	67
AtENGase 1 (*Arabidopsis thaliana*)	Asn-174, Glu-176	Trp-240	67
Endo-CE (*Caenorhabditis elegans)*	Asn-152, Glu-154	Trp-199	67
HsENGase (*Homo sapiens*)	Asn-235, Glu-237	Trp-282	67
Endo-M (*Mucor hiemalis*)	**Asn-175, Glu-177**	**Trp-228**	67
Endo-Om (*Ogatea minuta*)	Asn-194 e Glu-196	**Trp-295**	70

2.2. ENGases applications

The use of enzymes for industrial applications increased over the years, since they have high catalytic efficiency, allow to operate in conditions that are not harmful to the environment and are easily produced by large-scale fermentation[73].

Since glycoproteins are produced in the form of a heterogeneous mixture of glycoforms, currently all recombinantly produced therapeutic glycoproteins, including monoclonal antibodies, are produced and

administered as a mixture of glycoforms[74]. As such, there is an urgent need for access to unique and pure glycoforms of glycoproteins and glycopeptides. ENGases have increasingly become a focus of research studies due to their application in the engineering of pharmaceutical glycoproteins, which are intended to present glycans with homogeneous structures, so that their pharmacokinetics and pharmacodynamics are not affected[65]. These potent biocatalysts have been used in the convergent synthesis of glycopeptides. For example, Wang *et al.*[75] performed the convergent synthesis of CD52 glycopeptides mediated by Endo-A and Endo-M. In addition, ENGases have also been used in the remodelling of glycans linked to immunoglobulin G (IgG), to maximize the functions of these antibodies[76]. Great advances have been made to make ENGases optimal and preferred biocatalysts for synthetic reactions of glycopeptides and glycoproteins[77], among them: the development of synthetic oxazolines as activated donor substrates[35] and the production of glycosynthases that can use activated oxazolines as substrates for transglycosylation reactions and almost do not present hydrolytic activity in relation to the product generated[72,78]. However, there are still few ENGases available commercially and they are quite costly, rendering their application at industrial level unbearable. Among these are Endo F1, Endo F2 and Endo F3 from *Elizabethkingia myicola* (GH18 family), Endo-H from *Streptomyces plicatus* (GH85 family) and Endo-M from *M. hiemalis* (GH85 family)[73].

Given the great diversity of organisms that produce ENGases and the variety of *N*-glycans in which they can act, the discovery and identification of new enzymes is being expanded[65]. Recently, while studying the *N*-glycan profile of *A. gossypii*, Aguiar *et al.*[18] detected the presence of *N*-glycans with only one GlcNAc, indicating that this fungus may produce an ENGase capable of cleaving β-1,4 glycosidic bonds between two residues of GlcNAc. Through a search in the CAZy database for genes that could code for glycosyl hydrolases and glycosyltransferases, they identified in this fungus a gene (*AFR597W*) that could code for an ENGase of the GH85 family, which has no homologue in the closely related yeast *Saccharomyces cervisiae*[18]. In addition, through transcriptomic studies, Aguiar *et al.*[18] detected the presence of transcripts of the *AFR597W* gene. Furthermore, Murakami *et al.*[52] detected endogenous ENGase activity in crude extracts of the yeast *Zygosaccharomyces rouxii*. As such, the characterization of a new ENGase from the GH85 family present in this yeast will also be important. UniProt Database presents a gene that could code for an ENGase in *Z. rouxii*, *ZYRO0B07216g*.

3. *A. gossypii*

A. gossypii is a filamentous hemiascomycete belonging to the *Saccharomycetaceae* family. This fungus was first isolated and characterized by Nowell in 1916 as a phytopathogenic fungus of the cotton plant (*Gossypium hirsutum*)[79]. In the last 30 years, this fungus has been exploited industrially due to its natural production capacity of riboflavin (vitamin B2)[80–82]. In addition, *A. gossypii* has been considered as a cellular factory for the production of other value-added compounds of commercial interest[83,84], such as recombinant proteins[85,86], SCO (Single Cell Oil)[87,88], flavours (nucleoside monophosphate of inosine and guanosine)[93] and aromas (2-phenylethanol alcohol[94] and γ-lactones[91]).

A. gossypii exhibits a very simple life cycle as depicted in **Figure 8**. The main developmental phases of this filamentous fungus, which presents multinucleated hyphae, are: i) the isotropic growth phase during germination, where a germinal spherical cell is formed, ii) the polar growth phase, which begins with the bipolar branch of the germ cell and formation of mycelium that in turn also undergoes polar and dichotomous branching (this only at the tips), and iii) the sporulation phase. The uninucleate spores of *A. gossypii* present a haploid genome[92,93].

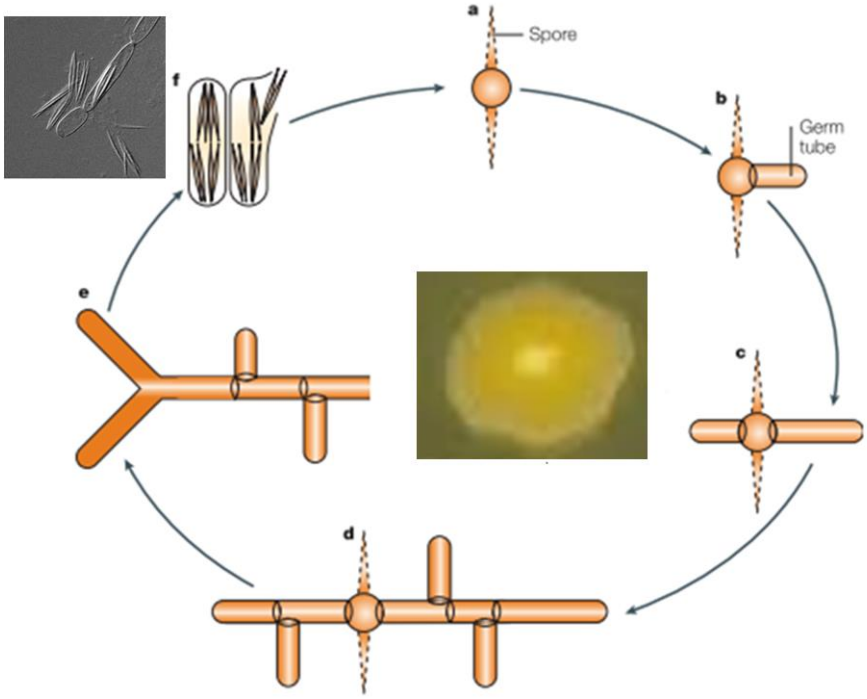


Figure 8. Life cycle of *A. gossypii.* The main phases of the filamentous fungus *A. gossypii* life cycle are illustrated: a) and b) represent the isotropic growth phase; c), d) and e) represent the polar growth phase; and f) represents the sporulation phase. Figure adapted from [93].

This fungus has a higher phylogenetic relationship with yeasts than with other filamentous fungi[94], which makes it a suitable model organism for cellular biology studies of multicellular filamentous fungi[92,93]. The properties that make this fungus suitable as an experimental system include its small genome (which did not undergo whole genome duplication)[95], ease of genetic manipulation and relatively fast growth in defined liquid or solid culture media[96].

The 9.12 Mb genome of *A. gossypii* consists of seven chromosomes, and 95 % of the protein coding sequences found in this fungus have homologues in the genome of *S. cerevisiae*, most of them (4324 open reading frames; ORF) in synthenic locations. This genetic similarity with the well-studied baker's yeast has facilitated the identification of genes encoding proteins with potential activity of biotechnological interest and allowed the reconstruction of the evolutionary past of both these organisms[95]. Moreover, since 259 protein coding genes of *A. gossypii* (5.4 %) have no homologue in baker's yeast (NOHBY), the characterization of these ORFs is important to better understand the physiology of this fungus. Noteworthy, *AFR597W* was identified as NOHBY655[97].

4. *Z. rouxii*

The yeast *Z. rouxii* was first assigned as *Saccharomyces rouxii*[98] and then classified as an hemiascomycete. Based on rRNA sequence comparisons, *Z. rouxii* was found at the time to be one of the yeasts phylogenetically most closely related to *S. cerevisiae*[99,100]. Later, after genome sequencing, it was demonstrated that this species did not undergo the whole genome duplication event that resulted in the genus *Saccharomyces*[101]. *Z. rouxii* and phylogenetically closely related species are now referred as *Z. rouxii* complex, since many works suggested that they exhibit mosaic genome structure with respect to some nuclear and mitochondrial phylogenetic markers[102,103].

At the biotechnological level, *Z. rouxii* is considered one of the main spoilage microorganisms in food industry, which contaminates many food products containing high concentration of sugar and/or salt or low pH, such as sugar syrups, honey, fruit juices, and salad dressings[104].This is due to its halotolerance and osmotolerance, which enables it to grow in environments with high concentration of salts and/or sugars that are usually hostile to most yeast growth[105]. Therefore, this yeast can be employed in the elaboration of balsamic vinegar, soy sauce and miso paste[106]. Given its preference for fructose over glucose, *Z. rouxii* is considered a fructophilic yeast. Moreover, since it can adapt and grow at low pH values, it is also considered an acid-tolerant yeast[101].

Regarding the life cycle of *Z. rouxii*, it is known that this species exists as haploid cells in most of the vegetative phase. Wickerham and Burton[107] suggested that this yeast is heterothallic and has the ability to produce few spores from individual colonies. The haploid life cycle of *Z. rouxii* is schematized in **Figure 9**[108].

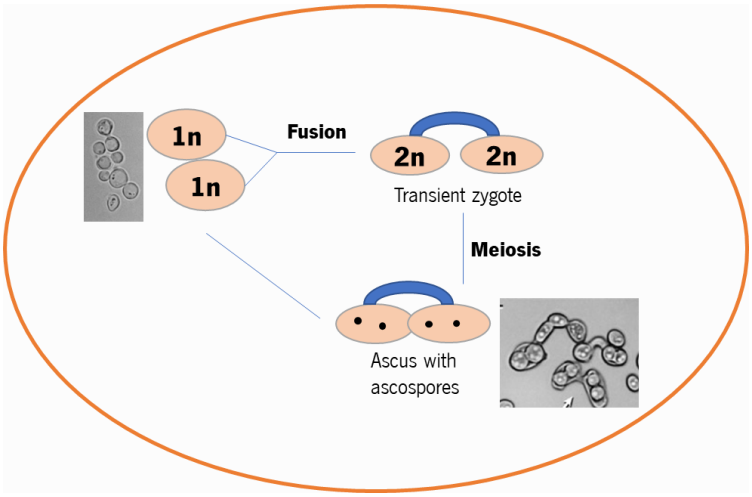


Figure 9. Life cycle of *Z. rouxii*. In most of the life cycle, *Z. rouxii* cells remain in a haploid state and then reproduce by budding, like most conventional yeasts, leading to the formation of a transient heterokaryotic zygote. Afterwards, the diploid zygote undergoes meiosis and sporulation and the haploid state is finally restored. Figure adapted from[108].

The haploid vegetative cells reproduce by budding, like most conventional yeasts, leading to the formation of a transient heterokaryotic zygote. Afterwards, meiosis and sporulation occur and the haploid state is finally restored[108]. However, some haploid *Z. rouxii* yeasts can shift to diploidism and ascosporulation can occur without the formation of the a transient heterokaryotic zygote[109].

5. Recombinant protein production in *Escherichia coli*: advantages and challenges

Currently, there are several hosts used for recombinant protein production, among them, bacteria, yeast, insect, plant and animal cells. *E. coli* is the most common host used for prokaryotic and eukaryotic recombinant protein production, due to its fast growth kinetics, low-cost and high yield of recombinant proteins, well-characterized genome and panoply of available molecular tools for its manipulation[110,111]. The ease to produce recombinant proteins in *E. coli* contributed to several structure-activity studies[70,112,113] and it was also remarkable for the development of the pharmaceutical industry[114]. In the context of recombinant ENGase production, *E. coli* has been used as host to produce Endo-M[62], Endo-

H[115], Endo-A[116], Endo-BH[57], Endo-CE[54] and Endo-CC from *Coprinopsis cinerea*[117]. Note that among these, Endo-M and Endo-CC belong to eukaryotic organisms.

Despite all the advantages of using *E. coli* as an expression system, there are still some drawbacks associated with the expression of recombinant proteins in this organism, among which, lack of enzymatic machinery to perform eukaryotic posttranslational modifications, inefficient formation of disulfide bonds, absence of chaperones to perform correct folding of the expressed proteins and lack of efficient secretion systems. Some of the above-mentioned limitations result, sometimes, in the production of proteins with low stability and solubility inside inclusion bodies[110,118].

Over the years, many strategies have been developed to improve recombinant protein production in *E. coli* , among them: i) the use of genetically improved *E. coli* strains, such as, C41(DE3), Origami and SHuffle[119,120]; ii) optimization of the inducer concentration[121–123]; iii) optimization of the induction temperature[124]; iv) co-expression of chaperones and foldases[118]; v) selection of the ideal growth phase to induce the expression in each strain[125]; vi) the addition of glucose to the medium, resulting in catabolic repression of the *lac* operon[126]; and vii) the addition of additives to the culture medium, such as Triton X-100, urea, ammonium sulphate, arginine, glycerol, and ethylenediamine tetra acetic acid (EDTA)[118]. Moreover, the expression of fusion proteins is often used to improve solubility and facilitate the purification of recombinant proteins. Common fusion tags used are maltose binding protein (MBP)[127], glutathione S-transferase (GST)[128], small ubiquitin-like modifier (SUMO)[129], thioredoxin A (TrxA)[130], NusA (N-utilizing substance A)[131] and poly-histidine (His-tag) for protein purification[132].

OBJECTIVES

ENGases have been studied because of their usefulness in structural and functional glycobiology studies, as well as in the engineering of pharmaceutical glycoproteins, which need to present homogeneous glycans in their structure. However, few ENGases have been characterized thus far and even fewer are available commercially, which has limited their application at the industrial level. Therefore, it is necessary to continue the search for new ENGases with interesting properties, which can be produced at low-cost in enough quantity to be applied in the industry in an affordable way. Moreover, a better clarification of the physiological importance of ENGases is also necessary.

Recently, Aguiar *et al.*[17] identified a gene in the filamentous fungus *A. gossypii* that should code for an ENGase of the GH85 family. Moreover, a gene homologue to that exists in the genome of the yeast *Z. rouxii* and endogenous ENGase activity was already detected in crude cell extracts of this yeast[51].

Envisioning the characterization of the putative ENGases above-mentioned, the objectives of the present master thesis include: i) *in silico* analysis of the potential ENGase activity of these proteins, namely the construction of homology-based models and the identification of possible catalytic residues through multiple comparison alignments, ii) construction of expression plasmids for the recombinant production of these ENGases in *E. coli* , iii) optimization of their recombinant production and purification, iv) optimization of protocols for the analysis of ENGase activity, namely in *A. gossypii* strains, vi) disruption of the *A. gossypii* gene *AFR597W*, which should code a putative ENGase of the GH85 family, and assessment of physiological alterations in the generated mutant.

MATERIALS AND METHODS

MATERIALS AND METHODS

1. Bioinformatics

The automatic annotations retrieved from the databases CAZy, Uniprot and KEGG (Kyoto Encyclopedia of Genes and Genomes) were manually curated using Blastp, Pfam and the PDB[133]. In this way, distant homologies, not evident by tools typically included in automated function annotation pipelines, were sought. Sequence alignments and dendrograms were performed and constructed with Clustal Omega[134], and FigTree, Version 1.4.4 was used to visualize the dendrograms. Homology-based models were constructed using the SWISS-MODEL server (ExPASy)[135]. Molecular visualisation was performed using The PyMol Molecular Graphics System, Version 2.1.0 Schrödinger, LLC. Homology-based models generated were superposed in the SuperPose server, Version 1.0. Molecular weight (MW) was predicted using the ExPASy ProtParam tool[136]. Dissulfide bonds were predicted using DiANNA 1.1 web server[137].

2. Strains and media

A. gossypii ATCC 10895, kindly provided by Prof. P. Philippsen (University of Basel) was used as the background for deletion purposes and cellular extracts analysis. *A. gossypii* was grown at 30 ℃ on agar-solidified *Ashbya* full medium (AFM; 10 g/L tryptone, 10 g/L yeast extract, 20 g/L glucose and 1 g/L myo-inositol), containing 200 μg/mL geneticin (G418; Sigma-Aldrich), for selection and maintenance of transformants[138]. *A. gossypii* spores were prepared and stored as follows. Mycelia grown for up to 2 days on AFM was scraped onto agar-solidified SPA sporulation medium (3 g/L soy peptone, 3 g/L yeast extract, 3 g/L malt extract, 20 mL/L corn step liquor, 10 g/L glucose and 40 g/L agar, pH 6.8) and let grow for another 2-4 days. Mycelia was then digested with 4 mg/mL of lysing enzymes from *Trichoderma harzanium* (Sigma-Aldrich) for 2-3 h at 37 ℃, spores were recovered by low-speed centrifugation (10 min at 4000 rpm) in a table-top refrigerated centrifuge and after two washing steps with spore buffer (8 g/L NaCl, 200 g/L glycerol and 0.25 g/L Tween 20) they were stored at -80 ℃ in the same buffer. For cellular extracts analysis, *A. gossypii* was grown on defined minimal medium[18,139].

Bacterial strains used in this work as cloning (*E. coli* NZY5α) and expression hosts (the rest) are represented in **Table 2**. For long-term storage at -80 ℃, bacterial cells collected from agar-solidified LB

were suspended in 30 % (w/v) glycerol. For short-term storage, bacterial cells were maintained at 4 ℃ for up to 3 weeks in selective agar-solidified LB.

Table 2. Bacterial strains used in this work as cloning and expression hosts.

Strain	Genotype	Source
E. coli NZY5α	*fhu*A2Δ(*arg*F-*lac*Z)U169 *pho*A *glnV*44 Φ80 Δ(*lac*Z)M15 *gyr*A96 *rec*A1 *rel*A1 *end*A1 *thi*-1 *hsd*R17	NZYTech
E. coli Origami 2 (DE3)	*Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL* F′[*lac*$^{+}$ *lacI*^{q} *pro*] (DE3) *gor522*::Tn*10 trxB* (Str^{R}, Tet^{R})	Novagen
E. coli BL21 (DE3)	F^{-} *ompT gal dcm lon hsd*SB(rB- mB-) λ(DE3 **lacI lacUV5-T7 gene 1 ind1 sam7 nin5*])	NZYTech
E. coli SHuffle T7 Express	fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal λatt::pNEB3-r1-cDsbC (SpecR, lacIq) ΔtrxB sulA11 R(mcr-73::miniTn10–TetS)2 [dcm] R(zgb-210::Tn10 --TetS) endA1 Δgor Δ(mcrC-mrr)114::IS10	New England Biolabs

The referred strains were grown in selective agar-solidified or liquid medium containing 100 μg/mL ampicillin (Sigma-Aldrich) for transformants selection. Media used in this work for bacterial growth are represented in **Table 3**.

Table 3. Media used in this work for bacterial growth.

Medium	Components	Source
Luria-Bertani (LB)	Enzymatic digest of casein (10.0 g/L), yeast extract (5.0 g/L) and sodium chloride (10.0 g/L)	Liofilchem
NZY Auto-Induction (AI) LB medium	Not available	NZYTech
Terrific broth (TB)	Tryptone (12 g/L), yeast extract (24 g/L), $K_2HPO_4.3H_2O$ (12.4 g/L), KH_2PO_4 (2.2 g/L) and glycerol (5 mL/L)	Cold Spring Harbor Protocols

When indicated, LB medium was supplement with 1 M sorbitol, 0.8 M NaCl, 0.05 M sodium phosphate buffer pH 7.4 (30 min before induction), 0.5 % (w/v) glycerol or 0.2 M arginine.

3. Molecular Biology

3.1. DNA constructs

3.1.1. Cloning and expression plasmid construction of putative AgENGase and ZrENGase-encoding genes

Based on the amino acid sequence of the putative AgENGase (UniProt Acession Number: Q752H6) and the putative ZrENGase (UniProt Acession Number: C5DRB8), the genes coding for these putative proteins were synthesised *in vitro* with codons optimized for expression in *E. coli* (service performed by NZYTech). DNA sequences were provided by the supplier cloned into pUC57. *E. coli* NZY5α competent cells (approximately 10 μL) were transformed with 0.25 μL of pUC57-*AFRR97W* (pUC57-AgENGase or pUC57-*ZYRO0B07216g* (pUC57-ZrENGase) according to the instructions provided by the supplier of the cells (NZYTech). The cells from each transformation were plated on LB agar plates containing 100 μg/mL of ampicillin and incubated overnight at 37 ℃. Next day, isolated colonies were randomly selected from each plate and streaked onto half LB selective plate for plasmid extraction with the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) according to the manufacturer's instructions, with the exception that another centrifugation step was performed after putting the eluate of the first elution step back in the column to increase the yield of obtained DNA. The assessment of DNA quantity and quality was performed in NanoDrop One (Thermo Scientific).

After quantification of plasmid DNA, 3 to 5 μg of pUC57-AgENGase and pUC57-ZrENGase were digested with *Nco*I and *Xho*I (New England BioLabs). 3 to 5 μg of the expression vector pETM20 (EMBL; **Table 4 and Figure 10**) was also digested with the same enzymes following the same protocol, to ensure correct directionality of the insert. Digestion mixes were then separated in a 1 % agarose gel and the bands corresponding to the desired digested fragments were purified with the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions, with the exception that DNA was eluted with 15 μL Buffer EB (10 mM Tris-Cl pH 8.5) plus 15 μL of ultra-pure water (MilliQ system; Sigma-Aldrich). Agarose gels were prepared by mixing agarose (GenON) in Tris-Acetate-EDTA (TAE; 40 mM Tris base, 20 mM glacial acetic acid and 2 mM EDTA) buffer to a final concentration of 1 % (w/v) and heated in a microwave until boiling. For visualization of DNA in the gel, 3 μL Green Safe Premium (NZYTech) were added to the agarose solution before its polymerization. DNA fragments were separated in these gels at 90 V for 1 h and stained DNA was visualized in a UV transilluminator. The NZYDNA Ladder III (NZYTech) was used as a DNA MW marker.

Table 4. pETM20 vector features.

pETM vector	Fusion protein	Tag size (aa)	Tag MW (kDa)	pI	Promoter	Selection	Protease cleavage site
20	M-*Trx*-His_6-TEV(*ENLYFQ↙G*) AMG-Gene of interest	136	14.7	5.66	T7/*lac*	ampicilin	TEV

aa – amino acids, Da – Dalton, MW – molecular weight, pI – isoeletric point

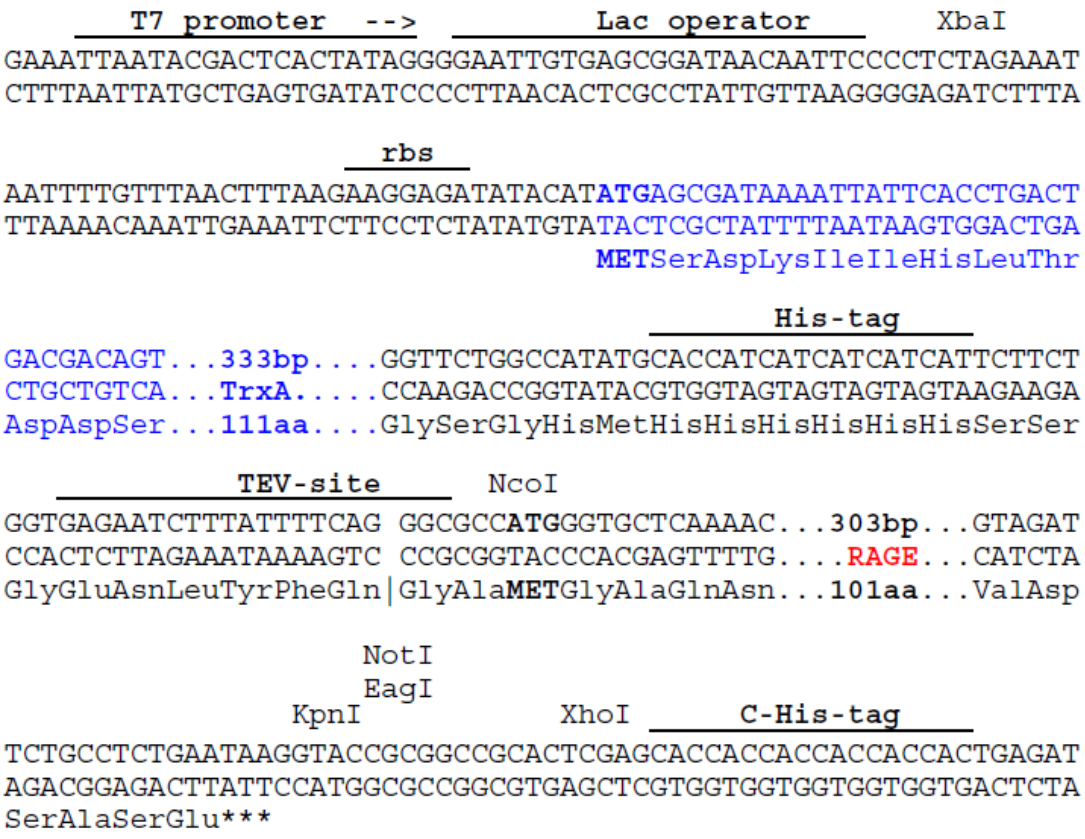


Figure 10. pETM20 vector sequence features.

Ligations into pETM20 were performed overnight at 15 ℃, using a molar ratio of 1:3 and T4 DNA ligase (Promega), according to manufacturer's instructions. A schematic representation of the steps followed in this protocol is represented in **Figure 11**.

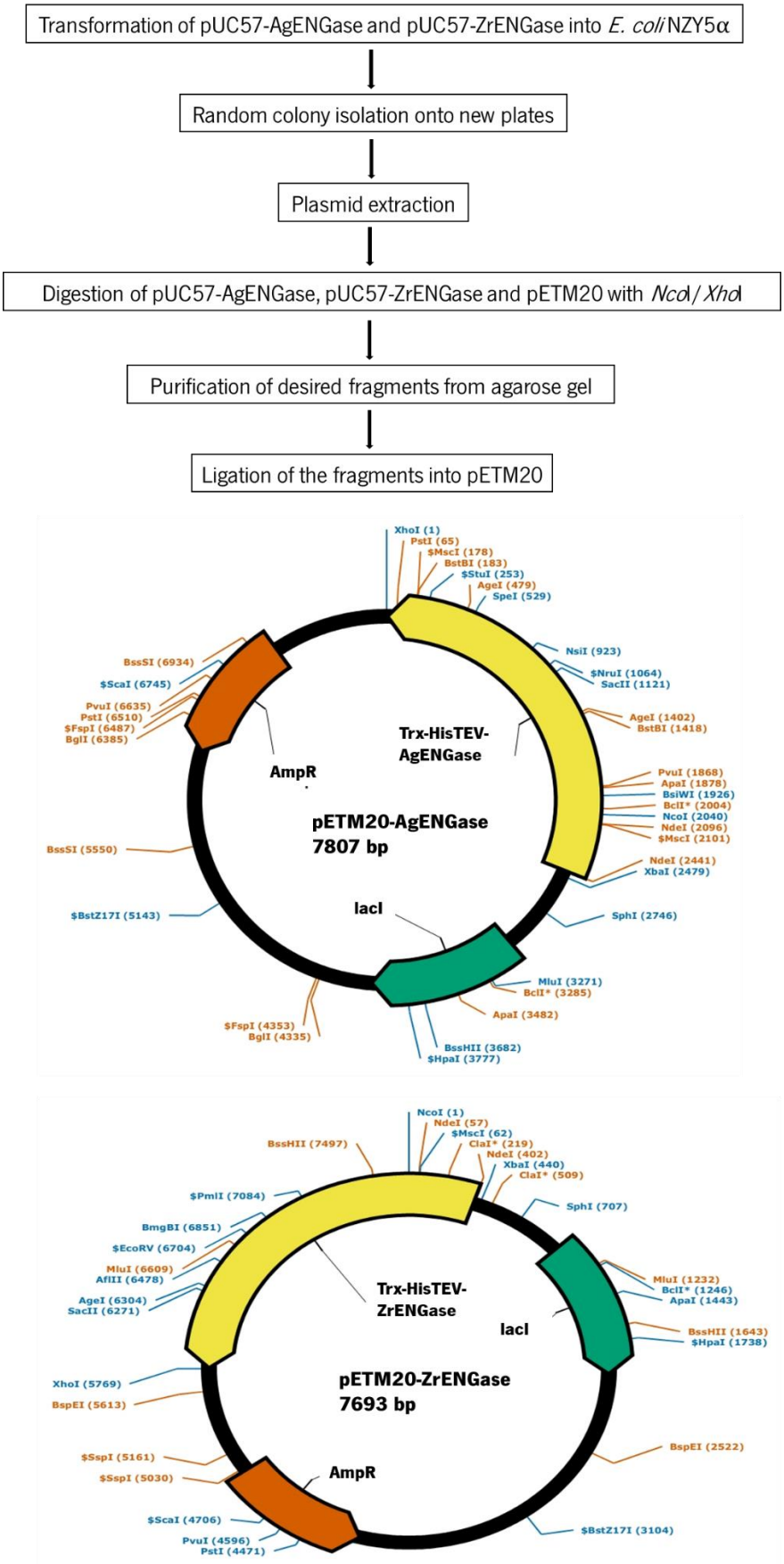


Figure 11. Schematic representation of the steps followed for the construction of pETM20-*ZYR00B07216g* (pETM20-ZrENGase) and pETM20-*AFR597W* (pETM20-AgENGase). PlasmaDNA software was used for *in silico* plasmid construction and representation.

After transformation of 50 µL of *E. coli* NZY5α competent cells with the ligations according to the cells manufacturer's instructions, successful insertion of the genes into pETM20 was initially assessed by colony PCR, using 0.3 µM of each primer T7.FW and T7.RV (**Table 5**), 0.2 U/µL NZYTaq II 2x Green Master Mix (NZYTech) and ultra-pure water up to 50 µL. PCR conditions were: 3 min at 95 ℃, 30 cycles of 94 ℃ for 30 s, 50 ℃ for 30 S, 72 ℃ for 2 min and 30 s, and a final extension at 72 ℃ for 10 min. Subsequently, positive transformants were also confirmed by restriction analysis of the isolated plasmid DNA. To this end, 1-5 µg of DNA was mixed with the NEB restriction enzymes of choice (pETM20-AgENGase was digested with *Pst*I and pETM20-ZrENGase was digested with *Xba*I and *Xho*I), the recommended 10 x CutSmart Buffer (New England BioLabs) and ultra-pure water up to a total volume of 35 µL. Reactions were incubated for 4 h and then analysed in 1 % agarose gels.

Table 5. Primers used in this work. Overhang sequences are underlined.

Primer	Sequence
T7.FW	´GATCCCGCGAAATTAATACGACTCACTATAG
T7.RV	CAAGGGGTTATGCTAGTTATTGCTCAGCGG
AFR597W-FW	CTATGTTGCGCTACGTTTTTGAGTCATACGACGAGCTTCTGGATTGGTTTGAGGGGTGCAGCAGGTCGACAA CCCTTAAT
AFR597W-RV	GGCCAGAAGTTCACGAAAGTAATACCCTTCAAGCGATGTTCTTCAAGAACAAGCCACATAGCATAGGCCACT AGTGGATC
F3.R	CTAGCTATAAGATCCAACGTCG
V3-Kan_FW	TCGCAGACCGATACCAGGATC
F3.F	ATGGACTTTGTCCGTGC

The constructed expression vectors were finally transformed into the *E. coli* expression strains BL21 (DE3), Origami 2 (DE3) and SHuffle T7 Express. 10 µL of competent cells were gently mixed with 1 µL of plasmid DNA and incubated on ice for 30 min. After that, the cells were heat-shocked for 40 s at 42°C and immediately transferred to ice. After 2 min on ice, 480 µL of room temperature SOB medium (Super Optimal Broth; NZYTech) plus 10 µL of 1 M glucose (Biochem) were added to the transformed cells, which were then incubated for 1 h at 37 ℃ and 220 rpm before being plated onto selective LB plates containing the antibiotics necessary to maintain the selective pressure. After overnight incubation at 37°C (or 30 ℃ for *E. coli* SHuffle T7 Express), two randomly selected colonies from each transformation were streaked onto selective LB plates and grown overnight before storage.

3.1.2. Disruption of *AFR597W* gene in *A. gossypii*

For disruption of the *AFR597W* gene, the 1990 bp *lox*P-*GEN3*-*lox*P cassette was amplified from pUGGEN3 using primers AFR597W-FW and AFR597W-RV (**Table 5**) as described by Aguiar *et al.*[140]. The composition of PCR mix was as follows: 5 µL of pUGGEN3 (63.8 ng/µL), 0.25 µM of each primer, 0.2 U/µL NZYTaq II 2x Green Master Mix (NZYTech) and ultra-pure water up to 200 µL. PCR conditions were: 3 min at 95 ℃, 30 cycles of 94 ℃ for 45 s, 50 ℃ for 45 S, 72 ℃ for 1 min, and a final extension at 72 ℃ for 10 min. PCR fragments were then purified using the QIAquick PCR Purification Kit (QIAGEN). The amplified deletion module comprised the resistance cassette flanked by *lox*P sequences and 60 bp guide sequences with homology to the 5' and 3' regions of the target locus, to ensure recombination in the correct location and direction in the genome.

3.2. *A. gossypii* transformation and screening of transformants

A. gossypii ATCC10895 was transformed by electroporation as described by Aguiar *et al.*[140] with 15 µg of the PCR amplified deletion cassette. Transformant colonies were selected on selective AFM plates containing 200 µg/mL of G418, transferred to fresh selective AFM plates, and grown for 5 days at 30 ℃[140].

To verify if the transformant colonies were positive, a piece of mycelium from the border ring of each colony was collected into a microtube containing 30 µL of DNA extraction buffer (0.05 M carbonate buffer pH 9.6, 2 % (w/v) polyvinylpyrrolidone 40, 0.2 % (w/v) bovine serum albumin (BSA) and 0.05 % (v/v) Tween 20). After vortexing, the sample was incubated at 95 ℃ for 10 min to disrupt the mycelia and liberate total DNA. Following centrifugation in a table-top microcentrifuge at 10,000 rpm for 1 min at room temperature, diagnostic PCR reactions were carried out in 25 µL reaction mixtures using 1-2 µL of DNA template, 0.5 µM of each primer, 0.2 U/µL NZYTaq II 2x Green Master Mix (NZYTech). PCR conditions were: 3 min at 95 ℃, 30 cycles of 94 ℃ for 30 s, 50 ℃ for 1 min, 72 ℃ for 30 s, and a final extension at 72 ℃ for 10 min. Correct integration of the *lox*P-*GEN3*-*lox*P cassette in the *AFR597W* locus was verified using the primers F3.F/F3.R and V3-Kan_FW/F3.R (**Table 5**)[140].

After confirmation that the clones had the correct integration in the genome, to isolate homokaryotic clones, a piece of mycelium grown in selective AFM plates for 2 days was streaked onto SPA sporulation plates and after 2 days of growth at 30 ℃, the mycelium was collected and digested as described above for 30 min to 1 h. After two washing steps with spore buffer, the digested mycelia was

resuspended in sterile spore buffer and plated onto AFM plates containing 1000 μg/mL of G418. After 2 days of growth, a piece of mycelium from the border ring of each colony was analysed by diagnostic PCR as described before[146].

To assess the growth of the *A. gossypii afr597w* mutant (*afr597wΔ*::*lox*P-*GEN3*-*lox*P), a pinhead of mycelia from fresh selective AFM plates (containing 1000 μg/mL of G418) was digested as described above for 30 min, washed and resuspended in 100 μL of sterile spore buffer, and then used as inoculum. *A. gossypii* ATCC10895 transformed with a plasmid conferring resistance to G418 was treated similarly for use as control. Agar-solidified AFM plates (90 mm diameter) containing 1000 μg/mL of G418 were then inoculated with 7 μL of digested mycelia and incubated at 30 ℃ for 4 days. Colony radial growth was determined by measuring the diameter of colonies in two perpendicular directions as described in [141].

4. Production and purification of recombinant ENGases

A pre-inoculum was prepared by inoculating 5 mL of selective LB medium with biomass from a fresh plate of *E. coli* BL21 (DE3), Origami 2 (DE3) or SHuffle T7 Express transformed with pETM20-AgENGse or pETM20-ZrENGase. After overnight growth at 37 ℃ and 200 rpm, 500 μL of each pre-inoculum were used to inoculate 50 mL of selective AI, LB or TB media (in 250 mL shake flasks). To promote biomass formation, the cultures prepared in AI medium were incubated first at 37 ℃ (for *E. coli* Origami 2 (DE3) or BL21 (DE3)) or 30 ℃ (for *E. coli* SHuffle T7 Express) and 200 rpm for 5 h and then, to promote recombinant protein production, for another 19 h at 20 or 37 ℃ and 150 rpm. In the case of cells grown in LB medium, biomass formation was promoted at 37 ℃ (for *E. coli* Origami 2 (DE3) or BL21 (DE3)) or 30 ℃ (for *E. coli* SHuffle T7 Express), and expression was promoted at 37, 20, 16, 18 or 4 ℃ after induction with 1 mM (for expression at 37 ℃), 0.2 mM or 0.1 mM of IPTG (Isopropyl β-D-1-thiogalactopyranoside) when OD_{600} was between 0.2-0.3 or 0.4-0.6. In the case of cells grown in TB medium, biomass formation was promoted at 37 ℃ (for *E. coli* Origami 2 (DE3) or BL21 (DE3)) or 30 ℃ (for *E. coli* SHuffle T7 Express), and expression was promoted at 20 ℃ after induction with 0.2 mM of IPTG when OD_{600} was between 0.8-1.2. Cells were then harvested by cebtrifugation, fresh weight was calculated, and the biomass pellet frozen at -20 ℃. A quarter of the total biomass obtained from culture centrifugation (approximately 200 mg) was thawed and incubated for 20 min at 25 ℃ with 1.5 mL of NZY Bacterial Cell Lysis Buffer (NZYTech) containing 10 mg/mL of lysozyme, 2 mg/mL of DNase I and

1 mM PMSF (phenylmethylsulfonyl fluoride). When indicated, the lysis buffer was supplemented with 0.3 M NaCl, 5 % (w/v) glycerol, 1 M sucrose, 0.2 M arginine and/or 2 M urea. After incubation, cell debris were pelleted by centrifugation for 15 min at 4 ℃ and 10,000 rpm.

For small-scale purification of the histidine-tagged recombinant proteins, HisPur™ Ni-NTA Resin (Thermo Scientific) was used, following the batch protocol at room temperature indicated by the manufacturer. The equilibration and washing buffer used was composed by 20 mM sodium phosphate buffer pH 7.4, 0.5 M NaCl and 40 mM imidazole, while the elution buffer contained 300 mM imidazole. After purification, imidazole was removed from the fractions containing the protein of interest using a PD-10 Desalting Column (GE Healthcare) according to manufacturer's instructions for the gravity protocol. Samples were collected at each purification step for further analysis by SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis).

For medium-scale purification of the recombinant proteins, HisTrap HP column (5 mL; GE Healthcare) was used according to manufacturer's instructions. Samples were collected in each step for further SDS-PAGE analysis. After purification, desalting was performed as before.

4.1. SDS- and Native-PAGE Analysis

Total protein fractions for SDS-PAGE analysis were prepared from 100 µL of culture as follows. After centrifugation at 10,000 rpm (4 ℃) for 10 min, the pellet was resuspended in 10 µL of 5x SDS-PAGE Sample Loading Buffer (NZYTech) plus 40 µL of distilled water. Soluble protein fractions were obtained after cell lysis and recovery of 16 µL the resulting supernatants, to which 5x SDS-PAGE Sample Loading Buffer (NZYTech) was added prior to loading. Insoluble fractions were prepared by resuspending the cell debris obtained after cell lysis in distilled water plus 5x SDS-PAGE Sample Loading Buffer (NZYTech). Prior to loading, 20 µL of these samples were boiled at 100 ℃ for 10 min and then 15 µL used to load SDS-PAGE gels.

For SDS-PAGE analyses, proteins were resolved on 15 % SDS-PAGE gels under denaturing conditions, using a 4 % stacking gel and the MW marker PageRuler™ Unstained Broad Range Protein Ladder (5-250 kDa; Thermo Scientific). Gels were run in Tris-Glycine-SDS buffer (0.025 M Tris, 0.192 M glycine and 0.1 % (w/v) sodium dodecyl sulphate (SDS)) at 15 mA for ~30 min. Gels were stained with Coomassie brilliant blue G-250 using the method described in [142].

For Native-PAGE analyses, proteins were resolved on 10 % Native-PAGE gels under nondenaturing conditions, using a 4 % stacking gel. No protein marker was used since proteins will migrate according to their charge, size and conformation. Prior to loading, the samples were resuspended in 5x Native-PAGE Sample Loading Buffer (0.313 M Tris-HCl pH 6.8, 50 % (w/v) glycerol and 0.05 % (w/v) bromophenol blue). Gels were run in Tris-Glycine buffer (0.025 M Tris and 0.192 M glycine) at 15 mA for approximately 30 min and then the *in gel* zymographic assay described by Magalhães *et al.*[143] was used for *in situ* detection of β-galactosidase. Briefly, gels were incubated in a solution of 0.02 % (w/v) X-gal in 0.2 M sodium acetate buffer (pH 4.5) for 5 h at room temperature with slow shaking, after which they were washed with distilled water.

5. ENGase/chitinase activity detection

5.1. Analysis of ENGase deglycosylation activity

The following glycoproteins were used as substrate to assess ENGase deglycosylation activity: RNase B from bovine pancreas (Sigma-Aldrich), β-galactosidase from *Aspergillus niger* (Maxilact, DSM) and invertase from *S. cerevisiae* (Sigma-Aldrich), all of which contain oligomannose type *N*-glycans.

For SDS-PAGE mobility shift assays, a 10 μL substrate protein solution was prepared, containing 1 % (v/v) Triton X-100, 10 mM EDTA, and 5 μg RNase B, 4 μg β-galactosidase or 0.5 mg invertase. *N*-glycan digestion was then performed as follows: the 10 μL substrate protein solution was combined with 1 μL of commercially obtained Endo-H (New England Biolabs), 4 μL of 250 mM sodium acetate buffer (pH 5.0) containing 1.25 mM NaCl and incubated overnight at 37 °C. Deglycosylation of glycoproteins was monitored by mobility shift of bands on 15 % SDS-PAGE gels after staining with Coomassie Brilliant Blue.

For High-Performance Liquid Chromatography (HPLC) assays, a 300 μL of a substrate protein solution was prepared, containing 1 % (v/v) Triton X-100, 10 mM EDTA, and 120 μg β-galactosidase or 15 mg invertase. *N*-glycan digestion was then performed as follows: 100 μL of substrate protein solution was combined with commercially obtained Endo-H (100 U/reaction with 1 × G5 reaction buffer; New England Biolabs) and 40 μL of 250 mM sodium acetate buffer (pH 5.0) containing 1.25 mM NaCl[52], and incubated overnight at 37 °C. Deglycosylation of glycoproteins was monitored by HPLC using a Jasco chromatograph equipped with an Evaporative Light Scattering Detector (ELSD SEDEX 85, SEDERE) and

a Prevail Carbohydrate ES column (5 μm, 250 × 4.6 mm, Altech). A mixture of acetonitrile-water (60:40 % (v/v)), pumped at 0.9 mL/min, was used as mobile phase. The injection volume was defined as 20 μL[144].

5.2. Analysis of β-*N*-acetylglucosaminidase and chitobiosidase activity in *A. gossypii*

A pre-inoculum of *A. gossypii* ATCC 10895 was prepared as follows: 10^7 spores were inoculated into 250 mL shake-flasks containing 50 mL of AFM and grown overnight (14-16 h) at 30 ℃ and 200 rpm. Next day, mycelia from 20 mL culture were collected by filtration and resuspended in 100 mL of defined minimal medium to an optical density at 600 nm ($OD_{600\ nm}$) of ~1. The 100 mL inoculum was then equally distributed in two sterile 250 mL flasks and the mycelia were grown for 24 h at 30 ℃ and 200 rpm. After 24 h of growth, the $OD_{600\ nm}$ was recorded and culture supernatants were separated from the mycelia by filtration. Before determining extracellular enzymatic activity, culture supernatants were concentrated approximately 60x in Amicon® Ultra-15 10,000 MWCO centrifugal filter devices (Millipore), with simultaneous buffer exchanged to 50 mM sodium phosphate buffer pH 7.4 containing 1.25 M NaCl. The biomass collected from each flask culture was equally distributed in two 2 mL tubes and stored at -80 ℃ until lyophilization.The lyophilized biomass was then grinded with a pestle and resuspended in 1.25 mL of protein extraction buffer (50 mM sodium phosphate buffer pH 7.4, 1.25 M NaCl, 1 M PMSF, 0.5 % (v/v) Triton X-100 and 1x complete protease inhibitor cocktail (Roche)). After centrifugation at 10,000 rpm (4 ℃) for 10 min, the soluble fraction of the crude cell extracts was recovered and used for further intracellular enzymatic activity tests.

The total protein content in the supernatants and cellular extracts was quantified according to the Bradford Assay - Standard Procedure for Microtiter Plates (Bio-Rad). Dye Reagent Concentrate (Bio-Rad) was diluted 1:4 in bi-distilled water and the solution was then filtered through a 0.22 μm filter and stored a 4 ℃ for further use. To construct the calibration curve (**Appendix C**), six dilutions of the Pierce™ Bovine Serum Albumin Standard (2 mg/mL; Thermo Scientific) were prepared: 0.05 mL, 0.1 mg/mL, 0.2 mg/mL, 0.3 mg/mL, 0.4 mg/mL and 0.5 mg/mL. Protein solutions were assayed in triplicate. The reaction was started by the addition of 200 μL of diluted dye reagent to each well of the microtiter plate already containing 10 μl of each standard or sample solution. Reactions were incubated at room temperature for 10 min and absorbance at 595 nm was then measured.

Fluorimetric detection of β-*N*-acetylglucosaminidase and chitobiosidase activity was performed using the Chitinase Assay Kit (Sigma-Aldrich) with the substrates 4-methylumbelliferyl *N*-acetyl-β-D-glucosaminide and 4-methylumbelliferyl *N,N'*-diacetyl-β-D-chitobioside, respectively. Briefly, the enzymatic hydrolysis of the substrates releases 4-methylumbelliferone (4MU), which upon ionization in basic pH, can be measured fluorimetrically. Solutions required to perform the assay were prepared according to manufacturer's instructions. For each substrate, a separate activity assay was performed. 10 µL of samples tested corresponded to culture supernatants and intracellular content after *A. gossypii* mycelium lysis. Reactions were performed in duplicate according to the manufacturer's instructions. According to kit manufacturer, one unit of chitinase activity will release 1 µmole of 4MU from a given substrate per minute at pH 5.0 at 37 °C. The ENGase/chitinase activity was calculated using a standard curve prepared from the fluorescence readings of the 5 standard solutions (**Appendix D**).

RESULTS AND DISCUSSION

RESULTS AND DISCUSSION

1. Bioinformatic characterization

ENGases of the GH85 family are a class of enzymes (EC 3.2.1.96) that, in addition to hydrolytic activity against the diacetylchitobiose core of *N*-glycans, can also display transglycosylation activity[65]. According to the CAZy database, family GH85 currently contains 768 members, of which only 11 have been characterized thus far. This study focused on the study of two new putative ENGases of this family: Q752H6 from the filamentous fungus *A. gossypii* (AgENGase) and C5DRB8 from the yeast *Z. rouxii* (ZrENGase). The putative AgENGase was recently annotated as being involved in the KEGG pathway (ec00511) of *N*-glycan biosynthesis and metabolism[145]. The putative ZrENgase is predicted to be involved in the same pathway. These putative ENGases share a sequence identity of 34.3 %.

The putative ZrENGase (GenBank Accession Number: XP_002495262.1) is encoded from the ZYRO0B07216g ORF (GenBank Accession Number: XM_002495217.1) present on chromosome B of *Z. rouxii*.

The putative AgENGase is annotated in the NCBI database (GenBank Accession Number: NP_986144) as being translated from the *AFR597W* ORF (GenBank Accession Number: NM_212280.3) present in chromosome VI of *A. gossypii*, which is annotated as containing a -1 frameshift. Therefore, one must bare in mind that the 677 amino acid (aa) sequence annotated as being encoded by this gene may not correspond to the actual translation occurring in *A. gossypii* cells. Nevertheless, transcripts of this gene have been detected in different culture conditions[97,145], indicating that the gene is actually expressed. When analysing through a Blastp search the sequence of the homologues of *AFR597W* in the species *Eremothecium cymbalariae* and *Eremothecium sinecaudum*, it was verified that the ribosomal frameshift found in *A. gossypii* is not conserved among the genus *Eremothecium*.

In order to identify in the putative AgENGase and ZrENGase conserved residues that may be important for the hydrolytic and transglycosylation activity of ENGases of the GH85 family, a multiple sequence alignment between different ENGases of this family (**Table 6**) was made using the Clustal Omega tool. Furthermore, a dendogram was built based on the multiple alignment performed (**Figure 12**).

Table 6. ENGases of the GH85 family. The UniProt Accession Number and the origin of the ENGases used to perform a multiple alignment with the putative ENGases of *A. gossypii* and *Z. roxii* are represented.

ENGase	Origin	UniProt Accession Number
Endo-D	*Streptococcus pneumoniae*	Q93HW0
Endo-BB	*Bifidobacterium longum*	Q8G4P3
Endo-BH	*Bacillus halodurans*	Q9KER4
Endo-A	*Arthrobacter protophormiae*	Q9ZB22
Endo-M	*Mucor hiemalis*	Q9C1S6
Endo-Om	*Ogatea minuta*	R4WHQ8
ZrENGase	*Zygosaccharomyces rouxii*	C5DRB8
AgENGase	*Ashbya gossypii*	Q752H6
Endo-CE	*Caenorhabditis elegans*	Q19089
DmENGase	*Drosophila melanogaster*	Q86B46
AtENGase 2	*Arabidopsis thaliana*	Q9SRL4
AtENGase 1	*Arabidopsis thaliana*	F4JZC2
HsENGase	*Homo sapiens*	Q8NFI3
MmENGase	*Mus musculus*	Q8BJF1

It should be noted that genome or gene sequences of *A. protophormiae*, *B. halodurans* C-125, *B. longum* NCC2705, *M. hiemalis*, *C. elegans*, *D. melanogaster*, *A. thaliana* ORF At3g11040 (AtENGase2) and *A. thaliana* ORF At5g05460 (AtENGase1) have been updated in the NCBI database since the publication of the articles that report the characterization of these enzymes and present now a different GenBank Accession Number (U59168.1, BA000004.3, AE014295.3, AB060586.1, AB079783.1, AY075262.1, AC009991.5, AB010692.1, respectively). As such, the putative or experimentally determined residues involved in the ENGase activity will appear in a different position in the sequence represented below, as compared to the published data.

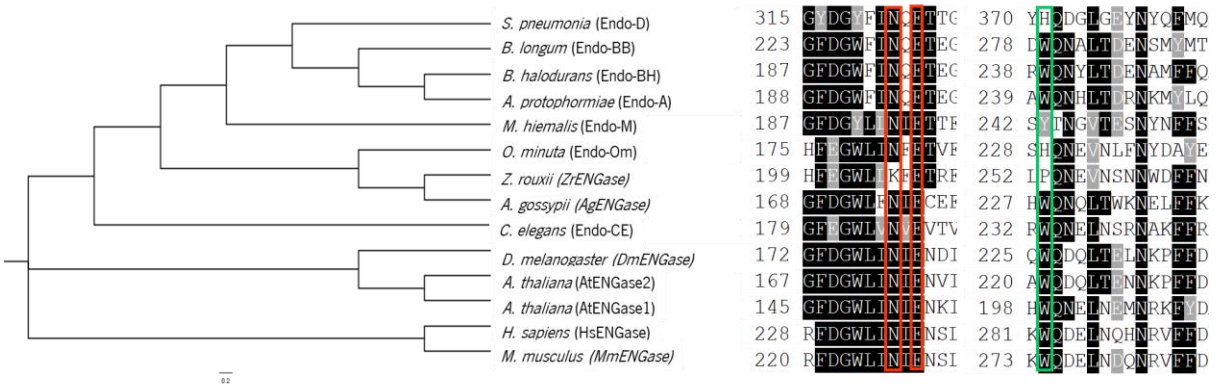


Figure 12. Multiple alignment of several ENGases of the GH85 family, namely the putative ENGases of *A. gossypii* and *Z. rouxii,* and corresponding dendogram. Multiple alignment was performed using the Clustal Omega tool and the dendogram was built based on the multiple alignment performed and representend using the FigTree tool. Residues conserved between the homologues are shaded in black. The putative or experimentally determined residues involved in the hydrolytic (red box) and transglycosylation activity (green box) are represented. Accession numbers for the represented ENGases are described in **Table 6**. Data included in the poster presented at the 4th Meeting of Medicinal Biotechnology (**Appendix E**).

Through analysis of the dendogram it is possible to verify that the putative ENGases of *A. gossypii* and *Z. rouxii* are phylogenetically close between each other and to Endo-Om from *O. minuta* (**Figure 12**). Multiple alignment analysis allows to identify putative catalytic residues involved in ENGase activity. As such, regarding the putative ENGase from *A. gossypii,* it was found that it conserves two catalytic residues that have been attributed to the hydrolytic activity of ENGases of the GH85 family (Asn-175 and Glu-177) and it also conserves a Trp (228) residue that has been reported to be involved in transglycosylation activity[52]. On the hand, the putative ENGase of *Z. rouxii* only conserves a Glu in the position 208 (equivalent to Glu-177 in AgENGase) and does not conserve the Trp residue important for transglycosylation activity (**Figure 12**). The putative residues involved in hydrolytic and transglycosylation activity are included in the catalytic domain for ENGases of the GH85 family predicted by the Pfam database (region between amino acid number 155 and 399 in AgENGase and region between amino acid number 93 and 375 in ZrENGase). These results indicate that the putative ENGase from *A. gossypii* will be more likely to exhibit transglycosylation and hydrolytic activity than the putative ENGase of *Z. rouxii*, despite ENGase activity has already been detected in crude extracts of this yeast[52].

Another important residue described in *O. minuta* and the plant pathogen fungus *M. hiemalis*, which is thought to be involved in substrate recognition and is required for hydrolytic activity against complex oligosaccharides only in eukaryotic ENGases, is the hydrophobic Trp residue at position 281/295 in *O. minuta/M. hiemalis*[52,146]. This residue is conserved in the putative ENGases of *A. gossypii* and *Z. rouxii*. In addition to the Trp-281/295 residue conserved in *O. minuta*/ *M. hiemalis*, the Glu-144/151 residue is also conserved in these organisms and it is thought to have an identical function to

that of the Trp-281/295 residue. This residue is conserved in the putative ENGase of *Z. rouxii* but not in *A. gossypii* (**Figure 13**). This result may indicate that the putative ENGase of *Z. rouxii* is more likely to exhibit hydrolytic activity towards complex *N*-glycans.

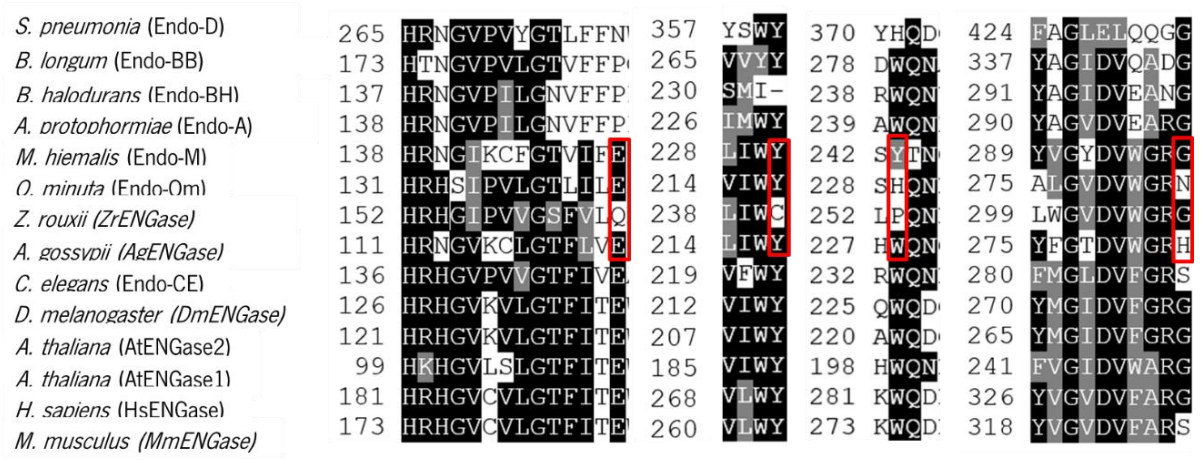


Figure 13. Multiple alignment of several ENGases of the GH85 family, namely the putative ENGases from *A. gossypii* and *Z. rouxii*. Multiple alignment was performed using Clustal Omega tool. Residues conserved between the homologues are shaded in black. Important residues involved in the catalysis besides the catalytic residues that may be conserved in Endo-M, Endo-Om and the putative ZrENGase and AgENGase are represented in a red box. Accession numbers for the represented ENGases are described in **Table 6**.

The Tyr-231 residue in Endo-M was found to be highly conserved between Endo-A, Endo-Om and the putative AgENGase, but not in the putative ZrENGase (**Figure 13**). It is thought that this residue may be involved in the formation of the hydrolytic product from the oxazolium intermediate[62,69,70]. In *Z. rouxii*, the residue is replaced by a cysteine, an amino acid with polar neutral side chain. In addition, the Tyr-243 residue that may act as substrate gate-keeper at the active site of Endo-Om[70] is not conserved in the putative AgENGase and ZrENGase (**Figure 13**).

To elucidate the possible residues that may be involved in the hydrolytic activity of the putative AgENGase and ZrENGase, two homology-based models were constructed using the SWISS-MODEL software to predict the 3D structure of these proteins. The template from which the models were built was an Endo-A mutant (PDB: 2VTF.1.A; **Figure 14**). The putative AgENGase presented 25.58 % amino acid identity with the template and a coverage in the region of the catalytic domain of 0.44. On the other hand, the putative ZrENGase presented less amino acid identity with the template (18.32 %) and a higher coverage in the region of the catalytic domain (0.85).

Figure 14. Endo-A structure used as template for the homology-based models of AgENGase and ZrENGase. Endo-A mutant (PDB: 2VTF.1.A) structure was used as a template for the construction of the homology-based models of AgENGase and ZrENGase. This structure was visualized in The Pymol Molecular Graphics System, Version 2.1.0 Schrödinger, LLC.

Qualitative Model Energy ANalysis (QMEAN) around zero indicateS good agreement between the model structure and experimental structures of similar size, while scores of -4.0 or below are an indication of models with low quality. The model generated for the putative AgENGase (**Figure 15A**) has a QMEAN of -3.93, which is acceptable, while the model generated for the putative ZrENGase (**Figure 15D**) has a QMEAN of -4.72, which is very low. This very low QMEAN of the model generated for the putative ZrENGase is due to the low similarity between the template protein sequence and the protein sequence with which the model was constructed.

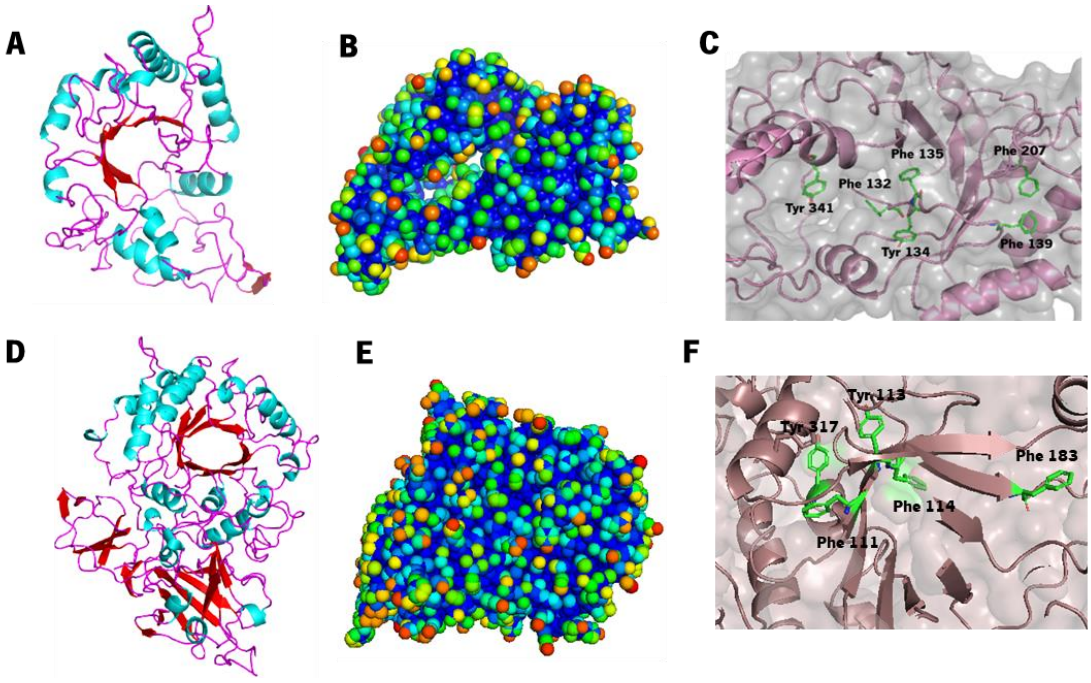


Figure 15. Prediction of the 3D structure of the putative AgENGase and ZrENGase from homology-based models. The models (A and D) were built using the SWISS-MODEL server and were visualized using The Pymol Molecular Graphics System, Version 2.1.0 Schrödinger, LLC. **A-** Predicted model for the possible AgENGase 3D structure, **B-** Representation of the solvent accessible surface of the putative AgENGase; **C-** Representation of residues that may establish hydrophobic interactions around the catalytic pocket (Tyr-341 Phe-132, Tyr-134, Phe-135, Phe-207 and Phe-139), **D-** Predicted model for the possible 3D structure ZrENGase, **E-** Representation of the solvent accessible surface of the putative ZrENGase, **F-** Representation of residues that may establish hydrophobic interactions around the catalytic pocket (Phe-111, Tyr-317, Tyr-113, Phe-114 and Phe-183). Data included in the poster presented in the 4ª Meeting of Medicinal Biotechnology (**Appendix E**).

Regarding the secondary structure of the predicted models, the putative AgENGase presents five β sheets surrounded by eight α helices (**Figure 15A**). This secondary structure is different from the secondary structure predicted by X-ray crystallography for the Endo-A mutant used as a template, which has a TIM barrel (β_8/α_8) characteristic of the ENGases of the GH85 family[64,70]. On the other hand, the possible ZrENGase may have a complete TIM barrel, according to the model (**Figure 15D**). This result may indicate that the putative ZrENGase is more likely to present ENGase activity than the putative AgENGase, although the predicted model is not very feasible (very low QMEAN) and the multiple alignment results indicate that putative AgENGase is more likely to present hydrolytic and transglycosylation activity.

In the case of the putative AgENGase, it is possible to visualize in the structure the hydrophobic pocket where the substrate is positioned during catalysis (**Figure 15B**), whereas in the case of putative ZENGase is not possible (**Figure 15E**). This is probably because the structure generated in the latter model is more complex.

The possible residues that may establish hydrophobic interactions around the catalytic pocket of AgENGase and ZrENGase are residues conserved between several ENGases of the GH85 family. Only the Phe-139 residue of the putative AgENGase is not conserved in the putative ZrENGase; it is substituted by a cysteine residue (**Figure 15C, F**).

In addition, the generated models were superposed using the SuperPose server and the model resulting from the superposition was visualized in Pymol tool (**Figure 16**).

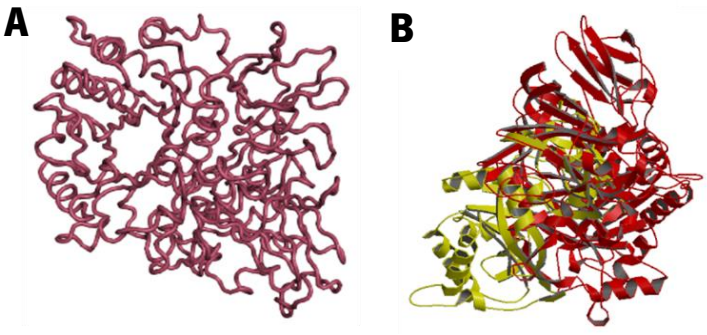


Figure 16. Overlapping of the generated models for the putative AgENGase and ZrENGase. A- Model resulting from superposition of the predicted model for the possible AgENGase with the predicted model for the possible ZrENGase; **B-** Overlapping of the secondary structure of the possible AgENGase (in yellow) with that of the possible ZrENGase (in red).

For the SuperPose model resulting from overlapping the predicted structures for the putative AgENGase and ZrENGase, an identity of 22.2 % between the sequences and an overall root-mean-square

deviation (RMSD) value of 33.75 Å was obtained. This high RMSD value indicates that there is a high average distance between the atoms of the two overlapping proteins, i.e., their structures are apparently quite different (**Figure 16**), probably due to poor models' quality.

To characterize both AgENGase and ZrENGase, the proteins were produced recombinantly in *E. coli*. The determination of several physical and chemical parameters (theorical isoelectric point, predicted MW and extinction coefficient) of the recombinant putative ENGases was done using ProtParam tool (Expasy). Dissulfide bonds were predicted using DiANNA web server. The results obtained are summarized in **Table 7**. According to the UniProt database, putative ENGases may be located in the cytoplasm.

Table 7. Summary of several computed features for the putative ENGases. The determination of several physical and chemical parameters (theorical pI, predicted MW and extinction coefficient) of the putative ENGases was done using ProtParam tool. Dissulfide bonds were predicted using DiANNA web server.

Protein	Protein size (aa)	pI	Molecular weight (Da)	Extinction coefficient (M^{-1} cm^{-1} at 280 nm)	Predicted dissulfide bonds
AgENGase	811	6.12	91844	157510 Abs 0.1 % (=1 g/L) - 1.715*	6
ZrENGase	773	6.46	88039	140565 Abs 0.1 % (=1 g/L)- 1.597*	7

pI- Isoelectric point, aa- amino acids, Da- Dalton; * assuming all pairs of Cys residues form cystines

2. Production and purification of recombinant AgENGase and ZrENGase

As an initial attempt to produce ZrENGase in *E. coli*, the native ORF encoding this protein was amplified from the genome of *Z. rouxii* NRRLY 2547, cloned into pETM10 and pETM20 and expressed in *E. coli* BL21 (DE3), Rosetta 2 (DE3) and Origami 2 (DE3). Cloning into pETM20 present advantages over pETM10, since it allows the production of the protein of interest in fusion with the TrxA tag at the N-terminal, which results in increased solubility of the recombinant protein[158,15147,148]. *E. coli* BL21 (DE3) is the most widely used T7 expression strain and its deficiency in both *lon* and *ompT* proteases result in decreased protein degradation during the purification process[156,15].

E. coli Rosetta 2 (DE3) is a derivative of BL21 designed to enhance the expression of recombinant eukaryotic proteins that have rare codons for *E. coli*. *E. coli* Origami 2 (DE3) is a K-12 derivative that

contains mutations in the glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) genes, which enhance its capability to correctly fold eukaryotic proteins with disulfide bonds in the cytoplasm[112]. Different production media (LB and AI), induction temperatures (37, 23 and 18 ℃) and inducer concentrations (1 mM and 0.2 mM IPTG) were tested with the different strains, but from the analysis of total protein extracts there were no conclusive evidences indicating the production of ZrENGase in any of the conditions tested (**Appendix F**). Most of the protein bands visible in the gels were present in the insoluble fraction (data not shown) and the protein of interest seemed to be "masked" by other *E. coli* intrinsic proteins, even after purification (**Appendix G**).

Since it is known that the presence of rare codons for *E. coli* on the target sequence can impair recombinant protein production[119] and the native ORFs encoding ZrENGase and AgENGase, have a considerable number of rare codons for *E. coli*, as determined using the ATGme tool[149], codon-optimized DNA sequences were used for subsequent plasmid constructions.

2.1. Expression plasmid construction

The newly constructed plasmids pETM20-AgENGase and pETM20-ZrENGase containing codon-optimized sequences of the gene of interest were transformed into *E. coli* NZY5α and positive transformants were confirmed by colony PCR (**Figure 17A**). The pET system was chosen for cloning the genes of interest, since it is one of the most powerful and widely used for *E. coli* expression[160].

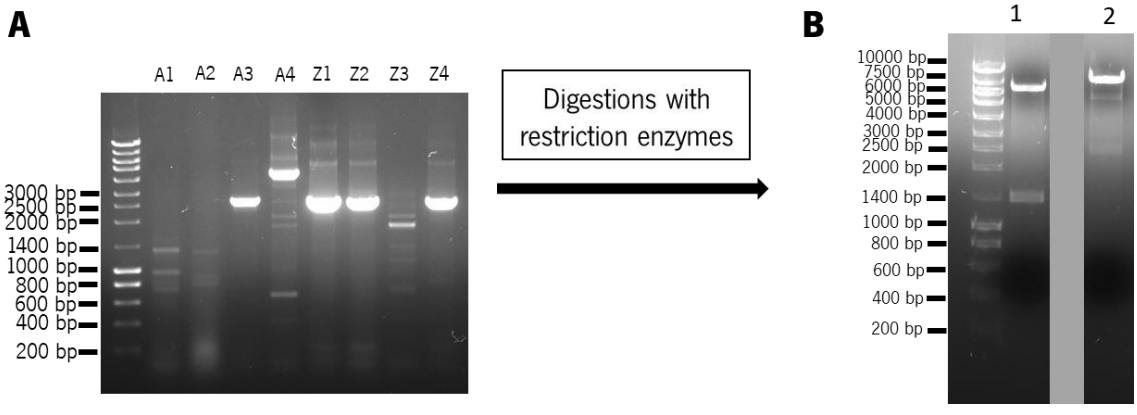


Figure 17. Confirmation of *E. coli* NZY5α clones containing pETM20-AgENGase and pETM20-ZrENGase. A- Colony PCR products were assessed by electrophoresis on a 1 % TAE agarose gel and the image was acquired with an UV transilluminator. A1-A4 correspond to the clones tested for pETM20-AgENGase transformants (one expected amplicon with 2045 bp) and Z1-Z4 correspond to the clones tested for pETM20-ZrENGase transformants (one expected amplicon with 1931 bp). **B-** After plasmid extraction, digestion with restriction enzymes was performed to verify if the insert was in the correct orientation. DNA fragments were separated by electrophoresis on a 1 % TAE agarose gel and the image was acquired with an UV transilluminator. Lane 1 corresponds to pETM20-AgENGase digested with *Pst*I (two expected fragments with 6445 and 1362 bp) and lane 2 corresponds to pETM20-ZrENGase digested with *Xba*I and *Xho*I (two expected fragments with 5329 and 2364 bp).

According to the constructions predicted size (2045 bp for pETM20-AgENGase and 1931 bp for pETM20-ZrENGase), it was verified by colony PCR that clones A3 (corresponding to pETM20-AgENGase) and Z1, Z2 and Z4 (corresponding to pETM20-ZrENGase) were positive (**Figure 17A**). Clones A3 and Z1 were chosen for further experiments.

Next, the plasmids extracted from these clones were digested with selected restriction enzymes in order to confirm if the insert was in the correct orientation according to the reading frame (**Figure 17B**). pETM20-AgENGase was digested with *Pst*I and pETM20-ZrENGase was digested with *Xba*I and *Xho*I. Predicted fragments size after digestion were confirmed by electrophoresis and the results indicated that the inserts were in correct orientation.

These plasmids were then transformed into *E. coli* BL21 (DE3), Origami 2 (DE3) and SHuffle T7 Express. This last engineered strain was also used because, like Origami 2 (DE3), it has enhanced capability for the formation of correct dissulfide bonds in recombinant proteins of eukaryotic origin[156].

2.2. Optimization of production and purification of the recombinant proteins

Since the putative AgENGase and ZrENGase may have dissulfide bonds (as predicted by DiANNA web server in section 1 of Results and Discussion), their production was initially tested in *E. coli* Origami 2 (DE3) and SHuffle T7 Express. Strains were grown in LB and AI media at 37 ℃ and 20 ℃, which are the optimum growth temperatures for *E. coli* Origami 2 (DE3) and SHuffle T7 Express, respectively, and protein production was induced in LB medium using 1 mM IPTG at 37 ℃ and 0.2 mM at 20 ℃ (**Figure 18**).

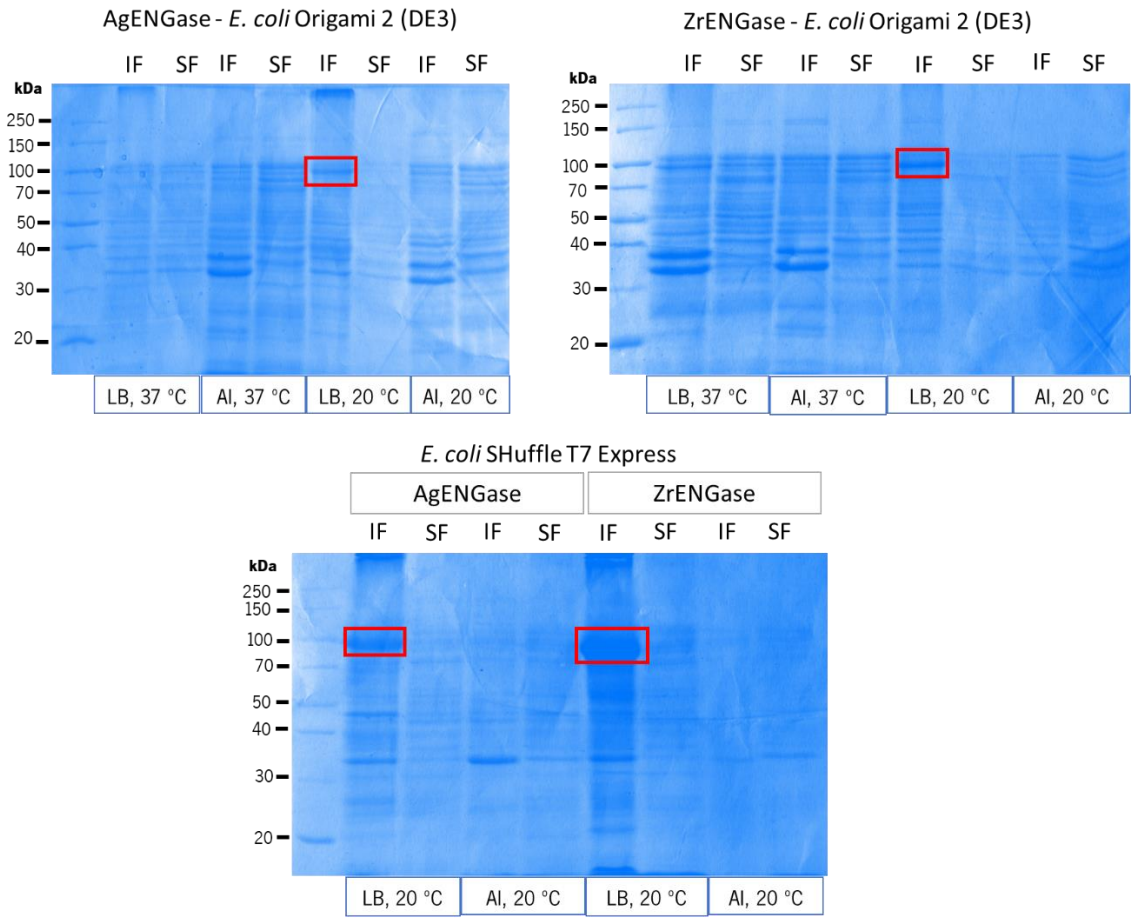


Figure 18. Recombinant production of AgENGase and ZrENGase in *E. coli* Origami 2 (DE3) and SHuffle T7 Express in LB and AI media at 37 and 20 °C. Insoluble (IF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes. Note that insoluble fraction at 37 °C was diluted 1:5 in order to facilitate band visualization.

By comparing the band patterns of insoluble and soluble fractions in the different conditions tested and the non-observation of bands with the expected sizes (92 kDa for TrxA-His-tag-TEV-AgENGase kDa and 88 kDa for TrxA-His-tag-TEV-ZrENGase), it was verified that at 37 °C no ENGases were evidently produced by *E. coli* Origami 2 (DE3; **Figure 18**). However, for better understanding if the recombinant proteins were being or not produced, a control experiment should have been performed with the transformed strains containing an empty vector.

E. coli is usually grown at 37 °C, but this temperature does not facilitate the production of all proteins[128]. Since the putative ENGases may be complex proteins with a high MW, this result is not unexpected. On the other hand, at 20 °C, it seems that both recombinant ENGases are being produced by both strains in LB medium, but the production seems to be mostly insoluble (red boxes in **Figure 18**). Usually, reduction of the induction temperature may facilitate the soluble expression of recombinant proteins[123].

LB medium is the standard medium used for *E. coli* growth. AI medium is, in turn, a more recent formulation that has proved to improve protein production[123,161]. However, it was verified that when both strains were grown in AI medium both ENGases were not produced. There were some bands in the gels that seem to be the proteins of interest. However, the band pattern in the soluble fraction of both ENGases is very similar, indicating that those bands may correspond to native *E. coli* proteins whose production is augmented in this medium (**Figure 18**). Collins *et al.*[123] suggested that the impairment in the production of recombinant proteins in AI medium may be due to a premature induction, resulting in a reduced cell growth and increased metabolic stress.

Since the induction at 20 ℃ seemed to be most favourable (red boxes in **Figure 18**) to produce the putative ENGases, small-scale batch purifications were performed using the soluble fractions collected from the productions of both strains in LB and AI medium. Nevertheless, the proteins of interest were not visible in any of the analysed eluates, since proteins were already not present in the soluble fraction (**Appendix H**).

TB medium was also tested as production media at 20 ℃ for both strains (**Figure 19**). This medium was proved to favour higher cell densities, since it allows the extension of the exponential growth phase of *E. coli*. As such, total protein expression may be favoured. Note that, when cells are grown in TB medium, the induction of protein expression is effectuated at the late-exponential growth phase (OD_{600} 0.8-1-2)[162].

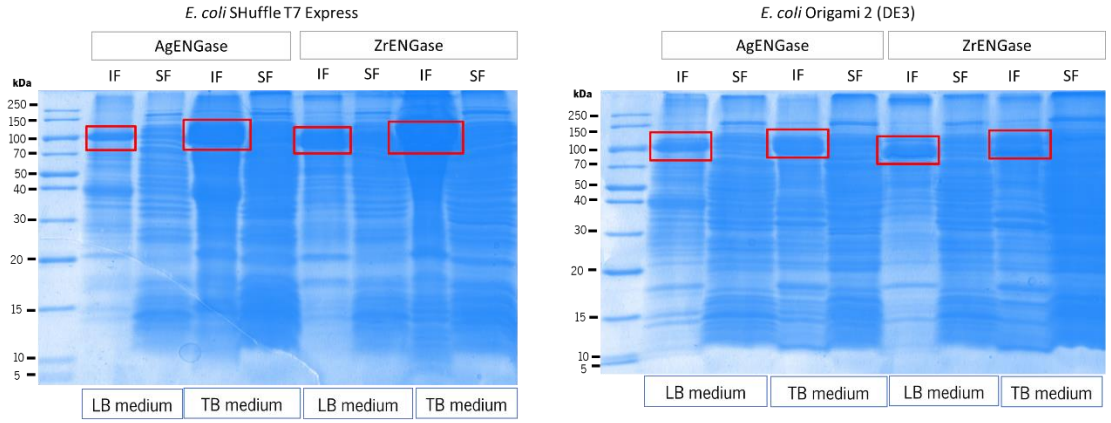


Figure 19. Recombinant production of AgENGase and ZrENGase in *E. coli* Origami 2 (DE3) and SHuffle T7 Express in LB and TB media at 20 ℃. Insoluble (IF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes.

It was verified that the putative ENGases are evidently produced by *E. coli* Origami 2 (DE3) and SHuffle T7 Express in TB and LB media at 20 ℃, but their production seems to be mostly insoluble, possibly in inclusion bodies (**Figure 19**). When comparing the production between LB and TB media, TB medium increases the total protein production and, in particular, the production of the proteins of interest, even though in insoluble form (red boxes in **Figure 19**). As overexpression of total proteins occurs when cells are grown in TB medium and considering that the proteins of interest have a high apparent MW[118], this fact may explain why the putative ENGases are being produced in inclusion bodies. Also, the purification process of soluble fractions from productions in TB medium appeared to be more difficult than in the case of LB medium (**Appendix I**).

Small-scale purifications under mild reducing and denaturing conditions (40 mM DTT and 0.5 % (w/v) SDS) of the soluble fractions of both strains' productions in LB and TB media at 20 ℃ were performed and it was verified that the proteins of interest were not evidently present in the analysed eluates (**Appendix I**). However, a molecular band corresponding to the predicted MW of the TrxA-His-tag-TEV-AgENGase fusion (92 kDa) was observed among others in the eluates from the purifications of soluble fractions collected from productions in TB medium (Panels B and H in **Appendix I**).

Bearing in mind that strain BL21 (DE3) is widely used to produce recombinant proteins in *E. coli*[62], production of the putative ENGases in this strain was also tested (**Figure 20**).

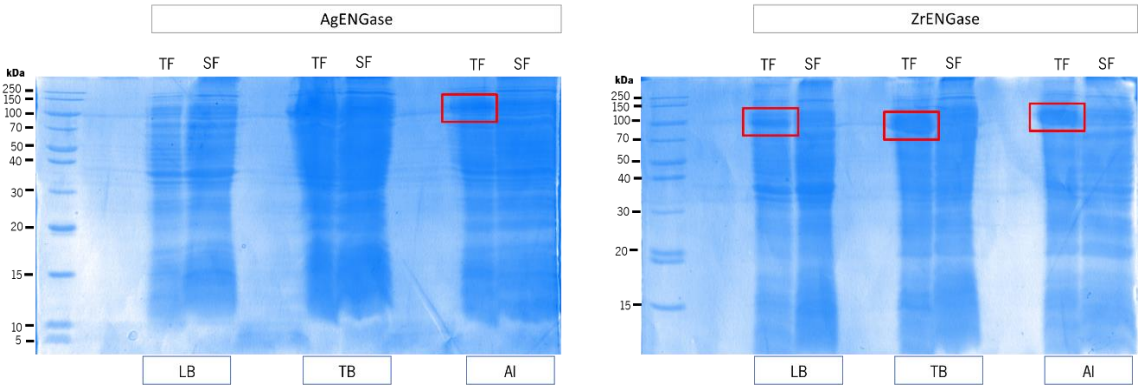


Figure 20. Recombinant production of AgENGase and ZrENGase in *E. coli* BL21 (DE3) in LB, TB and AI media at 20 ℃. Total (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes.

Analysing the obtained gels (**Figure 20**), it appears that both ENGases are produced by this strain, but mostly in an insoluble form. Regarding the differences between the media tested, it was again observed that the TB and AI media favour the production of higher protein concentration than the LB

medium (**Figure 20**), namely of the proteins of interest (although it is difficult to assess AgENGase production in TB medium from the acquired image).

Small-scale batch purifications of the soluble protein fractions from *E. coli* BL21 (DE3) productions in LB, TB and AI media at 20 ℃ were performed (**Appendix J**), but the proteins of interest were not present in the analysed soluble fractions, since they did not appear in the eluted fractions after IMAC purification.

In order to try to avoid the formation of inclusion bodies, production conditions were further optimized. Initially, the induction temperature and the inducer concentration were reduced, as these are among the most used strategies to improve soluble protein production[112]. At the lysis step, various reports indicated that the addition of NaCl in concentrations between 0.3 to 0.5 M enhance protein stability and solubility[163]. Also, the addition of 5 or 10 % (w/v) of glycerol is important for recombinant protein stability[164,165]. As such, both ENGases were produced in *E. coli* Origami 2 (DE3) in LB medium at 18 ℃ using 0.1 mM or 0.2 mM IPTG and lysis conditions were optimized by the addition of 0.3 M NaCl and 5 % (w/v) glycerol to the lysis buffer. *E. coli* Origami 2 (DE3) was the strain chosen to produce the putative ENGases under these conditions, since it was observed that it was the strain with best production results at 20 ℃ (**Figure 21**).

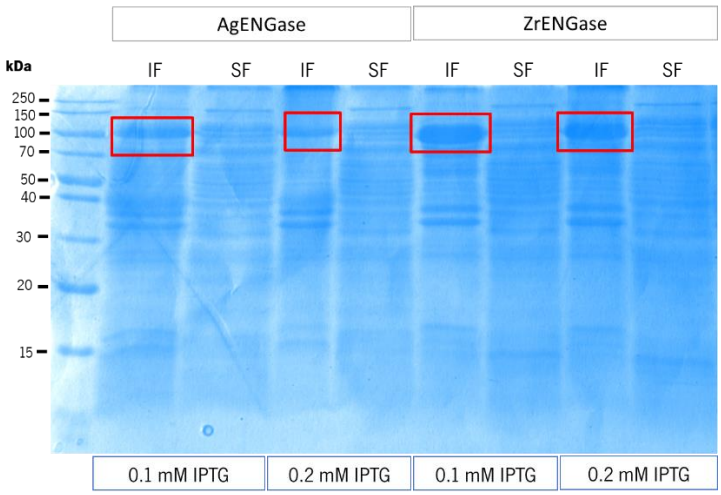


Figure 21. Recombinant production of AgENGase and ZrENGase in *E. coli* Origami 2 (DE3) in LB medium at 18 ℃. Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes.

It was verified that there are no significative differences between using 0.1 mM and 0.2 mM of IPTG at 18 ℃ (**Figure 21**). Comparing to the production of both ENGases in *E. coli* Origami 2 (DE3) at

20 ℃, it was verified that at 18 ℃ the total protein concentration is lower in LB medium (**Figure 19** and **21**). Despite the optimizations performed, the proteins of interest continued to be produced in an insoluble form (red boxes in **Figure 21**). So, the slight reduction in temperature may have not been enough for reducing the formation of inclusion bodies.

Considering that the putative ENGases were still being produced in inclusion bodies, induction temperature was again reduced for 16 ℃ and the induction was performed with 0.2 mM IPTG, at the early-log phase (OD_{600} 0.2-0.3) for 24 h[128]. Both ENGases were produced in *E. coli* Origami 2 (DE3) under these conditions (**Figure 22**). Also, ZrENGase was produced in *E. coli* Origami 2 (DE3) under the same conditions, with the difference that before induction it was added to the medium 1 M sorbitol, 0.8 M NaCl and 0.05 M sodium phosphate buffer pH 7.4 (**Figure 22**). Sorbitol has been proved to improve solubility of recombinant proteins[166–16157–159]. NaCl is responsible for maintaining the ionic strength of the media, also resulting in an increase of protein solubility[164,16155,160]. Sodium phosphate is responsible for stabilizing the produced proteins[170].

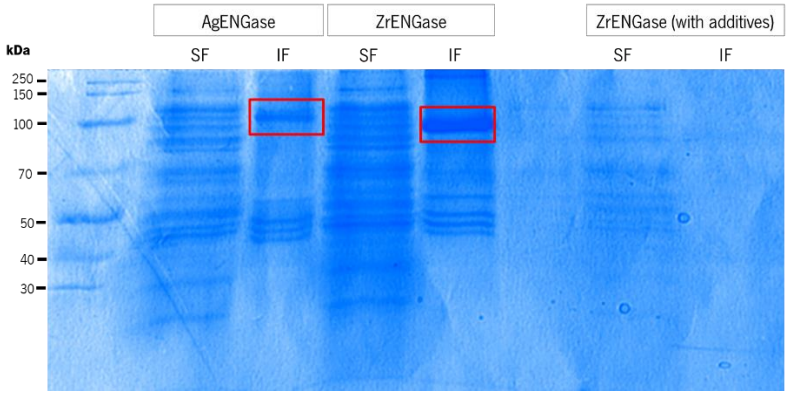


Figure 22. Recombinant production of AgENGase and ZrENGase in *E. coli* Origami 2 (DE3) in LB medium at 16 ℃ with and without additives (1 M sorbitol, 0.8 M NaCl, 0.05 M sodium phosphate buffer pH 7.4). Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes.

Analysing the obtained gel, it was verified that the putative AgENGase and ZrENGase (without additives in the medium) are still being produced in the insoluble fraction (red boxes in **Figure 22**). The addition of 1 M sorbitol and 0.8 M NaCl did not favour the production of the recombinant ZrENGase (**Figure 22**). The further reduction in temperature may have not been enough for reducing the formation of inclusion bodies, maybe due to inefficient lysis. As such, a dramatic reduction in the induction temperature was performed, since it was reported to favour soluble production of recombinant proteins in *E. coli* [128]. Both ENGases were produced in *E. coli* Origami 2 (DE3) at 4 ℃ during 72 h and induction

was performed at the early-log phase (**Appendix K**). The production of both ENGases was not favoured under these conditions comparing to the production at 16 ℃ (**Figure 22**) and both ENGases were still being produced in the insoluble fraction.

Optimization of lysis conditions was once again performed with different additives: 1 M sucrose and a combination of 1 M sucrose and 0.2 M arginine. Several studies have proved that arginine has a positive effect in the solubilization of proteins and prevention of aggregation[171,17162,163]. Sucrose may probably be responsible for enhancing protein stability[164]. A pellet collected from the production of the putative ZrENGase in *E. coli* Origami 2 (DE3) at 4 ℃ in LB medium was lysed with the referred additives, but the protein of interest remained in the insoluble fraction (red boxes in **Figure 23**). This result indicates that these lysis additives did not have a positive effect on the solubilization of inclusion bodies in the case of the putative ZrENGase.

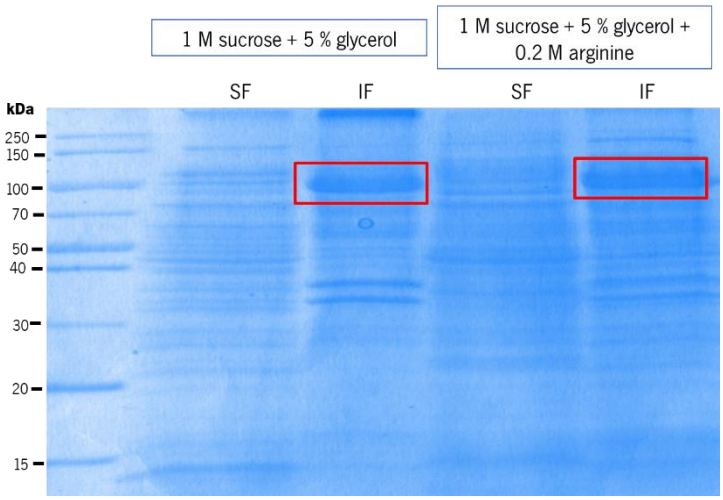


Figure 23. Lysis of a pellet collected from the production of the putative ZrENGase in *E. coli* Origami 2 (DE3) at 4 ℃ in LB medium with 1 M sucrose or a combination of 1 M sucrose and 0.2 M arginine. Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusion TrxA-His-tag-TEV-ZrENGase (88 kDa) are in the red boxes.

Since proteins were still present in inclusion bodies (red boxes in **Figure 23**), the use of low concentration of a denaturant in the lysis buffer, 2 M urea, was further tested. The use of a urea concentration lower than 6-8 M has proved to solve this issue in the case of other insoluble proteins[173]. A pellet collected from the production of the putative ZrENGase in *E. coli* Origami 2 (DE3) at 16 ℃ in LB medium and of the putative AgENGase in *E. coli* Origami 2 (DE3) at 4 ℃ in LB medium was lysed under these mild denaturation conditions. After 14 h of incubation with 2 M urea, soluble and insoluble fractions were analysed in gel (**Figure 24**).

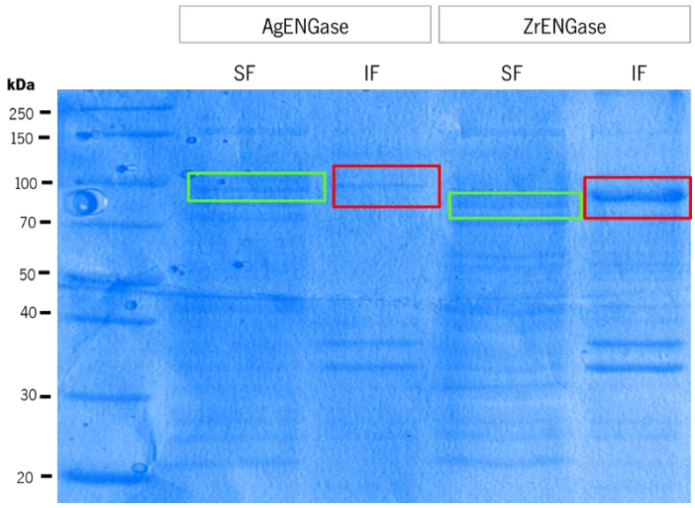


Figure 24. Lysis of pellets collected from the production of the putative ZrENGase in *E. coli* Origami 2 (DE3) at 16 ℃ in LB medium and of the putative AgENGase in *E. coli* Origami 2 (DE3) at 4 ℃ in LB medium in the presence of 2 M urea. Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) in the insoluble fraction are in the red boxes and in the soluble fraction are in green boxes.

It was verified that the addition of 2 M urea to the lysis of the pellets tested slightly improved the solubility of the putative ENGases in relation to the addition of 1 M sucrose and 0.2 M arginine (**Figure 23 and 24**). The results allow one to infer that these conditions did not allow the total solubilization of the inclusion bodies where putative ENGases were being produced.

Since no desirable results were being obtained from the addition of lysis additives, optimization of culture conditions was again tried. Considering the positive effects in protein solubility and stabilization of the addition of NaCl, glycerol and arginine to the culture medium, ZrENGase was produced in *E. coli* Origami 2 (DE3), BL21 (DE3) and SHuffle T7 Express at 16 ℃ in LB medium supplemented or not with 0.2 M NaCl, 0.5 % (w/v) glycerol and 0.2 M arginine. Expression was induced with 0.2 mM IPTG at OD_{600} between 0.2 to 0.3. After 24 h of induction, soluble and insoluble fractions were analysed in gel (**Figure 25**).

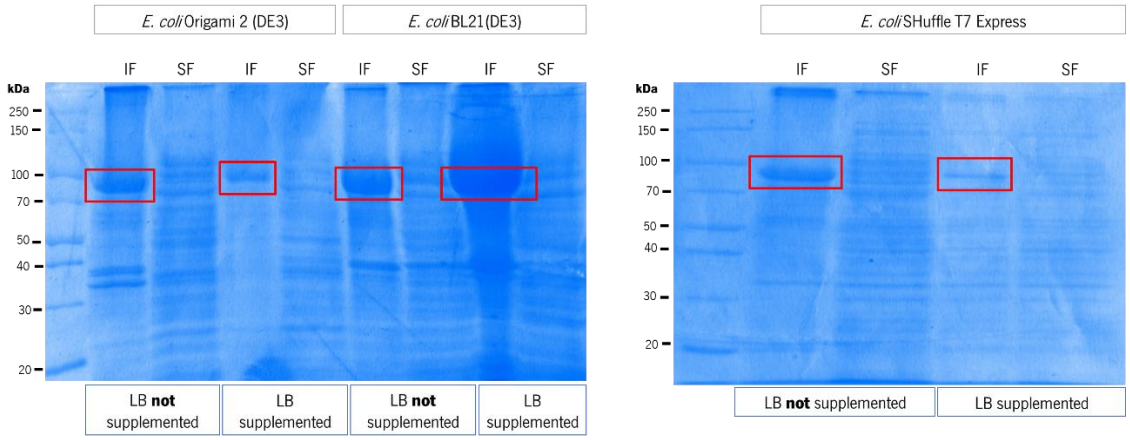


Figure 25. Recombinant production of ZrENGase in *E. coli* Origami 2 (DE3), BL21 (DE3) and SHuffle T7 Express in LB medium supplemented or not with 0.2 M NaCl, 0.5 % (w/v) glycerol and 0.2 M arginine at 16 ℃. Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes.

It was verified that when the putative ZrENGase is produced by *E. coli* Origami 2 (DE3) and BL21 (DE3), the supplementation of the LB medium did not favour the solubilization of the putative ENGase compared with non-supplemented LB medium, since the apparent relative quantity of the protein in the insoluble fraction when the medium is supplemented is higher than when the medium is not supplemented (**Figure 25**).

It was also noticed that there is no apparent difference between expressing the putative ZrENGase in *E. coli* BL21 (DE3) at 20 ℃ and 16 ℃ using the same inducer concentration (**Figure 20** and **25**). On the other hand, the expression of the putative ZrENGase in *E. coli* SHuffle T7 Express at 16 ℃ is favoured in relation to the expression at 20 ℃ (**Figure 18** and **25**).

A last attempt was to recover putative ENGases from the pellets of all the productions in *E. coli* Origami 2 (DE3) referred before, with exception of the ones at 37 ℃. All the pellets were lysed with lysis buffer supplemented with 2 M urea. All soluble fractions obtained were purified in a 5 mL IMAC column and equilibration and elution buffers were also supplemented with 2 M urea. Samples corresponding to the fraction that did not bind to the column, last washing step and fractions where proteins were eluted were analysed in gel (**Figure 26**).

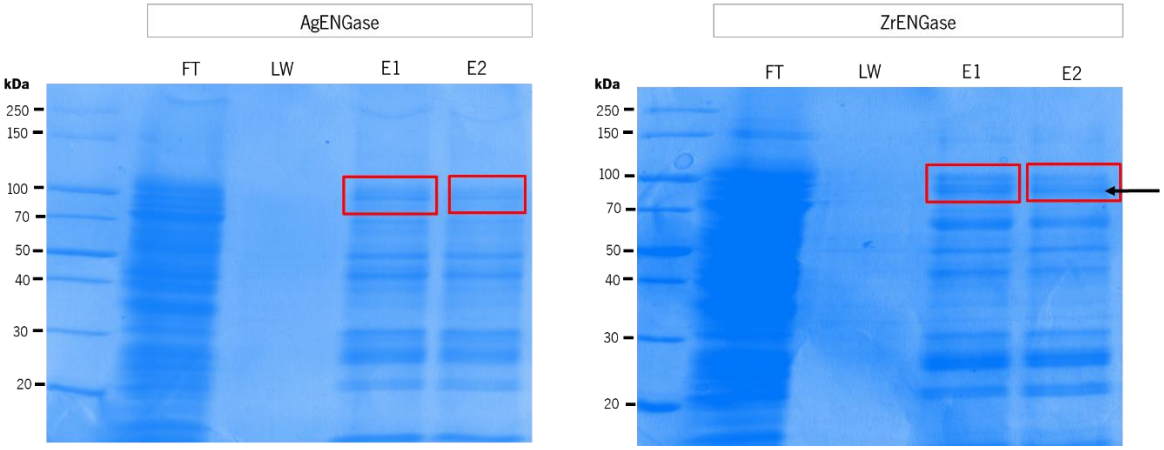


Figure 26. Column purification of the soluble protein fractions from productions in *E. coli* Origami 2 (DE3) at 20, 18, 16 and 4 ℃ of the putative ENGases. Fraction that did not bind to the column (FT), fraction from the last washing step (LW) and fractions were proteins were eluted (E1 and E2) were resolved in a 15 % acrylamide gel and stained with. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes. Black arrow is pointing to the band that may correspond to the putative ZrENGase.

According to the results obtained, it seems that both putative ENGases were purified under the conditions used (**Figure 26**). The apparent MW determined using MW Analysis Tool of Image Lab version 3.0 for the putative AgENGase was 103 kDa and for the putative ZrENGase was 96 kDa, which is close to the bioinformatically predicted MW for both fusion proteins.

The putative AgENGase was eluted after 1.2 column volume (CV) and the putative ZrENGase was eluted after 1 CV.

Analysing this data, it is possible to conclude that supplementation of lysis buffer with 2 M urea improved protein solubility.

Although it seemed that the putative ENGases were present in the elution fractions resulted from this purification, no ENGase activity was detected in gel, following the procedure described in section 4 of Materials and Methods.

Gathering all the results related to the analysis of the solubility of the putative ENGases, it is clear that further improvements in production, lysis and purification conditions will be needed.

3. Optimization of electrophoresis- and HPLC-based methodologies to analyse ENGase deglycosylation activity

In order to detect ENGase deglycosylation activity the amount of each substrate needed for the reactions was evaluated. 5 and 2.5 µg of RNase B from bovine pancreas, 2, 4, 8 and 38 µg of β-galactosidase from *A. niger* and 0.5 and 1 mg of invertase from *S. cerevisiae* were tested for performing the reaction with the model ENGase commercially available, Endo-H. It was verified that 5 µg of RNase B were sufficient to perform the reactions, while in the case of β-galactosidase, 4 µg of this substrate were adequate for the reaction tests (data not shown). In the case of invertase, it was verified that 0.5 mg is enough for ENGase activity detection in gel. After incubation of the substrates overnight at 37 ºC with Endo-H, samples were tested in gel (**Figure 27**).

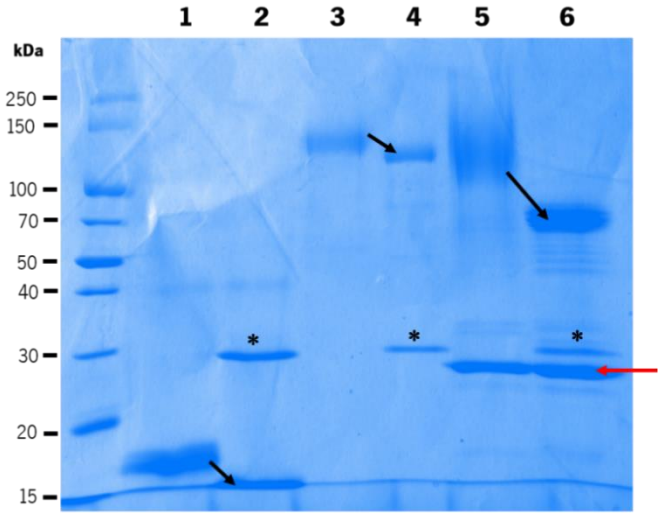


Figure 27. Gel electrophoresis detection of Endo-H ENGase deglycosylation activity using different substrates. Band shifts resulting from the reaction of the referred substrates with a model ENGase, Endo-H, were detected in a 15 % acrylamide gel. **1-** RNase B from bovine pancreas (negative control), **2-** RNase B deglycosylated with Endo-H, **3-** β-galactosidase from *A. niger* (negative control), **4-** β-galactosidase deglycosylated with Endo-H, **5-** invertase from *S. cerevisiae* (negative control), **6-** invertase deglycosylated with Endo-H. The black arrows indicate the deglycosylation shift of each substrate, the red arrow is pointing to possible contaminants of the β-galactosidase substrate solution and the asterisk represents the band corresponding to Endo-H.

RNase B is a glycoprotein containing a single *N*-linked glycosylation site and is composed of oligomannose *N*-glycans. The apparent molecular MW determined using MW Analysis Tool of Image Lab version 3.0 (**Figure 27**) for RNase B was around 14.7 kDa. After deglycosylation, the apparent MW was reduced to 13.1 kDa. β-galactosidase from *A. niger* is also a glycoprotein composed of oligomannose *N*-glycans and the apparent MW determined in gel (**Figure 27**) was 133 kDa. After deglycosylation, the

apparent MW was reduced to 121 kDa. Regarding invertase from *S. cerevisiae*, it is a glycoprotein composed of high-mannose type *N*-glycans and the apparent MW determined in gel (**Figure 27**) was about 109 kDa. After deglycoylation, the apparent MW was reduced to 75.4 kDa. The determined values were in accordance with the predicted by the manufacturers of the substrates (Sigma-Aldrich and Maxilact).

It was verified that Endo-H can hydrolyse the glycans present in the tested substrates (**Figure 27**). Further optimization of the amount of the putative ENGases that should be used for ENGase activity detection in gel will be necessary.

Since HPLC is a chromatographic technique widely used for the detection of carbohydrates and is more sensitive than SDS-PAGE electrophoresis, reactions of Endo-H with 120 µg β-galactosidase from *A. niger* and 15 mg invertase from *S. cerevisiae* were analysed by HPLC as described in section 5.1 of Materials and Methods. The detection of a peak in the deglycosylated sample with a different area from that observed in the negative control (glycosylated substrate) at the same retention time should correspond to the *N*-glycans released from the substrate. The chromatograms obtained from the analysis of the reaction of β-galactosidase with Endo-H as well as the negative control is represented in **Figure 28**.

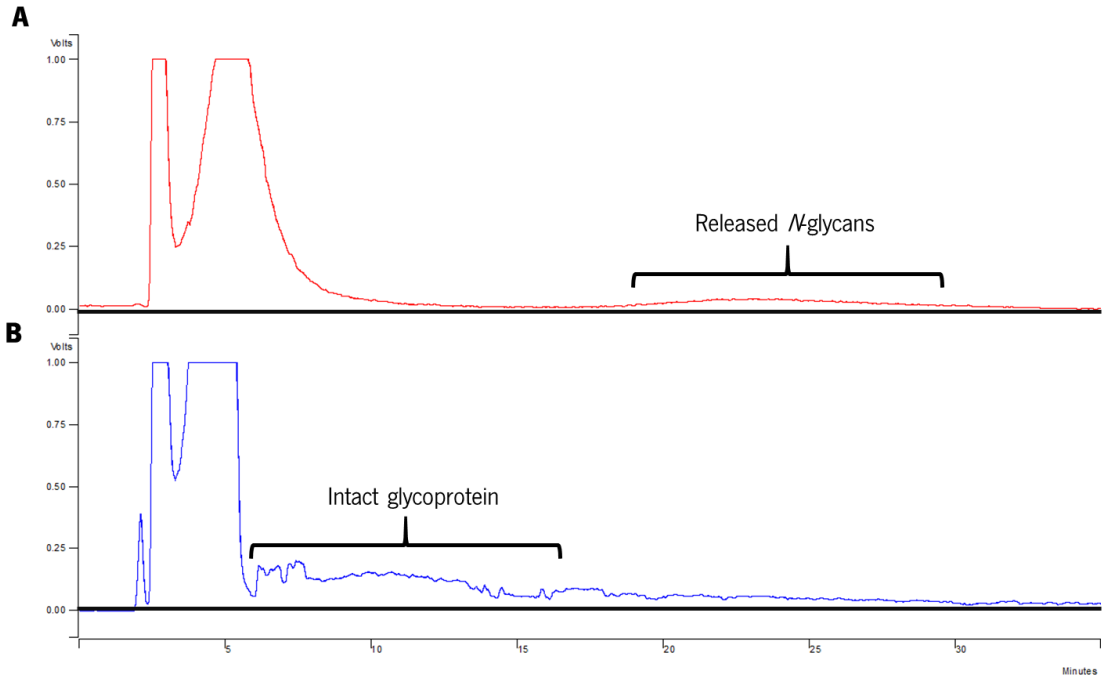


Figure 28. Chromatogram obtained from the analysis of the reaction of β-galactosidase with Endo-H (A) as well as the native control (B) by HPLC. Deglycosylation of glycoproteins was monitored by HPLC using a Jasco chromatograph equipped with an ELSD detector (SEDEX 85, SEDERE) and a Prevail Carbohydrate ES column (5 µm, 250 × 4.6 mm, Altech). A mixture of acetonitrile-water (60:40 % (v/v)), pumped at 0.9 mL/min, was used as mobile phase. The injection volume was defined as 20 µL.

As can been seen in the chromatogram, a peak with retention time of around 23 min that does not exist in the negative control (without enzyme) can be observed in the sample treated with enzyme, indicating the presence of *N*-glycans released from β-galactosidase (**Figure 28**).

The chromatograms obtained from the analysis of the reaction of invertase with Endo-H as well as the native control is represented in **Figure 29.**

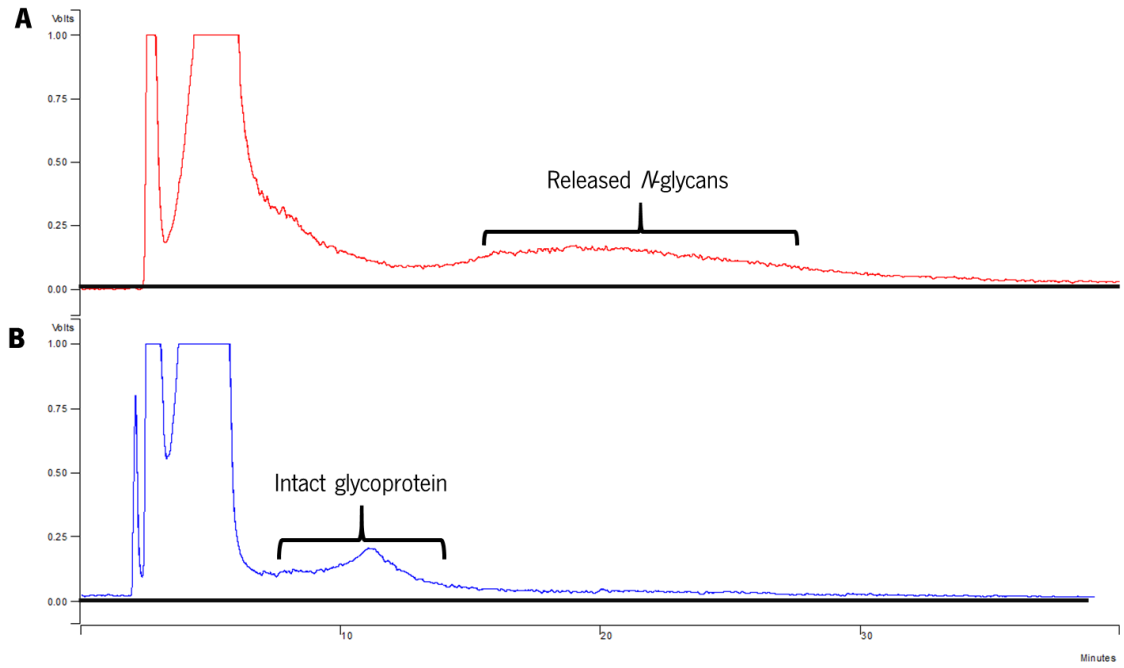


Figure 29. Chromatogram obtained from the analysis of the reactions of invertase with Endo-H (A) as well as the negative control (B) by HPLC. Deglycosylation of glycoproteins was monitored by HPLC using a Jasco chromatograph equipped with an ELSD detector (SEDEX 85, SEDERE) and a Prevail Carbohydrate ES column (5 µm, 250 × 4.6 mm, Altech). A mixture of acetonitrile-water (60:40 % (v/v)), pumped at 0.9 mL/min, was used as mobile phase. The injection volume was defined as 20 µL.

Once again, a peak with retention time of around 20 min that does not appear in the negative control (without enzyme) is detected in the sample treated with enzyme, indicating the presence of *N*-glycans released from invertase (**Figure 29**).

Considering these results, the protocol developed for the detection of the presence of ENGase hydrolytic activity by HPLC is working and ready to be used in future analyses of the activity of purified recombinant ENGases.

4. Analysis of intra and extracellular ENGase/chitinase activity in *A. gossypii*

A first attempt was made to assess endogenous ENGase activity in crude cell extracts prepared from *A. gossypii* and *Z. rouxii*, according to the protocol described by Murakami *et al.*[51] The commercially available ENGase Endo-H was used as a positive control. Commercially obtained β-galactosidase from *A. niger* was used as a substrate and the reactions were incubated overnight at 37 ℃. β-galactosidase zymograms were performed to detect a shift in the β-galactosidase MW in NATIVE-PAGE gels after incubation of this glycoprotein with the cellular extracts. This electrophoretic technique was used to allow the detection of ENGase activity without the need for protein purification steps. Although different substrate concentrations, different volumes of cellular extracts and several cellular disruption protocols were tested, no band shift was detected when β-galactosidase was incubated with crude cell extracts from *A. gossypii* and *Z. rouxii*. An example of those results is represented in **Appendix L**.

Since ENGase activity had already been detected in *Z. rouxii* crude cell extracts[52], subsequent attempts were made to detect ENGase activity in *A. gossypii* only. For that, the fungus was grown in defined minimal medium and, after 24 h, the culture supernatants were separated from the mycelia and each were treated as described in section 5.2 of Materials and Methods. The total protein content in the concentrated culture supernatants (extracellular) and crude cell extracts (intracellular) was then determined by the Bradford method (**Table 8**).

Table 8. Total protein concentration determined for intracellular content and concentrated culture supernatants.

Fraction	Total protein concentration (mg/mL x 10^{-4})	Standard deviation (x 10^{-4})
Intracellular content	4.49	1.22
Concentrated culture supernatants	2.55	1.03

The ENGase/chitinase activity was assessed using a commercial Chitinase Assay Kit that provides three substrates, but only two of them were tested in this work: 4-Methylumbelliferyl *N*-acetyl-β-D-glucosaminide and 4-Methylumbelliferyl β-D-*N,N'*-diacetylchitobioside hydrate (**Figure 30**). The second one allows the determination of chitobiosidase activity, while the first allows the determination of β-*N*-acetylglucosaminidase activity, since it possess a *N-N'*-diacetylchitobiose core upon which ENGases can act.

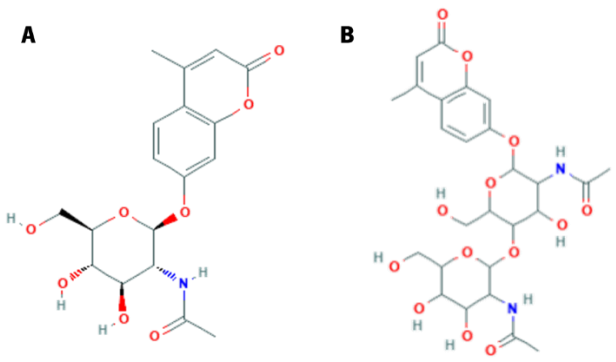


Figure 30. Substrates provided in the Chitinase Assay Kit used to detect β-*N*-acetylglucosaminidase/chitobiosidase activity. A- 4-Methylumbelliferyl *N*-acetyl-β-D-glucosaminide, B- 4-Methylumbelliferyl β-D-*N,N'*-diacetylchitobioside hydrate. Figure adapted from Pubchem.

Intra and extracellular ENGase/chitinase activity was determined from *A. gosspii* crude cell extracts and concentrated culture supernatants, respectively, using 4-Methylumbelliferyl *N*-acetyl-β-D-glucosaminide (A) and 4-Methylumbelliferyl β-D-*N,N'*-diacetylchitobioside hydrate (B) as substrates. The specific ENGase/chitinase activity calculated for each sample is represented in **Figure 31**.

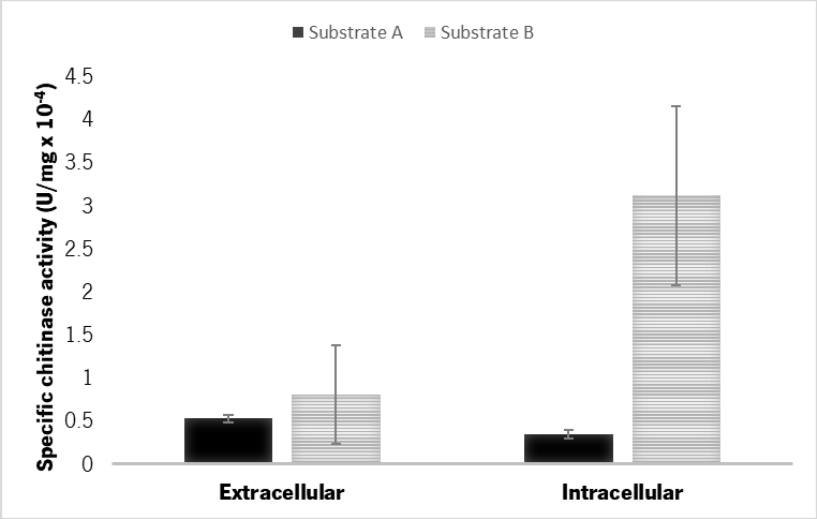


Figure 31. Specific ENGase/chitinase activity in *A. gosspii* crude cell extracts (intracellular) and concentrated culture supernatants (extracellular). Specific enzymatic activities were normalized per amount of total protein in the reaction mix. Error bars represent the standard deviation from two biological replicates.

Specific ENGase/chitinase activity detected extracellularly was only slightly different between substrates A and B (**Figure 31**). The only chitinase produced by *A. gossypii*, AgCts2 (UniProt Accession Number: Q75CT5)[165], is predicted to be located extracellularly, while the only putative AgENGase is predicted to be intracellular. Therefore, AgCts2 should be the main contributor for the enzymatic activity detected extracellularly. On the other hand, specific ENGase/chitinase activity detected intracellularly was 9 times higher towards substrate B than substrate A (**Figure 31**). This result could indicate that a putative

AgENGase able to act on the *N*-*N*-diacetylchitobiose core present in the structure of substrate B is produced intracellularly, thus contributing to the higher activity detected intracellularly towards this substrate. However, this hypothesis does not exclude the fact that other enzymes could be acting on the referred core, namely AgCts2 that could have been retained in the crude cell extract. Therefore, the future characterization of the ENGase/chitinase activity of the *A. gossypii afr597w* mutant generated within the scope of this work (see sections 3.2 and 5) will help to clarify this hypothesis.

5. Disruption of the *AFR597W* gene in *A. gossypii* and characterization of the generated mutant

Gene disruption technologies have been widely used to help scientists understand the physiological role of a given protein of interest[83]. Among these, gene disruption by homologous recombination coupled to the Cre-*loxP* technology has been used for the deletion of several genes of interest in the filamentous fungus *A. gossypii*[140,166]. As such, the same technology was used in this work to disrupt the sequence of the gene coding for the putative AgENGase (*AFR597W*) in *A. gossypii*. For the deletion, marker cassette *lox*P-*GEN3*-*lox*P was used to replace the gene of interest, as schematized in **Figure 32**. The chosen deletion cassette provides an easy way to select the generated mutants, which will be resistant to the antibiotic geneticin, and may be subsequently removed using Cre-*loxP* technology[140].

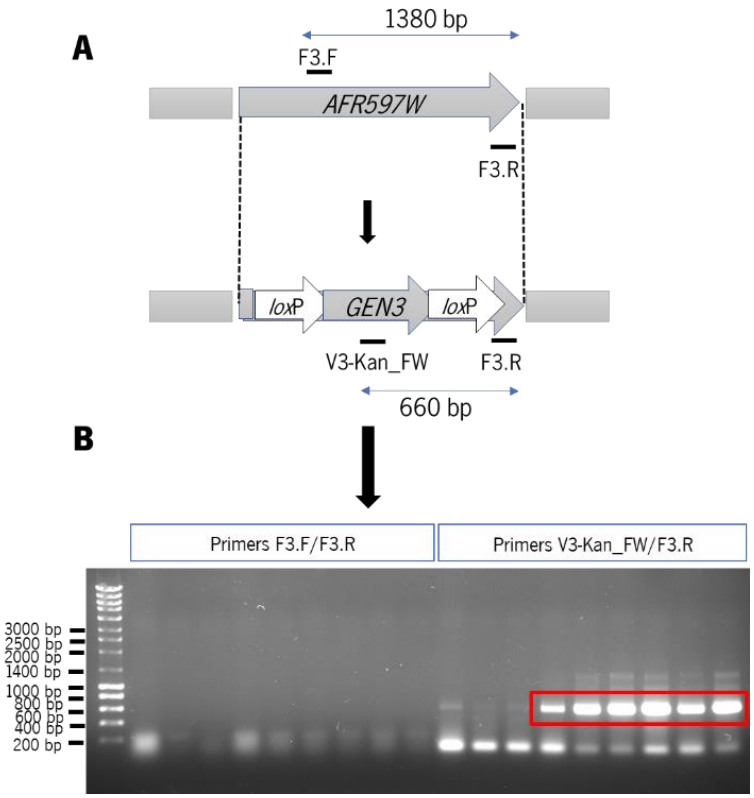


Figure 32. Schematic representation of the strategy used for the disruption of *AFR597W* in *A. gossypii*, and verification PCR of the generated mutants. (A) Construction of the *A. gossypii* mutant strain. The *AFR597W* gene was replaced by the *lox*P-*GEN3*-*lox*P cassette. **(B)** Verification of the integration of the *lox*P-*GEN3*-*lox*P cassette in the *AFR597W* locus and confirmation of its excision by colony PCR with primers F3.F/F3.R and V3-Kan_FW/F3.R.

From the nine transformants obtained, six were undoubtedly confirmed to containing nuclei with the *AFR597W* disrupted (**Figure 32**). After several trials of sporulation steps, it was verified that the generated *afr597w* mutants were not able to sporulate, indicating that this gene may be essential for this physiological process in *A. gossypii*. Despite the sporulation deficiency, no macroscopic differences were observed in the morphology of this mutant compared to the parent strain. However, the *afr597w* mutant was found to be more susceptible to lysis with the Lysing Enzymes used than the parent strain. This observation may indicate that the putative ENGase may affect the *N*-glycans present in the fungus cell wall.

Expression data determined by Aguiar *et al.*[145] (available in supplementary material) revealed that the expression of *AFR597W* remained constant in the different growth phases of the fungus in defined minimal medium, but in the late stationary phase, expression started to slightly increase, indicating that this gene may be important for sporulation. In a different study, Wasserstrom *et al.*[97] (supplementary material) verified that the expression of the *AFR597W* is 7.37 times higher when *A. gossypii* is grown in sporulation media than in rich medium (AFM). These data strongly support the hypothesis that this gene is important for sporulation in *A. gossypii*, as other ENGases have already been reported to be involved

in the sporulation process in some myxobacteria[167] and filamentous fungi. Namely, the deletion of the gene that codes for an ENGase of the GH18 family in the filamentous fungus *Trichoderma atroviride* was reported to affect sporulation, but in that case, it resulted in increased conidiation rate[168]. Although a decrease in growth rate was observed in *T. atroviride* after deletion of the gene that codes for an ENGase of the GH18 family, in the case of *A. gossypii*, no significant differences were observed in colony radial growth when the *afr597w* mutant strain was grown for 4 days compared with a control strain that also has resistance to G418 (**Figure 33**).

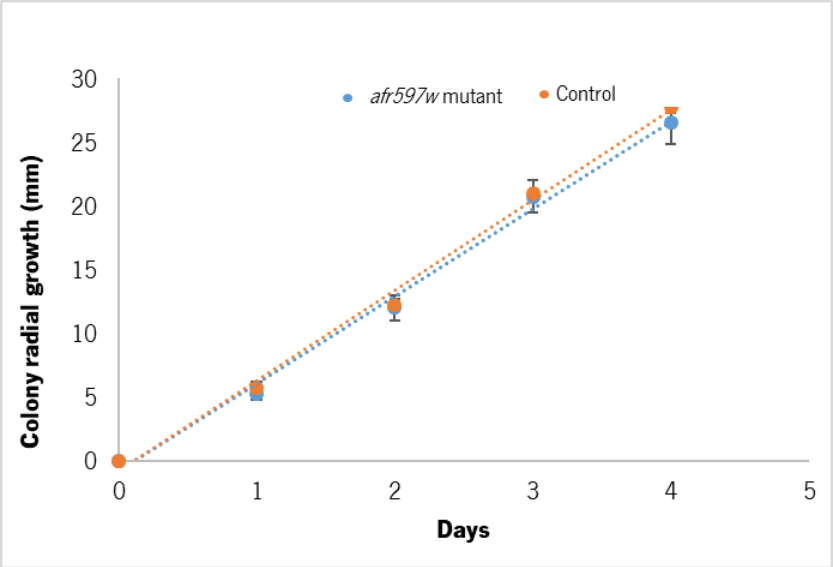


Figure 33. Colony radial growth measured in *A. gossypii afr597w* mutant strain and another *A. gossypii* mutant strain that also has resistance to G418 (control) after 4 days of growth on selective agar-solidified AFM. The values represent the mean of two biological replicas. Error bars represent the standard deviation of the biological replicas.

CONCLUSION

CONCLUSION

Given the range of pharmaceutical and biomedical applications of the ENGases, the interest in characterizing and producing these enzymes has been expanded. Bioinformatic approaches and recombinant production have played an important role in the identification and characterization of the activity of many ENGases, however, to understand the physiological role of these enzymes more functional studies are needed.

ENGase activity had been previously detected in *Z. rouxii* crude cell extracts and ENGase/chitinase activity was also detected here in *A. gossypii* crude extracts. Despite the optimizations made to the recombinant production of the putative proteins responsible for these activities (ZrENGase and AgENGase, respectively) in *E. coli*, obstacles were found related to their soluble production, which hampered further *in vitro* characterization of these enzymes. Therefore, improvements in their production and purification will still be required to generate enough protein for activity studies and eventual applications. In this context, the use of a different solubility fusion tag, the fusion of the putative ENGases with a signal sequence to translocate them to the periplasm of *E. coli* or even the expression in a eukaryotic host, such as *Pichia pastoris*, may be possible alternative strategies for solving production and purification issues. Mechanical methods for cell lysis, such as sonication, could be used alternatively and/or together with enzymatic lysis for further improving the recovery of soluble protein. Purification from inclusion bodies of the recombinant ENGases, using higher concentrations of urea and/or guanidine-HCl than those used here, may be a valid alternative for solving the issues faced during this work. Also, optimization of the washing step of column purification will be necessary.

The lack of sporulation capacity observed when the *AFR597W* NOHBY was disrupted in *A. gossypii* indicate that this gene may be essential for processes involved in the fungus lifecycle and sheds some light into its functionality. Further analysis of intracellular and extracellular ENGase activity as well as physiological characterization of the mutant strain will be necessary for better understanding the biological function of this uncharacterized enzyme.

In conclusion, although the extensive optimisation steps of the recombinant production and purification of putative ENGases have not allowed further activity tests, the obtained results are encouraging and represent a significant accomplishment that will be the basis for a more detailed investigation in the future.

REFERENCES

REFERENCES

1. Mulloy, B., Dell, A., Stanley, P. & Prestegard, J. Structural Analysis of Glycans. in *Essentials of Glycobiology* (eds. Varki, A., Cummings, R., Esko, J., Stanley, P., Hart, G., Aebi, M., Darvill, A., Kinoshita, T., Packer, N., Prestegard, J., Schnaar, R. & Seeberger, P.) 12,13 (Cold Spring Harbor Laboratory Press, 2017).
2. Varki, A. & Kornfeld, S. Historical Background and Overview. in *Essentials of Glycobiologyals* (eds. Varki, A., Cummings, R., Esko, J., Stanley, P., Hart, G., Aebi, M., Darvill, A., Kinoshita, T., Packer, N., Prestegard, J., Schnaar, R. & Seeberger, P.) 1 (Cold Spring Harbor Laboratory Press, 2017).
3. Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3,** 97–130 (1993).
4. Dwek, R. Glycobiology: Toward Understanding the Function of Sugars. *Chem. Rev.* **96,** 683–720 (1996).
5. Dube, D. & Bertozzi, C. Glycans in cancer and inflammation - Potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **4,** 477–488 (2005).
6. Helenius, A. & Markus, A. Intracellular Functions of *N*-Linked Glycans. *Science (80-.).* **291,** 2364–2369 (2001).
7. Live, D., Kumar, R., Beebe, X. & Danishefsky, S. Conformational influences of glycosylation of a peptide: a possible model for the effect of glycosylation on the rate of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 12759–12761 (1996).
8. van Zuylen, C., Kamerling, J. & Vliegenthart, J. Glycosylation beyond the Asn78-linked GlcNAc residue has a significant enhancing effect on the stability of the alpha subunit of human chorionic gonadotropin. *Biochem Biophys Res Commun* **232,** 117–120 (1997).
9. Arar, C., Carpentier, V., Le Caer, J., Monsigny, M., Legrand, A. & Roche, A. ERGIC-53, a membrane protein of the endoplasmic reticulum-Golgi intermediate compartment, is identical to MR60, an intracellular mannose- specific lectin of myelomonocytic cells. *Journal of Biological Chemistry* **270,** 3551–3553 (1995).
10. Spiro, R. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12,** 43R-56R (2002).
11. Springer, S. & Gagneux, P. Glycan evolution in response to collaboration, conflict, and constraint. *J. Biol. Chem.* **288,** 6904–6911 (2013).
12. Varki, A. *et al.* Symbol Nomenclature for Graphical Representation of Glycans. *Glycobiology* **25,** 1323–1324 (2015).
13. Zacchi, L. & Schulz, B. *N*-glycoprotein macroheterogeneity: biological implications and proteomic characterization. *Glycoconj. J.* **33,** 359–376 (2015).
14. Stanley, P., Schachter, H. & Taniguchi, N. *N*-gycans. in *Essentials of Glycobiology* (eds. Varki, A., Cummings, R., Esko, J., Stanley, P., Hart, G., Aebi, M., Darvill, A., Kinoshita, T., Packer, N., Prestegard, J., Schnaar, R. & Seeberger, P.) 5–10 (Cold Spring Harbor Laboratory Press, 2017).
15. Archer, D. & Peberdy, J. The molecular biology of secreted enzyme production by fungi. *Crit. Rev. Biotechnol.* **17,** 273–306 (1997).
16. Maras, M., Van Die, I., Contreras, R. & van den Hondel, C. Filamentous fungi as production

organisms for glycoproteins of bio-medical interest. *Glycoconj. J.* **16,** 99–107 (1999).

17. Gemmill, T. & Trimble, R. Overview of *N*- and *O*-linked oligosaccharide structures found in various yeast species. *Biochim. Biophys. Acta - Gen. Subj.* **1426,** 227–237 (1999).

18. Aguiar, T., Maaheimo, H., Heiskanen, A., Wiebe, M., Penttilä, M. & Domingues, L. Characterization of the *Ashbya gossypii* secreted *N*-glycome and genomic insights into its *N*-glycosylation pathway. *Carbohydr. Res.* **381,** 19–27 (2013).

19. Gagneux, P., Aebi, M. & Varki, A. Evolution of Glycan Diversity. in *Essentials of Glycobiology* 4–5 (Cold Spring Harbor Laboratory Press, 2017).

20. Tzelepis, G., Karlsson, M. & Suzuki, T. Deglycosylating enzymes acting on *N*-glycans in fungi: Insights from a genome survey. *Biochim. Biophys. Acta - Gen. Subj.* **1861,** 2551–2558 (2017).

21. Caramelo, J. & Parodi, A. A sweet code for glycoprotein folding. *FEBS Lett.* **589,** 3379–3387 (2015).

22. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **4,** 181–191 (2003).

23. Zacchi, L., Caramelo, J., McCracken, A. & Brodsky, J. Endoplasmic Reticulum-Associated Degradation and Protein Quality Control. *Encyclopedia of Cell Biology* **1,** (Elsevier Ltd., 2016).

24. Parodi, A. Role of *N*-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem J* **348 Pt 1,** 1–13 (2000).

25. Stolz, J. & Munro, S. The Components of the *Saccharomyces cerevisiae* Mannosyltransferase Complex M-Pol I Have Distinct Functions in Mannan Synthesis. *J. Biol. Chem.* **277,** 44801–44808 (2002).

26. Okamoto, R., Izumi, M. & Kajihara, Y. Decoration of proteins with sugar chains: recent advances in glycoprotein synthesis. *Curr. Opin. Chem. Biol.* **22,** 92–99 (2014).

27. Seitz, O. Glycopeptide synthesis and the effects of glycosylation on protein structure and activity. *Chembiochem* **1,** 214–46 (2000).

28. Wang, L. & Amin, M. Chemical and chemoenzymatic synthesis of glycoproteins for deciphering functions. *Chem. Biol.* **21,** 51–66 (2014).

29. Tejada, A., Brailsford, J., Zhang, Q., Shieh, J., Moore, M. & Danishefsky, S. Total Synthesis of Glycosylated Proteins. *Top Curr Chem* **362,** 1–26 (2015).

30. Kajihara, Y., Yamamoto, N., Okamoto, R., Hirano, K. & Murase, T. Chemical Synthesis of Homogeneous Glycopeptides and Glycoproteins. *Chem. Rec.* **10,** 80–100 (2010).

31. Yamamoto, N., Tanabe, Y., Okamoto, R., Dawson, P. & Kajihara, Y. Chemical synthesis of a glycoprotein having an intact human complex-type sialyloligosaccharide under the Boc and Fmoc synthetic strategies. *J. Am. Chem. Soc.* **130,** 501–510 (2008).

32. Chalker, J., Bernardes, G. & Davis, B. A 'tag-and-modify' approach to site-selective protein modification. *Acc. Chem. Res.* **44,** 730–741 (2011).

33. Wang, L. & Lomino, J. Emerging technologies for making glycan-defined glycoproteins. *Am. Chem. Soc.* **7,** 110–122 (2012).

34. Yu, H., Chokhawala, H. ., Huang, S. & Chen, X. One-pot three-enzyme chemoenzymatic approach to the synthesis of sialosides containing natural and non-natural functionalities. *Nat. Protoc.* **1,**

2485–2492 (2006).

35. Li, B., Zeng, Y., Hauser, S., Song, H. & Wang, L. Highly efficient endoglycosidase-catalyzed synthesis of glycopeptides using oligosaccharide oxazolines as donor substrates. *J. Am. Chem. Soc.* **127,** 9692–9693 (2005).

36. Witte, K., Sears, P., Martin, R. & Wong, C. Enzymatic glycoprotein synthesis: Preparation of ribonuclease glycoforms via enzymatic glycopeptide condensation and glycosylation. *J. Am. Chem. Soc.* **119,** 2114–2118 (1997).

37. Wang, L. Chemoenzymatic synthesis of glycopeptides and glycoproteins through endoglycosidase-catalyzed transglycosylation. *Carbohydr. Res.* **343,** 1509–1522 (2008).

38. Lomino, J., Naegeli, A., Orwenyo, J., Amin, M., Aebi, M. & Wang, L. A two-step enzymatic glycosylation of polypeptides with complex *N*-glycans. *Bioorganic Med. Chem.* **21,** 2262–2270 (2013).

39. Abe, H., Tomimoto, K., Fujita, Y., Iwaki, T., Chiba, Y., Nakayama, K. & Nakajima, Y. Development of *N*- and *O*-linked oligosaccharide engineered *Saccharomyces cerevisiae* strain. *Glycobiology* **26,** 1248–1256 (2016).

40. Kainz, E., Gallmetzer, A., Hatzl, C., Nett, J., Li, H., Schinko, T., Pachlinger, R., Berger, H., Reyes-Dominguez, Y., Bernreiter, A., Gerngross, T., Wildt, S. & Strauss, J. *N*-glycan modification in *Aspergillus* species. *Appl. Environ. Microbiol.* **74,** 1076–1086 (2008).

41. Wacker, M., Linton, D., Hitchen, P., Nita-lazar, M., Haslam, S., North, S., Panico, M., Morris, H., Dell, A., Wren, B. & Aebi, M. *N*-Linked Glycosylation *Campylobacter jejuni* and Its Functional Transfer into *E. coli*. *Science* **298,** 1790–1793 (2002).

42. Hiroko, A., Takaoka, Y., Chiba, Y., Sato, N., Ohgiya, S., Itadani, A., Hirashima, M., Shimoda, C., Jigami, Y. & Nakayama, K. Development of a valuable yeast strain using a novel mutagenesis technique for the effective production of therapeutic glycoproteins. *Glycobiology* **19,** 428–436 (2009).

43. Mrázek, H., Weignerová, L., Bojarová, P., Novák, P., Vaněk, O. & Bezouška, K. Carbohydrate synthesis and biosynthesis technologies for cracking of the glycan code: Recent advances. *Biotechnol. Adv.* **31,** 17–37 (2013).

44. Liu, L., Bennett, C. & Wong, C. Advances in glycoprotein synthesis. *Chem. Commun.* 21–33 (2006).

45. Maley, F., Trimble, R., Tarentino, A. & Plummer, T. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal. Biochem.* **180,** 195–204 (1989).

46. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42,** 490–495 (2014).

47. Suzuki, T. Complex, Two-way Traffic of Molecules Across the Membrane of the Endoplasmic Reticulum. *J. Biol. Chem.* **273,** 10083–10086 (1998).

48. Abbott, D., Macauley, M., Vocadlo, D. & Boraston, A. *Streptococcus pneumoniae* endohexosaminidase D, structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases. *J. Biol. Chem.* **284,** 11676–11689 (2009).

49. Takegawa, K., Nakoshi, M., Iwahara, S., Yamamoto, K. & Tochikura, T. Induction and purification

of endo-β-*N*-acetylglucosaminidase from *Arthrobacter protophormiae* grown in ovalbumin. *Appl. Environ. Microbiol.* **55,** 3107–3112 (1989).

50. Kato, T., Kitamura, K., Maeda, M., Kimura, Y., Katayama, T., Ashida, H. & Yamamoto, K. Free oligosaccharides in the cytosol of *Caenorhabditis elegans* are generated through endoplasmic reticulum-Golgi trafficking. *J. Biol. Chem.* **282,** 22080–22088 (2007).

51. Finn, R., Coggill, P., Eberhardt, R., Eddy, S., Mistry, J., Mitchell, A., Potter, S., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G., Tate, J. & Bateman, A. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44,** D279–D285 (2016).

52. Murakami, S., Takaoka, Y., Ashida, H., Yamamoto, K., Narimatsu, H. & Chiba, Y. Identification and characterization of endo-β-*N*-acetylglucosaminidase from methylotrophic yeast *Ogataea minuta*. *Glycobiology* **23,** 736–744 (2013).

53. Suzuki, T., Yano, K., Sugimoto, S., Kitajima, K., Lennarz, W., Inoue, S., Inoue, Y. & Emori, Y. Endo-beta-*N*-acetylglucosaminidase, an enzyme involved in processing of free oligosaccharides in the cytosol. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 9691–9696 (2002).

54. Kato, T., Fujita, K., Takeuchi, M., Kobayashi, K., Natsuka, S., Ikura, K., Kumagai, H. & Yamamoto, K. Identification of an endo-beta-*N*-acetylglucosaminidase gene in *Caenorhabditis elegans* and its expression in *Escherichia coli*. *Glycobiology* **12,** 581–587 (2002).

55. Kim, Y., Jahren, N., Stone, M., Udeshi, N., Markowski, T., Witthuhn, B., Shabanowitz, J., Hunt, D. & Olszewski, N. Identification and Origin of *N*-Linked-D-*N*-Acetylglucosamine Monosaccharide Modifications on *Arabidopsis* Proteins. *Plant Physiol.* **161,** 455–464 (2013).

56. Muramatsu, H., Tachikui, H., Ushida, H., Song, X., Qiu, Y., Yamamoto, S. & Muramatsu, T. Molecular Cloning and Expression of Endo-beta-*N*-Acetylglucosaminidase D, Which Acts on the Core Structure of Complex Type Asparagine Linked Oligosaccharides. *J. Biochem* **129,** 923–928 (2001).

57. Fujita, K., Takami, H., Yamamoto, K. & Takegawa, K. Characterization of Endo-β-*N*-acetylglucosaminidase from Alkaliphilic *Bacillus halodurans* C-125. *Biosci. Biotechnol. Biochem.* **68,** 1059–1066 (2004).

58. Haneda, K., Inazu, T., Mizuno, M. & Yamamoto, K. Chemoenzymatic Synthesis of Neoglycopeptides Using Endo β-*N*-acetylglucosaminidase from *Mucor hiemalis*. *Methods Enzymol.* **362,** 74–85 (2003).

59. Fan, J., Takegawa, K., Iwahara, S., Kondo, A., Kato, I., Abeygunawardana, C. & Lee, Y. Enhanced Transglycosylation Activity of *Arthrobacter protophormiae* Endo-beta-*N*-acetylglucosaminidase in Media Containing Organic Solvents. *J. Biol. Chem.* **270,** 17723–17729 (1995).

60. Yamamoto, K., Kadowaki, S., Watanabe, J. & Kumagai, H. Transglycosylation activity of *Mucor hiemalis* Endo-β-*N*-Acetylglucosaminidase which transfers complex oligosaccharides to the *N*-acetylglucosamine moieties of peptides. *Biochem. Biophys. Res. Commun.* **203,** 244–252 (1994).

61. Fujita, K., Sato, R., Toma, K., Kitahara, K., Suganuma, T., Yamamoto, K. & Takegawa, K. Identification of the catalytic acid-base residue of *Arthrobacter* endo-β-*N*-acetylglucosaminidase by chemical rescue of an inactive mutant. *J. Biochem.* **142,** 301–306 (2007).

62. Umekawa, M., Huang, W., Li, B., Fujita, K., Ashida, H., Wang, L. & Yamamoto, K. Mutants of *Mucor hiemalis* endo-β-*N*-acetylglucosaminidase show enhanced transglycosylation and

glycosynthase-like activities. *J. Biol. Chem.* **283,** 4469–4479 (2008).

63. Yamamoto, S., Muramatsu, H. & Muramatsu, T. Mutational studies on endo-β-*N*-acetylglucosaminidase D which hydrolyzes core portion of asparagine-linked complex type oligosaccharides. *Glycoconj. J.* **22,** 35–42 (2005).

64. Ling, Z., Suits, M., Bingham, R., Bruce, N., Davies, G., Fairbanks, A., Moir, J. & Taylor, E. The X-ray Crystal Structure of an *Arthrobacter protophormiae* Endo-β-*N*-Acetylglucosaminidase Reveals a $(\beta/\alpha)_8$ Catalytic Domain, Two Ancillary Domains and Active Site Residues Key for Transglycosylation Activity. *J. Mol. Biol.* **389,** 1–9 (2009).

65. Fairbanks, A. The ENGases: versatile biocatalysts for the production of homogeneous *N*-linked glycopeptides and glycoproteins. *Chem. Soc. Rev.* **46,** 5128–5146 (2017).

66. Withers, S. & Williams, S. CAZypedia. *Glycoside hydrolases* Available at: http://www.cazypedia.org/index.php/Glycoside_hydrolases#bibkey_Koshland1953. (Accessed: 28th March 2018)

67. Fujita, K., Kobayashi, K., Iwamatsu, A., Takeuchi, M., Kumagai, H. & Yamamoto, K. Molecular cloning of *Mucor hiemalis* endo-β-*N*-acetylglucosaminidase and some properties of the recombinant enzyme. *Arch. Biochem. Biophys.* **432,** 41–49 (2004).

68. Katoh, T., Katayama, T., Tomabechi, Y., Nishikawa, Y., Kumada, J., Matsuzaki, Y. & Yamamoto, K. Generation of a mutant *Mucor hiemalis* endoglycosidase that acts on core-fucosylated *N*-Glycans. *Journal of Biological Chemistry* **291,** (2016).

69. Yin, J., Li, L., Shaw, N., Li, Y., Song, J., Zhang, W., Xia, C., Zhang, R., Joachimiak, A., Zhang, H., Wang, L., Liu, Z. & Wang, P. Structural basis and catalytic mechanism for the dual functional Endo-β-*N*-acetylglucosaminidase A. *PLoS One* **4,** (2009).

70. Kitajima, T., Jia, Y., Komatsuzaki, A., Cui, J., Matsuzawa, F., Aikawa, S., Gao, X. & Chiba, Y. Structural modeling and mutagenesis of endo-β-*N*-acetylglucosaminidase from *Ogataea minuta* identifies the importance of Trp295 for hydrolytic activity. *J. Biosci. Bioeng.* **125,** 168–174 (2018).

71. Fujita, K. & Takegawa, K. Tryptophan-216 Is Essential for the Transglycosylation Activity of Endo-β-*N*-acetylglucosaminidase A. *Biochem. Biophys. Res. Commun.* **283,** 680–686 (2001).

72. Fan, S., Huang, W. & Wang, L. Remarkable transglycosylation activity of glycosynthase mutants of endo-D, an endo-β-*N*-acetylglucosaminidase from *Streptococcus pneumoniae*. *J. Biol. Chem.* **287,** 11272–11281 (2012).

73. Panke, S. & Wubbolts, M. Advances in biocatalytic synthesis of pharmaceutical intermediates. *Curr. Opin. Chem. Biol.* **9,** 188–194 (2005).

74. Jenkins, N., Parekh, R. & James, D. Getting the glycosylation right: Implications for the biotechnology industry. *Nat. Biotechnol.* **14,** 975–981 (1996).

75. Li, H., Singh, S., Zeng, Y., Song, H. & Wang, L. Chemoenzymatic synthesis of CD52 glycoproteins carrying native *N*-glycans. *Bioorganic Med. Chem. Lett.* **15,** 895–898 (2005).

76. Li, T., Tong, X., Yang, Q., Giddens, J. & Wang, L. Glycosynthase mutants of endoglycosidase S2 show potent transglycosylation activity and remarkably relaxed substrate specificity for antibody glycosylation remodeling. *J. Biol. Chem.* **291,** 16508–16518 (2016).

77. Wang, L. The Amazing Transglycosylation Activity of Endo-β-*N*-acetylglucosaminidases. *Natl.*

Institutes Heal. **23,** 33–52 (2011).

78. Fan, S., Huang, W., Wang, L., Umekawa, M., Li, C., Higashiyama, T., Huang, W., Ashida, H., Yamamoto, K. & Wang, L. Efficient glycosynthase mutant derived from *Mucor hiemalis* Endo-β-*N*-acetylglucosaminidase capable of transferring oligosaccharide from both sugar oxazoline and natural *N*-glycan. *J. Biol. Chem.* **285,** 11272–11281 (2012).

79. Ashby, S. & Nowell, W. The fungi of stigmatomycosis. *Ann. Bot.* **40,** 69–84 (1926).

80. Wickerham, L., Flickinger, M. & Johnson, R. Production of riboflavin by *Ashbya gossypii*. *Arch Biochem* **9,** 8–95 (1946).

81. Stahmann, K., Arst, H., Althöfer, H., Revuelta, J., Monschau, N., Schlüpen, C., Gätgens, C., Wiesenburg, A. & Schlösser, T. Riboflavin, overproduced during sporulation of *Ashbya gossypii*, protects its hyaline spores against ultraviolet light. *Environ. Microbiol.* **3,** 545–550 (2001).

82. Stahmann, K., Revuelta, J. & Seulberger, H. Three biotechnical processes using *Ashbya gossypii*, *Candida famata*, or *Bacillus subtilis* compete with chemical riboflavin production. *Appl. Microbiol. Biotechnol.* **53,** 509–516 (2000).

83. Aguiar, T., Silva, R. & Domingues, L. *Ashbya gossypii* beyond industrial riboflavin production: A historical perspective and emerging biotechnological applications. *Biotechnol. Adv.* **33,** 1774–1786 (2015).

84. Aguiar, T., Silva, R. & Domingues, L. New biotechnological applications for *Ashbya gossypii*: Challenges and perspectives. *Bioengineered* **8,** 309–315 (2017).

85. Ribeiro, O., Magalhães, F., Aguiar, T., Wiebe, M., Penttilä, M. & Domingues, L. Random and direct mutagenesis to enhance protein secretion in *Ashbya gossypii*. *Bioengineered* **4,** 322–331 (2013).

86. Ribeiro, O., Wiebe, M., Ilmén, M., Domingues, L. & Penttilä, M. Expression of *Trichoderma reesei* cellulases CBHI and EGI in *Ashbya gossypii*. *Appl. Microbiol. Biotechnol.* **87,** 1437–1446 (2010).

87. Ledesma-Amaro, R., Santos, M., Jiménez, A. & Revuelta, J. Strain design of *Ashbya gossypii* for single-cell oil production. *Appl. Environ. Microbiol.* **80,** 1237–1244 (2014).

88. Díaz-Fernández, David Aguiar, T. Q., Martín, V. I., Romaní, A., Silva, R., Domingues, Lucília Revuelta, J. L. & Jiménez, A. Microbial lipids from industrial wastes using xylose-utilizing *Ashbya gossypii* strains. *Bioresour. Technol.* **293,** 122054 (2019).

89. Ledesma-Amaro, R., Buey, R. & Revuelta, J. Increased production of inosine and guanosine by means of metabolic engineering of the purine pathway in *Ashbya gossypii*. *Microb. Cell Fact.* **14,** 1–8 (2015).

90. Ravasio, D., Wendland, J. & Walther, A. Major contribution of the Ehrlich pathway for 2-phenylethanol/rose flavor production in *Ashbya gossypii*. *FEMS Yeast Res.* **14,** 833–844 (2014).

91. Silva, R., Aguiar, T. Q., Coelho, E., Jiménez, A., Revuelta, J. L. & Domingues, L. Metabolic engineering of *Ashbya gossypii* for deciphering the *de novo* biosynthesis of γ-lactones. *Microb. Cell Fact.* **18,** 1–11 (2019).

92. Schmitz, H. & Philippsen, P. Evolution of multinucleated *Ashbya gossypii* hyphae from a budding yeast-like ancestor. *Fungal Biol.* **115,** 557–568 (2011).

93. Wendland, J. & Walther, A. *Ashbya gossypii*: A model for fungal developmental biology. *Nat. Rev. Microbiol.* **3,** 421–429 (2005).

94. Prillinger, H., Schweigkofler, W., Breitenbach, M., Briza, P., Staudacher, E., Lopandic, K., Molnár,

O., Weigang, F., Ibl, M. & Ellinger, A. Phytopathogenic filamentous (*Ashbya*, *Eremothecium*) and dimorphic fungi (*Holleya*, *Nematospora*) with needle-shaped ascospores as new members within the *Saccharomycetaceae*. *Yeast* **13,** 945–960 (1997).

95. Dietrich, F., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Mohr, C., Pöhlmann, R., Luedi, P., Choi, S., Wing, R., Flavier, A., Gaffney, T., Philippsen, P., Steiner, S. & Pohimann, R. The *Ashbya gossypii* Genome as a Tool for Mapping the Ancient *Saccharomyces cerevisiae* Genome. *Science* **304,** 304–307 (2004).

96. Hermida, L., Brachat, S., Voegeli, S., Philippsen, P. & Primig, M. The *Ashbya* Genome Database (AGD) - A tool for the yeast community and genome biologists. *Nucleic Acids Res.* **33,** 348–352 (2005).

97. Wasserstrom, L., Lengeler, K., Walther, A. & Wendland, J. Developmental growth control exerted via the protein a kinase Tpk2 in *Ashbya gossypii*. *Eukaryot. Cell* **14,** 593–601 (2015).

98. Boutroux, M. Sur la conservation des ferments alcoholiques dans la nature. *Ann Soc Nat Bot* **17,** 144–209 (1884).

99. Souciet, J. L. *et al.* Comparative genomics of protoploid *Saccharomycetaceae*. *Genome Res.* **19,** 1696–1709 (2009).

100. Wilmotte, A., Van de Peer, Y., Goris, A., Chapelle, S., De Baere, R., Nelissen, B., Neefs, J., Hennebert, G. L. & De Wachter, R. Evolutionary Relationships Among Higher Fungi Inferred from Small Ribosomal Subunit RNA Sequence Analysis. *Syst. Appl. Microbiol.* **16,** 436–444 (1993).

101. Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430,** 35–44 (2004).

102. James, S. A. & Stratford, M. *Zygosaccharomyces* Barker (1901). in *The Yeasts, a Taxonomic Study* (eds. Kurtzman, C. P., Fell, J. W. & Boekhout, T.) 937–947 (Elsevier, London, 2011).

103. Gordon, J. L. & Wolfe, K. H. Recent allopolyploid origin of *Zygosaccharomyces rouxii* strain ATCC42981. *Yeast* **25,** 449–456 (2008).

104. Pribylova, L., Montigny, J. De & Sychrova, H. Osmoresistant yeast *Zygosaccharomyces rouxii*: the two most studied wild-type strains (ATCC 2623 and ATCC 42981) differ in osmotolerance and glycerol. **1,** 171–180 (2007).

105. Solieri, L., Cassanelli, S. & Giudici, P. A new putative *Zygosaccharomyces* yeast species isolated from traditional balsamic vinegar. 403–417 (2007).

106. Deák, T. Yeasts in specific types of foods. in *Handbook of Food Spoilage Yeasts* (ed. Deák, T.) 117–201 (Boca Raton: CRC Press., 2007).

107. Wickerham, L. & Burton, K. Heterothallism in *Saccharomyces rouxii*. *J Bacteriol.* **80,** 492–495 (1960).

108. Solieri, L., Chand Dakal, T., Croce, M. A. & Giudici, P. Unravelling genomic diversity of *Zygosaccharomyces rouxii* complex with a link to its life cycle. *FEMS Yeast Res.* **13,** 245–258 (2013).

109. Solieri, L., Cassanelli, S., Croce, M. A. & Giudici, P. Genome size and ploidy level: New insights for elucidating relationships in *Zygosaccharomyces* species. *Fungal Genet. Biol.* **45,** 1582–1590 (2008).

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

111. Rosano, G. & Ceccarelli, E. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* **5,** 1–17 (2014).

112. Giannuzzi, G., Lobefaro, N., Paradies, E., Vozza, A., Punzi, G. & Marobbio, C. M. T. Overexpression in *E. coli* and purification of the *L. pneumophila* Lpp2981 protein. *Mol. Biotechnol.* **56,** 157–165 (2014).

113. Kim, M. J., Park, H. S., Seo, K. H., Yang, H. J., Kim, S. K. & Choi, J. H. Complete Solubilization and Purification of Recombinant Human Growth Hormone Produced in *Escherichia coli*. *PLoS One* **8,** (2013).

114. Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **36,** 1136–1145 (2018).

115. Robbins, P., Wirth, D. & Hering, C. Expression of the *Streptomyces* enzyme endoglycosidase H in *Escherichia coli*. *J. Biol. Chem.* **256,** 10640–10644 (1981).

116. Takegawa, K., Yamabe, K., Fujita, K., Tabuchi, M., Mita, M., Izu, H., Watanabe, A., Asada, Y., Sano, M., Kondo, A., Kato, I. & Iwahara, S. Cloning, sequencing, and expression of *Arthrobacter protophormiae* Endo-β-*N*-acetylglucosaminidase in *Escherichia coli*. *Arch. Biochem. Biophys.* **338,** 22–28 (1997).

117. Eshima, Y., Higuchi, Y., Kinoshita, T., Nakakita, S. & Takegawa, K. Transglycosylation activity of glycosynthase mutants of endo-β-*N*-Acetylglucosaminidase from *Coprinopsis cinerea*. *PLoS One* **10,** 1–15 (2015).

118. Kaur, J., Kumar, A. & Kaur, J. Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. *Int. J. Biol. Macromol.* **106,** 803–822 (2018).

119. Lebendiker, M. & Danieli, T. Production of prone-to-aggregate proteins. *FEBS Lett.* **588,** 236–246 (2014).

120. Robinson, M. P., Ke, N., Lobstein, J., Peterson, C., Szkodny, A., Mansell, T. J., Tuckey, C., Riggs, P. D., Colussi, P. A., Noren, C. J., Taron, C. H., Delisa, M. P. & Berkmen, M. Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. *Nat. Commun.* **6,** (2015).

121. Ang, I., Atte, A., Halim, B. & Jassal, J. Effect of temperature, inducer concentration, and *Escherichia coli* cytosolic redox state on MBP-PI2 expression. *J. Exp. Microbiol. Immunol.* **21,** 11–14 (2017).

122. García-Fraga, B., da Silva, A., López-Seijas, J. & Sieiro, C. Optimized expression conditions for enhancing production of two recombinant chitinolytic enzymes from different prokaryote domains. *Bioprocess Biosyst Eng.* **38,** 2477–86 (2015).

123. Collins, T., Azevedo-Silva, J., da Costa, A., Branca, F., Machado, R. & Casal, M. Batch production of a silk-elastin-like protein in *E. coli* BL21(DE3): key parameters for optimisation. *Microb. Cell Fact.* **12,** 1–16 (2013).

124. Piubelli, L., Campa, M., Temporini, C., Binda, E., Mangione, F., Amicosante, M., Terreni, M., Marinelli, F. & Pollgioni, L. Optimizing *Escherichia coli* as a protein expression platform to produce *Mycobacterium tuberculosis* immunogenic proteins. *Microb. Cell Fact.* **12,** 115 (2013).

125. San-Miguel, T., Pérez-Bermúdez, P. & Gavidia, I. Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. *Springerplus* **2,** 1–4 (2013).

126. Mansey, M., Ghareeb, K., Moghazy, A., Tawfick, M., Fouda, M., El Marzugi, N., Othman, N. & El

Enshasy, H. Glucose concentration affects recombinant interferon [alpha]-2b production in *Escherichia coli* using thermo-induction system. *J. Appl. Pharm. Sci.* **4,** 1–5 (2014).

127. Vu, T., Jeong, B., Krupa, M., Kwon, U., Song, J.-A., Do, B., Nguyen, M., Seo, T., Nguyen, A., Joo, C. & Choe, H. Soluble Prokaryotic Expression and Purification of Human Interferon Alpha-2b Using a Maltose-Binding Protein Tag. *J Mol Microbiol Biotechnol* **26,** 359–368 (2016).

128. Li, H., Liu, N., Wang, W., Wang, J. & Gao, W.-Y. Cloning and characterization of GST fusion tag stabilized large subunit of *Escherichia coli* acetohydroxyacid synthase I. *J. Biosci. Bioeng.* **121,** 21–26 (2016).

129. Saitoh, H., Uwada, J. & Azusa, K. Strategies for the Expression of SUMO-Modified Target Proteins in *Escherichia coli*. in *SUMO Protocols. Methods In Molecular Biology* (ed. Ulrich, H.) (Humana Press, 2009).

130. Xu, Z., Peng, L., Zhong, Z., Fang, X. & Cen, P. High-Level Expression of a Soluble Functional Antimicrobial Peptide, Human beta-Defensin 2, in *Escherichia coli*. *Biotechnol. Prog.* **22,** 382–386 (2006).

131. Turner, P., Holst, O. & Karlsson, E. Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in *Escherichia coli* by using a soluble fusion-tag and by tuning of inducer concentration. *Protein Expr. Purif.* **39,** 54–60 (2005).

132. Kutyshenko, V., Mikoulinskaia, G., Chernyshov, S., Yegorov, A., Prokhorov, D. & Uversky, V. Effect of C-terminal His-tag and purification routine on the activity and structure of the metalloenzyme, L-alanyl-D-glutamate peptidase of the bacteriophage T5. *Int. J. Biol. Macromol.* **124,** 810–818 (2019).

133. Rose, P. ., Bi, C., Bluhm, W. ., Christie, C. ., Dimitropoulos, D., Dutta, S., Green, R. ., Goodsell, D. ., Prlic, A., Quesada, M., Quinn, G. ., Ramos, A. ., Westbrook, J. ., Young, J., Zardecki, C., Berman, H. . & Bourne, P. . The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res.* **41,** D475-82 (2013).

134. Madeira, F., Park, Y. ., Lee, J., Buso, N., Gur, T., Madhusoodanan, N Basutkar, P., Tivey, A. R. ., Potter, S. ., Finn, R. . & Lopez, R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* gkz268 (2019).

135. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R. & Schwede, T. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46,** W296–W303 (2018).

136. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. ., Appel, R. . & Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* (ed. Walker, J. .) 571–607 (Humana Press Inc., 2005).

137. Ferrè, F. & Clote, P. DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. *Nucleic Acids Res.* **34,** W182–W185 (2006).

138. Wendland, J., Ayad-Durieux, Y., Knechtle, P., Rebischung, C. & Philippsen, P. PCR-based gene targeting in the filamentous fungus *Ashbya gossypii*. *Gene* **242,** 381–391 (2000).

139. Verduyn, C., Postma, E., Scheffers, W. . & Van Dijken, J. . Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* **8,** 501–517 (1992).

140. Aguiar, T. Q., Dinis, C. & Domingues, L. Cre-loxP-based system for removal and reuse of selection

markers in *Ashbya gossypii* targeted engineering. *Fungal Genet. Biol.* **68,** 1–8 (2014).

141. Silva, R., Aguiar, T. Q., Oliveira, R. & Domingues, L. Light exposure during growth increases riboflavin production, reactive oxygen species accumulation and DNA damage in *Ashbya gossypii* riboflavin-overproducing strains. *FEMS Yeast Res.* **19,** (2019).

142. Lawrence, A. & Besir, H. Staining of Proteins in Gels with Coomassie G-250 without Organic Solvent and Acetic Acid. *J. Vis. Exp.* **30,** e1350 (2009).

143. Magalhães, F., Aguiar, T., Oliveira, C. & Domingues, L. High-level expression of *Aspergillus niger* β-galactosidase in *Ashbya gossypii*. *Biotechnol. Prog.* **30,** 261–268 (2014).

144. Cardoso, B. B., Silvério, S. C., Abrunhosa, L., Teixeira, J. A. & Rodrigues, L. R. β-galactosidase from *Aspergillus lacticoffeatus*: A promising biocatalyst for the synthesis of novel prebiotics. *Int. J. Food Microbiol.* **257,** 67–74 (2017).

145. Aguiar, T. Q., Ribeiro, O., Arvas, M., Wiebe, M. G., Penttilä, M. & Domingues, L. Investigation of protein secretion and secretion stress in *Ashbya gossypii*. *BMC Genomics* **15,** 1137 (2014).

146. Yamamoto, K., Kadowaki, S., Fujisaki, M., Kumagai, H. & Tochikura, T. Novel Specificities of *Mucor hiemalis* Endo-β-*N*-acetylglucosaminidase Acting Complex Asparagine-Linked Oligosaccharides. *Biosci. Biotechnol. Biochem.* **58,** 72–77 (1994).

147. LaVallie, E., DiBlasio, E., Kovacic, S., Grant, K., Schendel, P. & McCoy, J. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnol. (N Y)* **11,** 187–93 (1993).

148. LaVallie, E., Lu, Z., Diblasio-Smith, E., Collins-Racie, L. & McCoy, J. Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. *Methods Enzym.* **326,** 322–40 (2000).

149. Daniel, E., Onwukwe, Goodluck U. Wierenga, R. K., Quaggin, S. E., Vainio, S. J. & Krause, M. ATGme: Open-source web application for rare codon identification and custom DNA sequence optimization. *BMC Bioinformaticsvolume* **16,** (2015).

150. Chomczynski, P. & Mackey, K. Reduced Background Expression and Improved Plasmid Stability with pET Vectors in BL21 (DE3). *Biotechniques* **19,** 942–5 (1995).

151. Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P. & Berkmen, M. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* **11,** 1 (2012).

152. Chen, W. B., Nie, Y., Xu, Y. & Xiao, R. Enhancement of extracellular pullulanase production from recombinant *Escherichia coli* by combined strategy involving auto-induction and temperature control. *Bioprocess Biosyst. Eng.* **37,** 601–608 (2014).

153. Jia, B. & Jeon, C. O. High-throughput recombinant protein expression in *Escherichia coli*: Current status and future perspectives. *Open Biol.* **6,** (2016).

154. Gräslund, S., Nordlund, P., Weigel, J., Hallberg, B. M., Bray, J., Gileadi, O. & Knapp, S. Protein production and purification. *Nat Methods* **5,** 135–146 (2008).

155. Leibly, D. J., Nguyen, T. N., Kao, L. T., Hewitt, S. N., Barrett, L. K. & van Voorhis, W. C. Stabilizing Additives Added during Cell Lysis Aid in the Solubilization of Recombinant Proteins. *PLoS One* **7,** (2012).

156. Chen, S., Manabe, Y., Minamoto, N., Saiki, N. & Fukase, K. Development of a simple assay system

for protein-stabilizing efficiency based on hemoglobin protection against denaturation and measurement of the cooperative effect of mixing protein stabilizers. *Biosci. Biotechnol. Biochem.* **80,** 1874–1878 (2016).

157. Oganesyan, N., Ankoudinova, I., Kim, S. H. & Kim, R. Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. *Protein Expr. Purif.* **52,** 280–285 (2007).

158. Prasad, S., Khadatare, P. B. & Roy, I. Effect of chemical chaperones in improving the solubility of recombinant proteins in *Escherichia coli*. *Appl. Environ. Microbiol.* **77,** 4603–4609 (2011).

159. Sandee, D., Tungpradabkul, S., Kurokawa, Y., Fukui, K. & Takagi, M. Combination of Dsb coexpression and an addition of sorbitol markedly enhanced soluble expression of single-chain Fv in *Escherichia coli*. *Biotechnol. Bioeng.* **91,** 418–424 (2005).

160. Chopra, A., Brasier, A., Das, M., Xu, X.-J. & Peterson, J. Improved synthesis of *Salmonella typhimurium* enterotoxin using gene fusion expression systems. *Gene* **144,** 81–85 (1994).

161. de Lencastre, Novaes LC Mazzola, P., Pessoa, A. J. & Penna, T. Citrate and phosphate influence on green fluorescent protein thermal stability. *Biotechnol Prog.* **27,** 269–72 (2011).

162. Ignatova, Z. & Gierasch, L. Inhibition of protein aggregation *in vitro* and *in vivo* by a natural osmoprotectant. *Proc Natl Acad Sci U S A.* **103,** 13357–61 (2006).

163. Ghosh, R., Sharma, S. & Chattopadhyay, K. Effect of arginine on protein aggregation studied by fluorescence correlation spectroscopy and other biophysical methods. *Biochemistry* **48,** 1135–43 (2009).

164. Jevševar, S., Gaberc-Porekar, V., Fonda, I., Podobnik, B., Grdadolnik, J. & Menart, V. Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. *Biotechnol. Prog.* **21,** 632–639 (2005).

165. Dünkler, A., Jorde, S. & Wendland, J. An *Ashbya gossypii* cts2 mutant deficient in a sporulation-specific chitinase can be complemented by *Candida albicans* CHT4. *Microbiol. Res.* **163,** 701–710 (2008).

166. Silva, R., Aguiar, T. Q. & Domingues, L. Blockage of the pyrimidine biosynthetic pathway affects riboflavin production in *Ashbya gossypii*. *J. Biotechnol.* **193,** 37–40 (2015).

167. Barreaud, J. P., Bourgerie, S., Julien, R., Guespin-Michel, J. F. & Karamanos, Y. An endo-*N*-acetyl-β-D-glucosaminidase, acting on the di-*N*-acetylchitobiosyl part of *N*-linked glycans, is secreted during sporulation of *Myxococcus xanthus*. *J. Bacteriol.* **177,** 916–920 (1995).

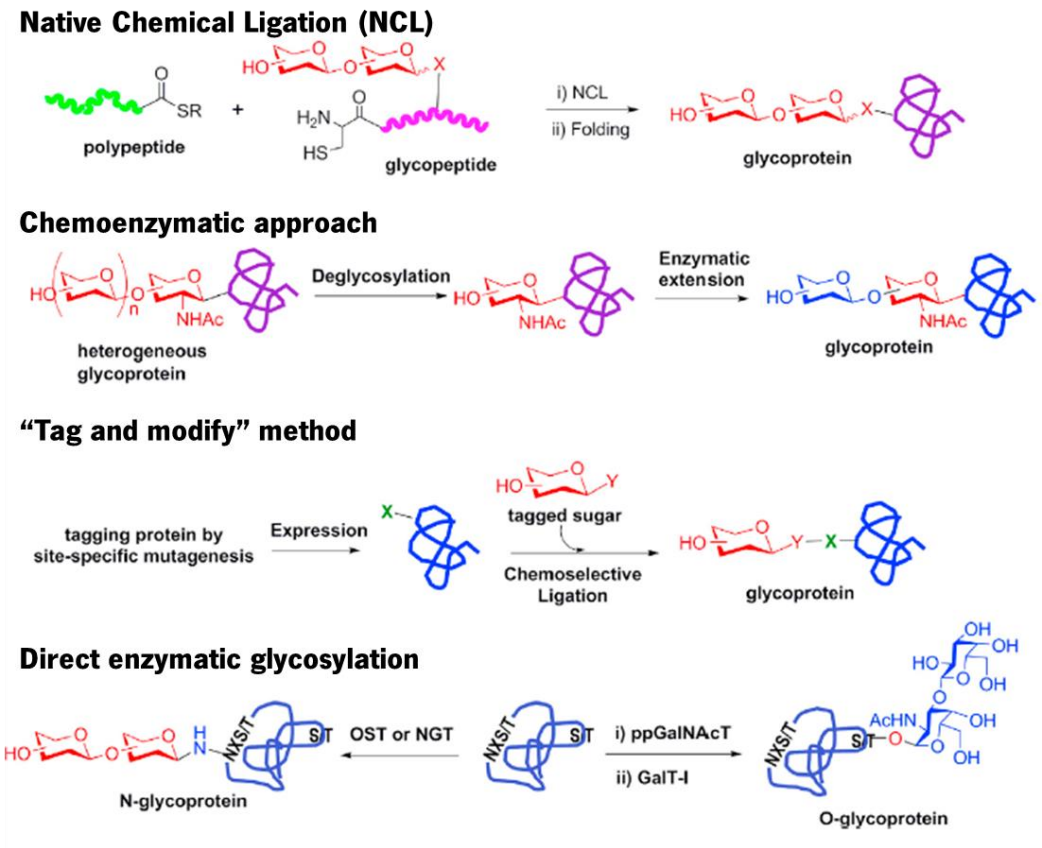
168. Tzelepis, G., Hosomi, A., Hossain, T. J., Hirayama, H., Dubey, M., Jensen, D. F., Suzuki, T. & Karlsson, M. Endo-β-*N*-acetylglucosamidases (ENGases) in the fungus *Trichoderma atroviride*: Possible involvement of the filamentous fungi-specific cytosolic ENGase in the ERAD process. *Biochem. Biophys. Res. Commun.* **449,** 256–261 (2014).

APPENDICES

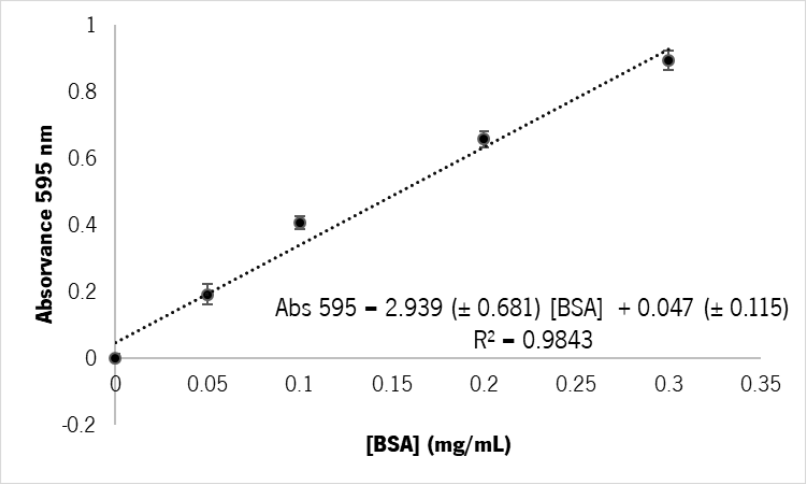
APPENDICES

Hexose	Glc	Man	Gal	Gul	Alt	All	Tal	Ido	
HexNAc	GlcNAc	ManNAc	GalNAc	GulNAc	AltNAc	AllNAc	TalNAc	IdoNAc	
Hexosamine	GlcN	ManN	GalN	GulN	AltN	AllN	TalN	IdoN	
Hexuronate	GlcA	ManA	GalA	GulA	AltA	AllA	TalA	IdoA	
Deoxyhexose	Qui	Rha		6dGul	6dAlt		6dTal		Fuc
DeoxyhexNAc	QuiNAc	RhaNAc			6dAltNAc		6dTalNAc		FucNAc
Di-deoxyhexose	Oli	Tyv		Abe	Par	Dig	Col		
Pentose		Ara	Lyx	Xyl	Rib				
Deoxynonulosonate		Kdn				Neu5Ac	Neu5Gc	Neu	Sia
Di-deoxynonulosonate		Pse	Leg		Aci		4eLeg		
Unknown	Bac	LDmanHep	Kdo	Dha	DDmanHep	MurNAc	MurNGc	Mur	
Assigned	Api	Fru	Tag	Sor	Psi				

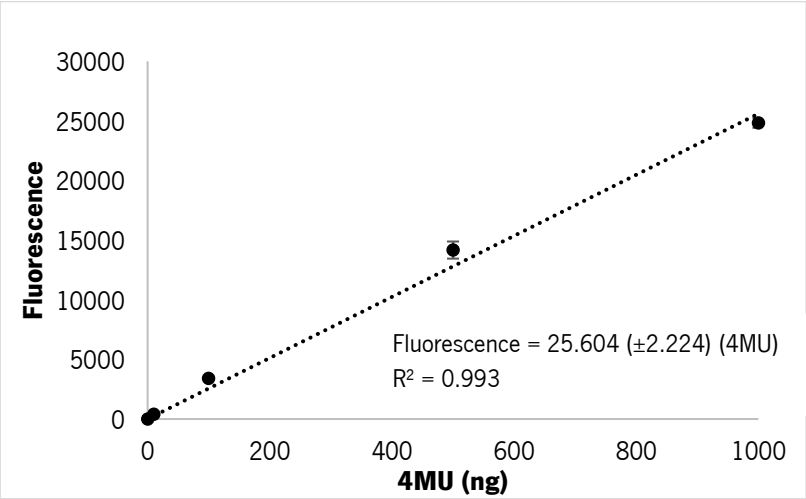
Appendix A. Monosaccharide symbol nomenclature. Figure adapted from [12].



Appendix B. Schematic summary of the common synthetic strategies used in engineering glycopeptides and glycoproteins. Figure adapted from [28].



Appendix C. Calibration curve built and used for the determination of total protein by the Bradford method, using BSA as standard. Different concentrations of the standard were used to build the curve: 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 mg/mL. Each point represents the average of 6 independent assays. Error bars represent standard deviation.



Appendix D. Calibration curve built and used for the determination of chitinase activity, using 4MU as standard. Different quantities of standard were used to build the curve: 10, 100, 500, 1000 ng. Each point represents the average of 6 independent assays. Error bars represent standard deviation.

Novel enzymes for the analysis and glycoengineering of therapeutic glycoproteins

University of Minho
School of Engineering

CENTRE OF BIOLOGICAL ENGINEERING

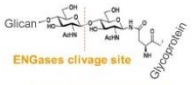
Marlene Baptista, Carla Oliveira, Lucília Domingues and Tatiana Q. Aguiar

CEB – Centre of Biological Engineering, University of Minho, Campus de Gualtar, Braga 4710-057, Portugal
E-mail: tatiana.aguiar@deb.uminho.pt

INTRODUCTION

- Many proteins of clinical and pharmaceutical interest are *N*-glycosylated, being their bioactivity, pharmacokinetics and pharmacodynamics affected by the glycan structures they carry. Thus, one of the biggest challenges in the pharmaceutical and biomedical areas is the manufacturing of glycopeptides and glycoproteins with homogenous and defined oligosaccharide structures.
- Endo-β-*N*-acetylglucosaminidases (ENGases) of the glycoside hydrolase (GH) family 85 are a class of enzymes (EC 3.2.1.96) that, in addition to hydrolytic activity against the diacetylchitobiose core of *N*-glycans, can also display transglycosylation activity [1].
- These enzymes have increasingly become a focus of interest due to their useful applications in the analysis and glycoengineering of therapeutic glycopeptides and glycoproteins.
- Although family GH85 currently contains 768 members, only 11 GH85 ENGases have been characterized thus far [2].
- Envisioning the identification of new GH85 ENGases with useful action, this study focused on two new putative ENGases of this family: Q752H6 from the filamentous fungus *Ashbya gossypii* and C5DRB8 from the yeast *Zygosaccharomyces rouxii*.
- **Previous results hinted at the existence of ENGase activity in these organisms [3,4], and therefore this work aimed at using *in silico* approaches to assess about the potential activity of their putative ENGases.**

N-N'-diacetylchitobiose core



Hidrolytic activity

Figure adapted from [1].

METHODOLOGY

Multiple alignments

Performed with Clustal Omega (EMBL-EBI).

Phylogenetic tree

Built in Clustal Omega (EMBL-EBI) using the Neighbor-Joining method and the results from the alignments.

Visualized in FigTree v1.4.4.

ENGase	Origin	UniProt Accession
Endo-D	*Streptococcus pneumonia*	Q93HW0
Endo-BB	*Bifidobacterium longum*	Q8G4P3
Endo-BH	*Bacillus halodurans*	Q9KER4
Endo-A	*Arthrobacter protophormiae*	Q9ZB22
Endo-M	*Mucor hiemalis*	Q9C1S6
Endo-Om	*Ogatea minuta*	R4WHQ8
ZrENGase	*Zygosaccharomyces rouxii*	C5DRB8
AgENGase	*Ashbya gossypii*	Q752H6
Endo-CE	*Caenorhabditis elegans*	Q19089
DmENGase	*Drosophila melanogaster*	Q86B46
AtENGase2	*Arabidopsis thaliana*	Q9SRL4
AtENGase1	*Arabidopsis thaliana*	F4JZC2
HsENGase	*Homo sapiens*	Q8NFI3
MmENGase	*Mus musculus*	Q8BJF1

Homology-based models construction

Models were constructed with the SWISS-MODEL tool (ExPASy), using as template the available structure of the Endo-A mutant (PDB: 2VTF.1.A).

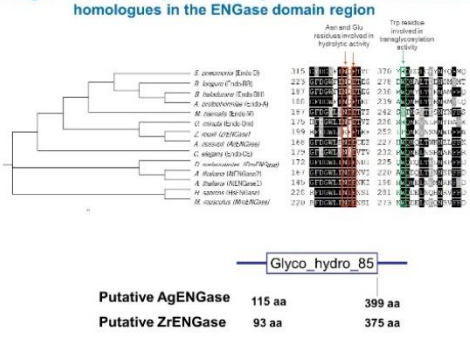
Pymol (Schrödinger) was used for the visualization and analysis of the models.

Domain search

ENGase domain search was performed using Pfam 32.0 (EMBL-EBI).

RESULTS

Multiple alignments revealed possible catalytic residues conserved between homologues in the ENGase domain region



Putative AgENGase	115 aa	399 aa
Putative ZrENGase	93 aa	375 aa

The homology-based 3D models generated for the putative *Z. rouxii* ENGase revealed the presence of a TIM barrel (β_8/α_8) common to several previously characterized ENGases of GH85 family

Endo-A

Putative AgENGase

Putative ZrENGase

ENGase	Amino acid identity related to the template (%)	Coverage in domain region
Putative AgENGase	25.58	0.44
Putative ZrENGase	18.32	0.85

FINAL REMARKS

- Despite the potential of ENGases for pharmaceutical and biomedical applications, there are few of these enzymes commercially available and the existing ones are quite expensive. As such, the discovery of new enzymes becomes essential.
- Multiple alignments and homology-based prediction of 3D structure of the putative ENGases presented here revealed the potential of these enzymes to catalyze hydrolytic and transglycosylation reactions.
- The cDNA fragments encoding the putative ZrENGase and AgENGase were already cloned in *E. coli* expression pETM vectors and transformed into different *E. coli* expression strains. We are working in the production and purification process for subsequent characterization of their activity.

References

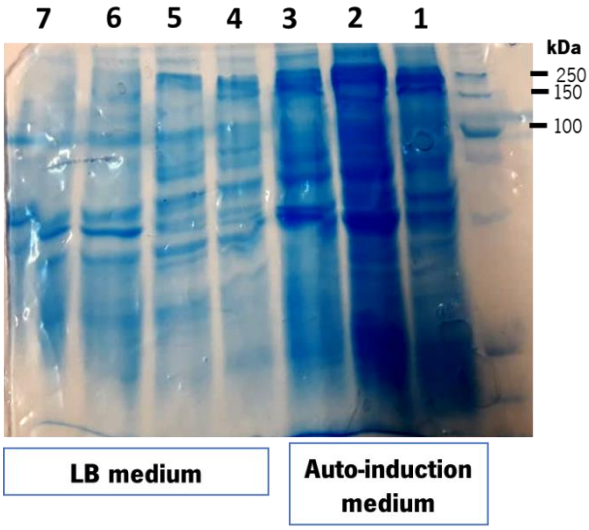
[1] Fairbanks, Chem Soc Rev 2017; 46:5128; [2] CAZy database, 2019/04/26; [3] Aguiar *et al.*, Carbohydr Res 2013; 381:19-27; [4] Murakami *et al.*, Glycobiology 2013; 23:736-744

Acknowledgments

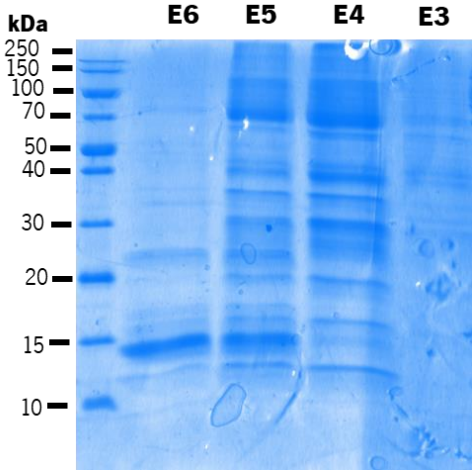
Study supported by FCT and by the European Regional Development Fund under the scope of Compete2020 and Norte2020 strategic funding of UID/BIO/04469/2019. Project EcoBioInks4SmartTextiles (PTDC/CTM-TEX/30298/2017 - POCI-01-0145-FEDER-030298) and BioTecNorte operation (NORTE-01-0145-FEDER-000004).

17 May 2019

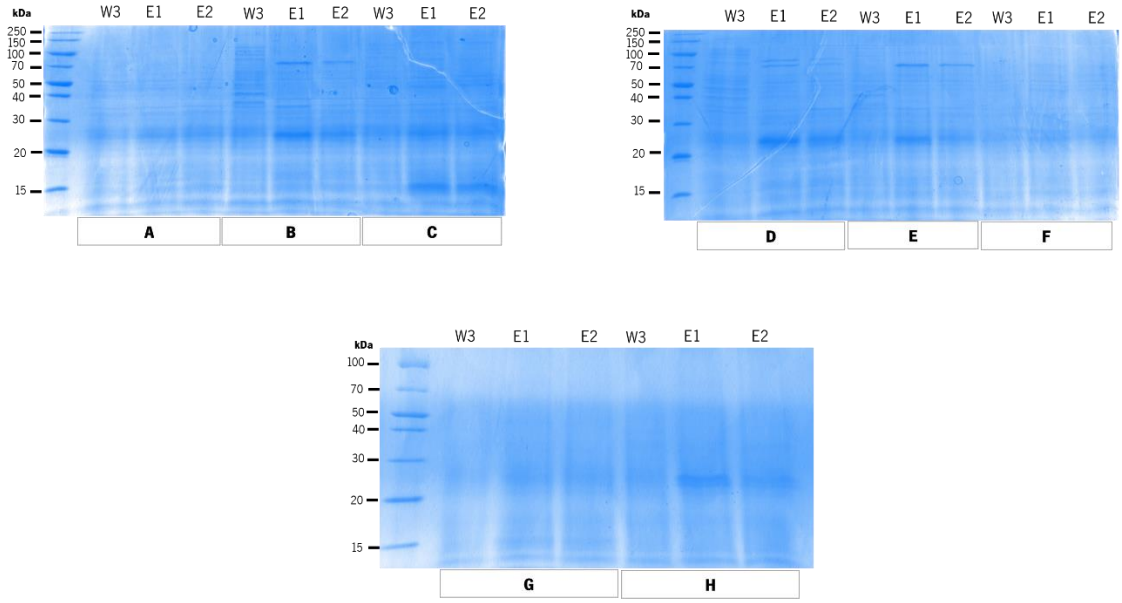
Appendix E. Poster presented at the 4th Meeting of Medicinal Biotechnology, Porto, 17/05/2019.



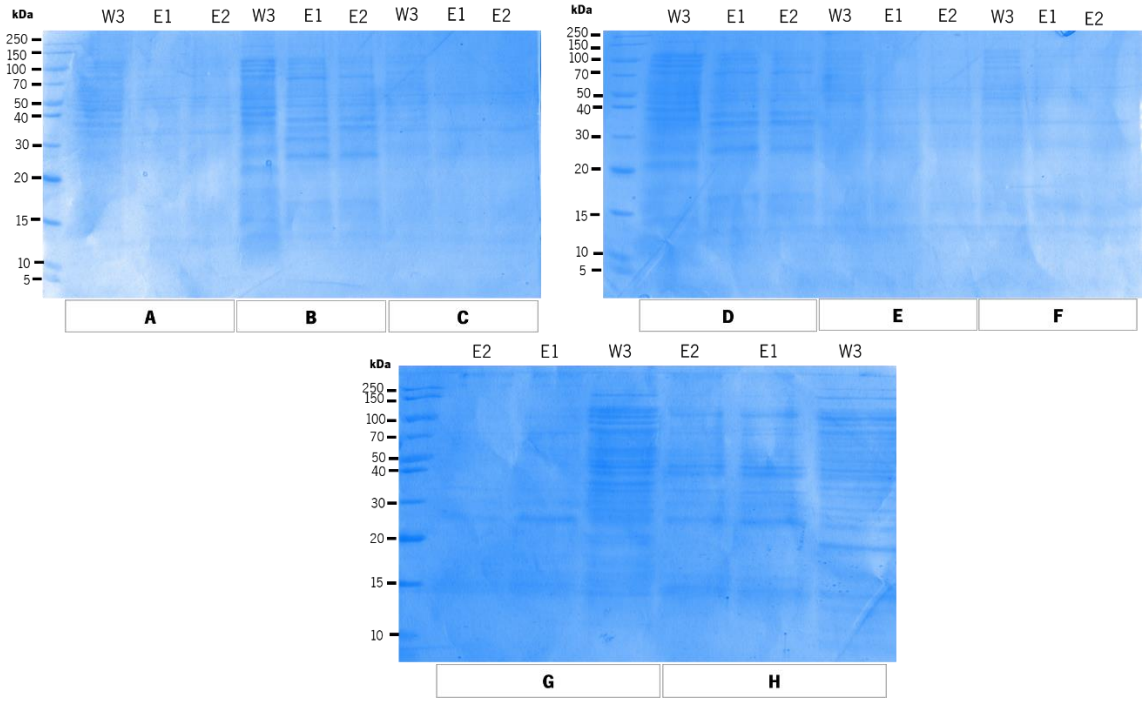
Appendix F. SDS-PAGE gels from productions of *E. coli* BL21(DE3) or Rosetta 2 (DE3) ttransformed with the pETM20 containing the gene encoding the ZrENGase with codons not optimized for expression in *E. coli*. 1- Total protein fraction recovered from a culture of *E. coli* Rosetta 2 (DE3) grown in AI medium, **2-** Total protein fraction recovered from a culture of *E. coli* Origami 2 grown in AI medium, **3-** Total protein fraction recovered from a culture of *E. coli* BL21 grown in AI medium, **4-** Total protein fraction recovered after 16 h of induction in LB medium of *E. coli* Origami 2, **5-** Total protein fraction recovered just before induction in LB medium of *E. coli* Origami 2, **6-** Total protein fraction recovered after 16 h of induction in LB medium of *E. coli* Rosetta 2 (DE3), **7-** Total protein fraction recovered just before induction in LB medium of *E. coli* Rosetta 2 (DE3).



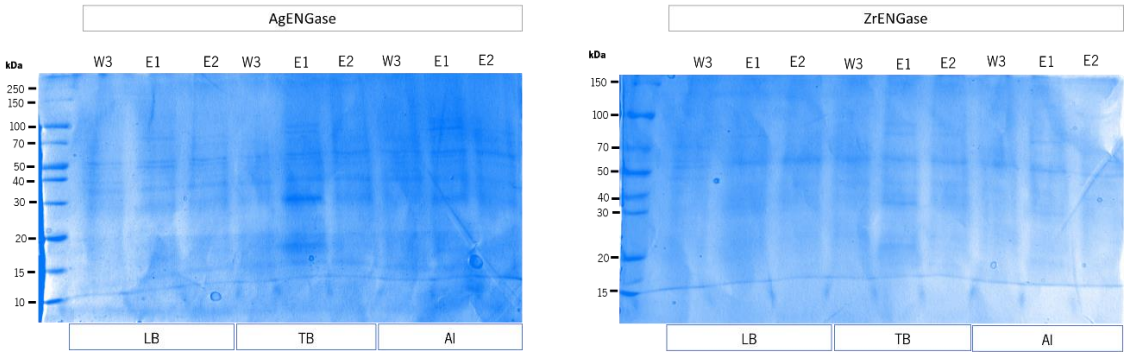
Appendix G. Small-scale purification of the fusion TrxA-His-tag-TEV-ZrENGase produced by *E. coli* Rosetta 2 (DE3) from pETM20 containing the gene encoding ZrENGase with codons not optimized for expression in *E. coli*. E3, E4, E5 and E6 correspond to eluted fractions. The predicted MW of the fusion protein is 88 kDa.



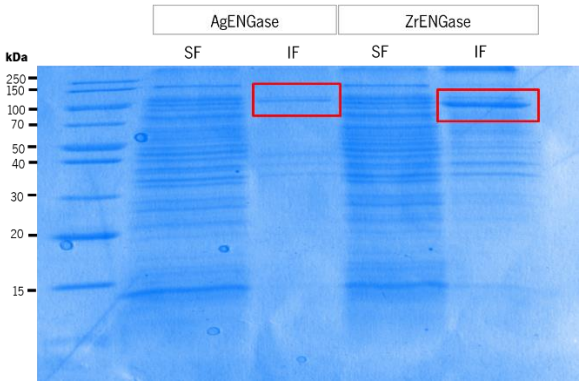
Appendix H. Small-scale batch purifications of the soluble fractions collected from *E. coli* Origami 2 (DE3) and SHuffle T7 Express productions in LB and AI media at 20 ℃. Samples were collected after the third washing step of the purification (W3) and first and second eluted fraction and were analysed in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. **A-** Soluble fraction of *E. coli* Origami 2 (DE3) with putative ZrENGase grown in LB medium, **B-** Soluble fraction of *E. coli* Origami 2 (DE3) with putative AgENGase grown in AI medium, **C-** Soluble fraction of *E. coli* SHuffle T7 Express with putative AgENGase grown in LB medium, **D-** Soluble fraction of *E. coli* SHuffle T7 Express with putative AgENGase grown in AI medium, **E-** Soluble fraction of *E. coli* SHuffle T7 Express with putative ZrENGase grown in AI medium, **F-** Soluble fraction of *E. coli* SHuffle T7 Express with putative ZrENGase grown in LB medium, **G-** Soluble fraction of *E. coli* Origami 2 (DE3) with putative AgENGase grown in LB medium, **H-** Soluble fraction of *E. coli* Origami 2 (DE3) with putative ZrENGase grown in AI medium.



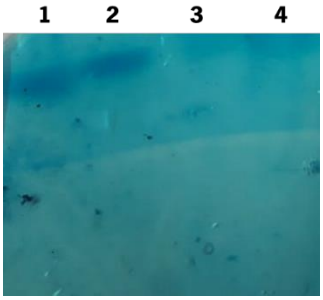
Appendix I. Small-scale batch purifications of the soluble fractions of collected from *E. coli* Origami 2 (DE3) and SHuffle T7 Express productions in LB and TB media at 20 ℃. Samples were collected after the third washing step of the purification (W3) and first and second eluted fraction and were analysed in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. **A-** Soluble fraction of *E. coli* Origami 2 (DE3) with putative AgENGase grown in LB medium, **B-** Soluble fraction of *E. coli* Origami 2 (DE3) with putative AgENGase grown in TB medium, **C-** Soluble fraction of *E. coli* Origami 2 (DE3) with putative ZrENGase grown in LB medium, **D-** Soluble fraction of *E. coli* Origami 2 (DE3) with putative ZrENGase grown in TB medium, **E-** Soluble fraction of *E. coli* SHuffle T7 Express with putative AgENGase grown in LB medium, **F-** Soluble fraction of *E. coli* SHuffle T7 Express with putative ZrENGase grown in LB medium, **G-** Soluble fraction of *E. coli* SHuffle T7 Express with putative ZrENGase grown in TB medium, **H-** Soluble fraction of *E. coli* SHuffle T7 Express with putative AgENGase grown in TB medium.



Appendix J. Small-scale batch purifications of the soluble fractions collected from *E. coli* BL21 (DE3) productions in LB, TB and AI media at 20 ℃. Samples were collected after the third washing step of the purification (W3) and first and second eluted fraction (E1 and E2). They were then analysed in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue.



Appendix K. Recombinant AgENGase and ZrENGase production in *E. coli* Origami 2 (DE3) in LB medium at 4 ℃ for 72 h. Insoluble (TF) and soluble (SF) protein fractions were resolved in a 15 % acrylamide gel and stained with Coomassie Brilliant Blue. Bands with the predicted MW of the fusions TrxA-His-tag-TEV-AgENGase (92 kDa) / ZrENGase (88 kDa) are in the red boxes.



Appendix L. Zymogram of glycosylated and de-glycosylated β-galactosidase. 1- Positive control (β-galactosidase deglycosylated with Endo-H), **2-** Negative control (β-galactosidase), **3-** β-galactosidase incubated with *A. gossypii* crude cell extract, **4-** β-galactosidase incubated with *Z. rouxii* crude cell extract.