COMPARATIVE PROTEIN ANALYSIS TO INVESTIGATE *CHLAMYDOMONAS REINHARDTII* AS A CELL BIOFACTORY

Lorenzo Barolo

PhD by research

University of Technology Sydney

School of Life Sciences, Climate Change Cluster (C3)

2020

Supervision:
Dr Mathieu Pernice (primary supervisor)
Dr Raffaela M. Abbriano, Dr Audrey S. Commault,
Dr Matthew P. Padula (co-supervisors)

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Lorenzo Barolo, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in the School of Life Science at the University of Technology Sydney. This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Date: 16/11/2020

ACKNOWLEDGEMENTS

Audrey, Raffie, Mathieu, this would have not happened without you. I'll never forget it.

I want to thank Doctor Albert Nal for his support.

No microalgae were harmed in the making of this thesis (kind of).

PREFACE

This thesis has been prepared for submission as a thesis by compilation, therefore the thesis contains a combination of submitted and publishable work. Consequently, there is a degree of repetition across different chapters, especially within the introductions and the materials and methods sections of Chapter 1, 2, 3, and 4. Chapter 1 was written for publication as a literature review. While in the final part of this Chapter the concept of glyco-engineering is introduced and extensively described, this was mainly a perspective and not the focus of this PhD thesis. Therefore, in my thesis, no data for glyco-engineering in microalgae was generated. Submitted works have been incorporated in this thesis and appear as they were presented to the journal with the following modifications: i) the font and format was changed to maintain consistency throughout the whole thesis, ii) figures and tables were re-numbered to reflect the chapter numbering, and iii) supplementary figures have been re-numbered.

LIST OF PUBLICATIONS INCLUDED IN THE THESIS

Chapter 1:

Perspectives for glyco-engineering of recombinant biopharmaceuticals from microalgae

Lorenzo Barolo, Raffaela M. Abbriano, Audrey S. Commault, Jestin George, Tim Kahlke, Michele Fabris, Matthew P. Padula, Angelo Lopez, Peter J. Ralph, and Mathieu Pernice

Submitted to the journal "Cells" on the 12th of February 2020 and accepted with minor revisions on the 16th of February

Chapter 2:

Proteomic analysis of *Chlamydomonas reinhardtii* strain "UVM4" reveals molecular reprogramming related to enhanced transgene expression

Lorenzo Barolo, Audrey S. Commault, Raffaela M. Abbriano Matthew P. Padula, Unnikrishnan Kuzhiumparambil, Mikael Kim, Peter J. Ralph and Mathieu Pernice

Submitted to the journal "Journal of Proteome Research" on the 20th of February 2020

Another article was submitted in association with my PhD, however it does not form a part of this thesis:

Effect of biphasic temperature regime on therapeutic recombinant protein in the green alga *Chlamydomonas reinhardtii*

Audrey S. Commault, Navpreet Kaur Walia, Michele Fabris, Lorenzo Barolo, Jack Adriaans, Peter J. Ralph, and Mathieu Pernice.

Submitted to the journal "Algal Research" on the 14th of February 2020

TABLE OF CONTENT

CERTIFICATE OF ORIGINAL AUTHORSHIP ii
ACKNOWLEDGEMENTS iii
PREFACE iv
LIST OF PUBLICATIONS INCLUDED IN THE THESIS iv
TABLE OF CONTENT vi
LIST OF FIGURES AND TABLES ix
THESIS ABSTRACT 1
CHAPTER 1: Perspectives for glyco-engineering of recombinant biopharmaceuticals from microalgae 2
1. ABSTRACT 4
2. INTRODUCTION 5
3. PRODUCTION OF BIOPHARMACEUTICALS IN ALTERNATIVE HOSTS 7
4. MICROALGAL BIOFACTORIES 7
5. POST-TRANSLATIONAL MODIFICATIONS AND GLYCOSYLATION 11
5.1. *N*-glycosylation 12
5.2. *O*-glycosylation 16
5.3. Computationally predicted distribution of microalgal *N*- and *O*- glycosylation enzymes 19
6. STRATEGIES FOR MANIPULATING PROTEIN GLYCOSYLATION 22
6.1. Protein engineering 24
6.1.1. Glycoprotein sequence engineering 24
6.1.2. Subcellular location engineering 25
6.1.3. Glycosylation pattern engineering 26
6.2. Cell glyco-engineering 26
6.2.1. Glyco-engineering by inhibitor interference 27
6.2.2. Genetic glyco-engineering 27
7. FUTURE PERSPECTIVE FOR GLYCO-ENGINEERING IN MICROALGAE 29
8. REFERENCES 32
CHAPTER 2: Proteomic analysis of *Chlamydomonas reinhardtii* strain "UVM4" reveals molecular reprogramming related to enhanced transgene expression 51
1. ABSTRACT 52
2. INTRODUCTION 53
3. MATERIALS AND METHODS 55
4. RESULTS AND DISCUSSION 60
4.1. Cell biology 60
4.1.1. Cell morphology and motility 60

4.1.2. Growth and acetate uptake 61
4.2. Proteomic analysis 62
4.2.1. Protein selection and annotation 62
4.2.2. CRY1-1 and ARG7 gene expression 64
4.2.3. Growth and photosynthesis 65
4.2.4. Cell wall 68
4.2.5. Nucleic acid processing, transcription, translation, and protein degradation 71
4.2.6. Post-translational modifications (PTMs) 78
5. CONCLUSIONS 79
6. REFERENCES 83
CHAPTER 3: Transgene expression elicits widespread proteomic reprogramming in *Chlamydomonas reinhardtii* 90
1. ABSTRACT 91
2. INTRODUCTION 92
3. MATERIALS AND METHODS 95
4. RESULTS AND DISCUSSION 99
4.1. Low transformation success rates and gene silencing mechanisms still impair genetic engineering in *C. reinhardtii* strain UVM4 99
4.2. Strain UMV4 is more affected by position effect than strain 137c 101
4.3. Proteomic analysis 105
4.3.1. Nuclear transformation negatively affects cytoskeleton and morphogenesis in both strains 106
4.3.2. Nuclear transgene insertion and expression induce chromatin remodelling 107
4.3.3. Genetic engineering enhances production of ribosomal proteins but decreases translation initiation and elongation 109
4.3.4. Chaperons and transporters show general lower abundance, particularly in strain UVM4 112
5. CONCLUSIONS 115
6. SUPPLEMENTARY FIGURES 118
7. REFERENCES 120
CHAPTER 4: Comparative glycomic analysis reveals differences in protein glycosylation between *Chlamydomonas reinhardtii* strains 137c and UVM4 126
1. ABSTRACT 127
2. INTRODUCTION 128
3. MATERIALS AND METHODS 131
4. RESULTS & DISCUSSION 134
4.1. *C. reinhardtii* growth 134

4.2. Strains 137c and UVM4 have different glycoprotein profiles 134
4.3. Glycosylation analysis shows immunogenic residues in both strain 137c and UVM4 135
4.3.1. Similarities and differences in glycosylation status between strains 137
4.3.2. Strain 137c shows a putative "complex" glycoform 140
5. CONCLUSION 143
6. REFERENCES 144
CHAPTER 5: Synthesis, limitations, future perspectives, and concluding remarks 147
1. SUMMARY OF THESIS 148
2. SYNTHESIS 150
2.1. Comparative proteomic analysis of *C. reinhardtii* strains 137c and UVM4 (Chapter 2) 150
2.2. Comparative proteomic analysis of C. reinhardtii strains 137c and UVM4 after genetic engineering (Chapter 3) 151
2.3. Glycosylation comparison of strains 137c and UVM4 (Chapter 4) 153
3. LIMITATIONS 155
4. FUTURE PERSPECTIVES AND CONCLUDING REMARKS 161
5. REFERENCES 164

LIST OF FIGURES AND TABLES

Table 1.1. Twenty years of recombinant biopharmaceutical production in microalgae. 9

Figure 1.1. The different *N*-glycosylation representative patterns among (A) humans, (B) established biofactories including yeasts, plants, CHO cells, and (C) specific *N*-glycosylation patterns in microalgae. Differences in specificity of yeast, CHO, plant, and microalgal Golgi glycosyltransferases and glycosidases lead to variations in the final glycosylation profiles compared to humans. Consequently, glycans *N*-linked to recombinant proteins produced in these biofactories differ from native human proteins, necessitating glycan engineering to produce efficient and safe biopharmaceuticals in these alternative host systems. Green circle = Mannose. Blue circle = Glucose. Yellow circle = Galactose. Blue square = GlcNAc. White diamond = Neu5Gc. Me = Methylated residue. Fuchsia diamond = Sialic acid. Purple star = Xylose. Red triangle = Fucose. α-Man I = α-mannosidase I. GnT I = *N*-acetylglucosaminyltransferase I. α-Man II = α-mannosidase II. GnT II = *N*-acetylglucosaminyltransferase II..... 14

Figure 1.2. A comparison of the different *O*-glycosylation patterns among humans, CHO cells, yeasts, plants, and microalgae. Experimental evidence for *O*-glycosylation in microalgae is limited to *C. reinhardtii* [115]. Native *C. reinhardtii* proteins possess a (Hyp-*O*-Ara-Ara) core and methylated residues, characteristics that differ significantly from human *O*-glycosylation patterns. Ser/Thr = serine or threonine. Hyp = hydroxyproline. Green circle = Mannose. Yellow circle = Galactose. Orange pentagon = Arabinose. Blue square = GlcNAc. Me = Methylated residue. Purple star = Xylose. Red triangle = Fucose. 18

Figure 1.3. Hypothetical presence of protein *N*-glycosylation and *O*-glycosylation enzymes in the representative microalgal species, *P. purpureum* (PP), *P. tricornutum* (PT), *Nannochloropsis gaditana* (NG), *C. reinhardtii* (CR), *B. braunii* (BB), and *C. vulgaris* (CV), compared with glycosylation enzymes from *H. sapiens* and *A. thaliana*. Enzymes are classified as *present* (dark blue)*, potentially present* (light blue) or *missing* (grey). The genome assemblies of genomes *C. reinhardtii* (GCA_000002595), *P. purpureum* (GCA_000397085), *P. tricornutum* (GCF_000150955)*, B. braunii* (GCA_002005505), *C. vulgaris* (GCA_001021125), and *N. gaditana* (GCA_000569095; GCF_000240725; GCA_001614215), were downloaded from the National Center for Biotechnology Information (NCBI). Potential *Open Reading Frames* (ORF) sequences were predicted on all assemblies using EMBOSS' *getorf* command [116]. Gene candidates were identified using the Basic Local Alignment Tool (BLAST) [117] searching all predicted ORFs for a list of candidate genes. Additionally, template genes were searched for functional domains using PFAM [118] and HMMER v3 [119]. Hit-domain motifs were downloaded and search for in the predicted ORFs. All candidate ORFs determined with BLAST and Pfam domain searches were finally searched and aligned on the NCBI BLASTp webpage. Sequences were classified as present if the candidate ORFs returned good hits (e-value $<1^{-05}$) to proteins of the same or closely related organism annotated with the exact function searched for. Sequences were classified as *potentially present* if: (1) only closely related proteins were identified using BLAST, but specific PFAM domains were present in the genome; (2) BLASTp returned good hits (e-value $<1^{-05}$) to closely related functional proteins in any other organism. All other genes were classified as missing. The glycoprofiles were hierarchically clustered using seaborn cluster map with Euclidean method on Python. 20

Figure 1.4. A schematic of different glyco-engineering strategies; (A) protein engineering and (B) cell engineering. Protein glyco-engineering approaches (A) can target (1) the recombinant DNA sequence (rDNA), (2) the sub-cellular location of the biopharmaceutical, and (3) the glycosylation pattern of the translated protein. Cell glyco-engineering strategies (B) can modify the activity of glycosylation

enzymes by (1) random genetic insertion, (2) targeted gene knock-in or knock-out, and (3) inhibitor interference. Green circle = Mannose. Blue circle = Glucose. Yellow circle = Galactose. Blue square = GlcNAc. Fuchsia diamond = Sialic acid. Yellow star = Xylose. Red triangle = Fucose..... 24

Figure 2.1. Optical microscope observations (magnification 100X) of strain 137c and strain UVM4. Scale bar: 10 µm..... 60

Figure 2.2. Comparison of 137c and UVM4 strains growth (cell density (A) and optical density (B)), and acetate consumption (C). Acetate intake was assessed by measuring the concentration of acetate in spent media over time. Data are mean ± SEM (n=3). *Significant differences between the two strains (Fisher's LSD, $p < 0.05$). 61

Figure 2.3. Venn diagrams and volcano plots of the common proteins in intracellular (A) and extracellular (B) proteomes comparison. The number of proteins in each category is given in brackets. Out of the common proteins, 227 and 297 proteins were found to be significantly different in the intracellular and extracellular proteomes, respectively. Differentially produced proteins were selected based on significance value ≥ 20 and fold change of $-1 \geq \log_2 \geq 1$. 64

Figure 2.4. Relative abundance of strain UVM4 intracellular proteins involved in photosynthetic mechanisms, using strain 137c protein abundance as reference. Proteins with a relative abundance > 1 are considered more abundant in strain UVM4 than in strain 137c (red) and vice-versa for relative abundances < -1 (blue), n = 3..... 67

Figure 2.5. Relative abundance of cell wall proteins in strain UVM4 using protein abundance in strain 137c as reference. The proteins showing higher abundance in strain UVM4 than in 137c are in red, while the less abundant proteins are in blue, n = 3..... 69

Figure 2.6. Relative abundance of strain UVM4 intracellular (A, B, C) and extracellular (D) proteins in comparison with strain 137c proteins (used as reference) involved in: (A) DNA accessibility, binding, and replication; (B) RNA binding, processing, and modification; (C) DNA transcription; (D) protein degradation. The proteins showing higher abundance in strain UVM4 than in strain 137c are in red, while the less abundant proteins are in blue, n = 3..... 73

Figure 2.7. Schematic diagram summarizing key regulations observed by comparative proteomics in this study, indicating molecular mechanisms potentially driving phenotypic differences and improved recombinant protein production in strain UVM4. 80

Figure 3.1. Dot plots of fluorescence-activated cell sorting (FACS) for transgenic cells of strains 137c and UVM4. The positive cells showing expression of the transgene were gated in green based on wild type cells fluorescence. Cells included in the "mVenus" yellow gate were selected for single cell sorting. 99

Figure 3.2. Percentage of mVenus positive cells in 11 transgenic lines of strain 137c (A) and 18 transgenic lines of strain UVM4 (B), gated based on their respective wild type behaviour. The data was obtained by flow cytometry analysis after 72 hours of growth under antibiotic selective pressure. 100

Figure 3.3. Comparison of growth profiles for wild type and transformant lines expressing mVenus for strains A) 137c and B) UVM4. Cells were grown in 500 mL of antibiotic-free media. 102

Figure 3.4. Normalised fluorescence in transgenic cell lines of strain 137c (A) and strain UVM4 (B) over time. Cells were grown in 500 mL of antibiotic-free media. Positive cells were detected based on wild type background fluorescence. Normalised fluorescence was calculated using equation (Eq1). 104

Table 3.1. Relative abundance of cytoskeleton proteins and proteins involved in morphogenesis after comparison of the transgenic lines with their respective wild type strain (n = 5). ND = Not Detected.....106

Table 3.2. Relative abundance of proteins involved in DNA binding and chromatin remodelling after comparison of the transgenic lines with their respective wild type strain (n = 5). ND = Not Detected. 108

Table 3.3. Relative abundance of ribosomal proteins and proteins involved in translation initiation and elongation. Protein relative abundance was determined by comparing the transgenic lines with their respective wild type strain (n = 5). ND = Not Detected. 110

Table 3.4. Relative abundance of proteins involved in protein folding and transport. Protein relative abundance was determined by comparing the transgenic lines with their respective wild type strain (n = 5). ND = Not Detected.....112

Figure 3.5. Schematic of altered pathways in *C. reinhardtii* strains 137c and UVM4 after transgene insertion and expression.....116

Figure 4.1. Extracellular proteome gel, stained with (A) Coomassie and (B) glycoprotein staining. Sample loaded: ~100 µg total protein, n=3. Intense staining of high molecular weight bands in both Coomassie and glyco-stained gel suggests overabundance of glycosylated extracellular proteins in strain UVM4 relative to strain 137c. Proteins with molecular weight lower than 50 kDa did not show detectable differences. Red arrows indicate the most intense bands detected.....135

Figure 4.2. Mass spectra of glycan population of extracellular proteins of (A) strain 137c and (B) strain UVM4. Green circle = Mannose. Yellow circle = Galactose. Blue square = GlcNAc. Yellow star = Xylose. Red triangle = Fucose. The relative peak abundance was calculated by normalizing the intensity of each peak to the intensity of the highest peak, which occurs at 1205.62 m/z ratio. 137

Figure 4.3. Immunogenic glycoforms identified for strains 137c and UVM4. Green circle = Mannose. Yellow circle = Galactose. Blue square = GlcNAc. Yellow star = Xylose. Red triangle = Fucose. A tick symbol (✓) indicates presence of the glycoform in the strain mass spectra. 137

Figure 5.1. Schematic diagram showing the altered molecular processes detected in strain UVM4. 151

Figure 5.2. Simplified schematic of protein production pathways in *C. reinhardtii*, showing the common pathways altered due to transgene expression in strains 137c and UVM4 152

Figure 5.3. Immunogenic glycans detected in strains 137c and UMV4. Green circle = Mannose; Blue square = *N*-Acetylglucosamine; Red triangle = Fucose; Yellow star = Xylose. 154

Figure 5.4. SDS-PAGE gel showing the extracellular proteomes of 137c and UVM4 strains stained with (A) Coomassie and (B) glycoprotein staining. Selected bands are highlighted (in red).....158

Table 5.1. Identified proteins from strain UVM4 gel band 1.....159

Table 5.2. Identified proteins from strain UVM4 gel band 2.....159

THESIS ABSTRACT

Chlamydomonas reinhardtii is a eukaryotic unicellular green microalga historically used as a model organism to describe and analyse fundamental biological processes including photosynthesis. Recently, this species has been utilised as a biofactory to successfully produce recombinant therapeutic proteins. *C. reinhardtii* has many advantages over traditional biofactories, such as *E. coli* (bacteria) and Chinese Hamster Ovary cells (CHO, mammalian). It has high growth rates at low production costs, it cannot be contaminated by human pathogens, it can effectively secrete recombinant proteins, and it possess a eukaryotic post-translational modification (PTM) machinery. Unfortunately, *C. reinhardtii* also displays two major disadvantages: recombinant protein yields can be low and glycosylation (a fundamental PTM) can be incorrect. Low yields and potential low quality of products are the only drawbacks that kept *C. reinhardtii* out of the recombinant biopharmaceutical market, now worth 140 billion US$.

The low recombinant protein yield issue was partially overcome in 2009 with the generation of a UV mutated strain called UVM4. This strain is now well-established and broadly-used for secreted recombinant protein production in *C. reinhardtii*, and is capable of yields up to 15 mg/L (3-fold higher than the non-mutated strain 137c). However, these yields are still far from the extensive ones obtained with CHO cells (up to 5 g/L). Interestingly, as frequently happens for strains generated by mutagenesis, the pathways altered by the mutation were not investigated. Therefore, the reasons for these higher recombinant protein yields produced by strain UVM4 are still unknown. Characterising the modified protein pathways in this strain might help to understand the causes for the general lower yields in *C. reinhardtii*, to subsequently optimise and finally completely overcome the issue. In addition, to fully validate strain UVM4 as a cell biofactory, it is also necessary to analyse recombinant protein quality, namely glycosylation. Incorrect glycosylation can lead to immunogenic biopharmaceuticals, therefore a complete glycosylation profiling of strain UVM4 is also required.

With the results obtained in my thesis, I provide a detailed analysis of the two major drawbacks in recombinant protein production in *C. reinhardtii*, unravel possible causes, provide potential solutions, and overall I corroborate this species as a future industrial cell biofactory.

CHAPTER 1: Perspectives for glyco-engineering of recombinant biopharmaceuticals from microalgae

Lorenzo Barolo[1], Raffaela M. Abbriano[1], Audrey S. Commault[1], Jestin George[1], Tim Kahlke[1], Michele Fabris[1,2], Matthew P. Padula[3], Angelo Lopez[4], Peter J. Ralph[1], and Mathieu Pernice[1]

AFFILIATIONS

[1] University of Technology Sydney, Climate Change Cluster, Broadway Campus, Ultimo NSW 2007, Sydney, Australia

[2] CSIRO Synthetic Biology Future Science Platform, Brisbane, Qld 4001, Australia

[3]School of Life Sciences and Proteomics Core Facility, Faculty of Science, University of Technology Sydney, Ultimo NSW 2007, Sydney, Australia

[4]Thomas-Oates Mass Spectrometry Group, Department of Chemistry, University of York, York, YO10 5DD, England

CORRESPONDING AUTHORS

Mathieu Pernice

University of Technology Sydney, Climate Change Cluster, Broadway Campus, Ultimo NSW 2007, Sydney, Australia

Email: mathieu.pernice@uts.edu.au

Lorenzo Barolo

University of Technology Sydney, Climate Change Cluster, Broadway Campus, Ultimo NSW 2007, Sydney, Australia

Email: lorenzo.barolo@student.uts.edu.au

KEYWORDS: glycosylation, microalgae, recombinant proteins, biopharmaceuticals

CONTRIBUTIONS: LB wrote the manuscript with contributions from all authors.

This article was submitted on the 12th of February 2020 to the journal “Cells” and accepted with minor revisions on the 16th of February

1. ABSTRACT

Microalgae exhibit great potential for recombinant therapeutic protein production, due to lower production costs, immunity to human pathogens, and advanced genetic toolkits. However, a fundamental aspect to consider for recombinant biopharmaceutical production is the presence of correct post-translational modifications. Multiple recent studies focusing on glycosylation in microalgae have revealed unique species-specific patterns absent in humans. Glycosylation is particularly important for protein function and is directly responsible for recombinant biopharmaceutical immunogenicity. Therefore, it is necessary to fully characterise this key feature in microalgae before these organisms can be established as industrially relevant microbial biofactories. Here, I review the work done to date on production of recombinant biopharmaceuticals in microalgae, experimental and computational evidence for *N*- and *O*-glycosylation in diverse microalgal groups, established approaches for glyco-engineering, and perspectives for their application in microalgal systems. The insights from this introduction may be applied to future glyco-engineering attempts to humanize recombinant therapeutic proteins and to obtain potentially cheaper, fully functional biopharmaceuticals from microalgae.

2. INTRODUCTION

Biopharmaceuticals are biological macromolecules that exhibit therapeutic actions in humans. This group of compounds includes essential molecules such as antibodies, hormones, and vaccines [1]. More than 60% of commercialized biopharmaceuticals are recombinant proteins [2], which are produced in genetically engineered host cells defined as biofactories [3]. Biofactories can include a broad spectrum of host organisms, spanning from prokaryotic to eukaryotic cells, and support a US$140 billion biopharmaceuticals market [4].

Currently, the host cell systems *Escherichia coli* (bacteria), *Saccharomyces cerevisiae* and *Pichia pastoris* (yeast), and Chinese hamster ovary (CHO) cells (mammalian) dominate the production of biopharmaceuticals, with the mammalian CHO cells covering >50% of the market [5]. However, these four cell systems have several disadvantages, including species-specific issues related to the nature of each biofactory. *E. coli* often incurs translational errors, accumulates inclusion bodies, and completely lacks the eukaryotic organelles and machinery necessary to produce fundamental post-translational modifications (PTMs) [6]. Challenges with using yeasts include inadequate protein secretion and incorrect protein PTMs [7,8]. Biopharmaceutical production with CHO cells is expensive because of the complex culturing requirements associated, difficult to scale, and susceptible to contamination with human viruses and prions [9,10]. These complications have prompted efforts to optimize these systems as well as identify more suitable host cell lines. In this context, microalgae have emerged as attractive novel expression systems for biopharmaceutical production.

Chlamydomonas reinhardtii (Chlorophyceae) has historically been the model microalgal species for biotechnological innovations [11], due to its tractability in the laboratory [12]. However, other species are increasingly being evaluated for their capacity to produce recombinant proteins. Microalgae including *Chlorella sp.* (Trebouxiophyceae) [13,14], *Dunaliella salina* (Chlorophyceae) [15], *Nannochloropsis oculata* (Eustigmatophyceae) [16,17] and the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) [18] have been reported to successfully express biopharmaceuticals.

A major challenge with producing biopharmaceuticals in non-human cells is obtaining correct PTMs of the recombinant protein. Since each host cell system possesses its own unique protein processing machinery at a translational and post-translational level, the same recombinant protein produced in different cell systems can display different and unique PTMs [19,20,21,22,23]. PTMs such as phosphorylation, acetylation, methionine oxidation, asparagine and glutamine deamination, disulphide bond formation, and glycosylation all play a significant role in stability, functionality, and activity of proteins [24]. Among these, glycosylation is a major PTM that is found on more than 50% of human proteins [25]. In biopharmaceutical production, over 40% of approved therapeutic proteins are glycosylated, highlighting the importance of understanding and controlling the glycosylation mechanism of non-human expression systems [26,27].

The cellular mechanisms that determine protein glycosylation patterns are complex and vary among different eukaryotic species [28]. Given the significant influence of glycosylation on yield, efficacy, pharmacokinetics, and immunogenicity of recombinant therapeutic proteins, it is essential to choose the right host expression system to successfully produce a functional biopharmaceutical [29]. Additionally, sub-optimal glycosylation of recombinant proteins can be overcome using glyco-engineering strategies [30]. Glyco-engineered biopharmaceuticals will present “humanized” glycans that will not be immunogenic to humans [4].

In this introduction, I provide a brief overview of various host systems, with a focus on the advantages and disadvantages of microalgae as biofactories for the production of recombinant therapeutic proteins. I also discuss strategies to overcome the challenges, with a focus on microalgae glycosylation status and its comparison to human glycosylation. Lastly, I assess the prospect of applying glyco-engineering techniques to optimize recombinant biopharmaceutical production in microalgal host systems.

3. PRODUCTION OF BIOPHARMACEUTICALS IN ALTERNATIVE HOSTS

Due to improvements in genetic engineering technologies and recombinant protein expression, the repertoire of biopharmaceuticals produced in biofactories is expanding. Increased capability to replicate the functional characteristics of larger, more complex proteins [31] has broadened the range of potential biopharmaceutical applications, including treatments for cancer and autoimmune diseases [32]. This is particularly evident for antibodies [31], which have been replicated in various forms including single-chain fragment variables (scFvs) [33] and antigen-binding fragments (Fabs) [34], amongst others. The choice of biofactory is primarily based on the type of biopharmaceutical (e.g. functional multi-domain mammalian proteins cannot be produced in bacteria as they require specific PTMs), but other factors such as production costs, yields, time to market, and the safety of the patient are also considered when selecting an expression host. The recent production of a biopharmaceutical used as a treatment for the Ebola virus in the Democratic Republic of the Congo has proven that plants (in this case *Nicotiana benthamiana*) can be a powerful alternative to common expression systems, especially for rapid and low-cost production of antibodies using technologies that are relatively simple to implement in developing countries [35]. Among the multitude of alternative expression hosts, photosynthetic systems have emerged with high potential to offer a rapid and cost-effective alternative to traditional hosts. The previously described traditional biofactories are all heterotrophic systems and rely on the addition of organic carbon sources to the culture media, whilst photosynthetic organisms can generate biomass using CO_2 and light, lowering not only carbon footprint but also media complexity and costs associated with recombinant protein production. In recent decades, plant cells and microalgal organisms have been analysed for their promising role as photosynthetic host systems, with the latter being the focus of this introduction and this thesis overall.

4. MICROALGAL BIOFACTORIES

Microalgae are unicellular photosynthetic organisms found in both terrestrial and aquatic environments [36]. The potential of microalgae to produce natural products with commercial value, ranging from food additives and health supplements to biofuels and cosmetics, has currently become widely recognized

[37]. Recent advances in genetic engineering have enabled expression of many recombinant biopharmaceuticals in microalgal hosts (Table 1.1).

Microalgal biofactories possess many unique advantages. Like plants, microalgae can be cultivated in large areas and require a lower up-front investment compared to bacterial or mammalian cell systems [38]. Additionally, microalgae do not harbor human pathogens and some species are generally recognized as safe (GRAS) [12]. They are easy to cultivate in bioreactors, and thus are less prone to airborne contamination issues [12]. Moreover, microalgae have higher growth rates and less complex media requirements compared to plants, and therefore potentially lower production costs [10,11,12]. Furthermore, microalgae are unicellular organisms that lack functional parts such as roots and leaves, making recombinant biopharmaceuticals production more uniform within the same batch and diminishing the energy and resources needed to generate additional biomass [12].

One of the better characterized and widely exploited microalgae for recombinant protein production is the biflagellate Chlorophyta *C. reinhardtii* [39]. This species has become the model algae to explore and understand the biological processes occurring within the green microalgal lineage due to an advanced genetic toolkit and the availability of fully sequenced nuclear and organellar genomes [12]. *C. reinhardtii* is now the most comprehensive microalgal platform for expression of recombinant proteins with promising industry applications in the bioenergy, biopharmaceutical, biomaterial, and nutraceutical sectors [11,40]. A recent toxicology study demonstrated that *C. reinhardtii* biomass is safe for human consumption [41], and although *C. reinhardtii*-derived products are not yet commercially available, extensive research on large-scale cultivation is bringing it closer to reality [42].

Recombinant protein production in *C. reinhardtii* has been achieved by engineering both the chloroplast and nuclear genomes. In the chloroplast, the level of transgenic expression can reach 20% of total soluble protein (TSP) [10]. However, recombinant proteins expressed in the chloroplast are retained inside the plastid and cannot be secreted [43]. Therefore, recombinant proteins expressed in the chloroplast are unable to pass through the secretion route via the endoplasmic reticulum and the Golgi

apparatus, and cannot be subjected to fundamental PTMs such as glycosylation [12,43]. Hence, chloroplast expression is only suitable for non-glycosylated proteins. Recombinant proteins expressed from the nucleus, on the other hand, can be targeted to the secretion route by adding specific signal peptides to the recombinant amino acid sequence, resulting in secretion and glycosylation of the recombinant protein [10]. Unfortunately, nuclear expression results in very low yields, due to random integration, low transformation efficiency, and gene silencing mechanisms [43,44].

As increasing knowledge and genetic tools become available, other microalgae species have been exploited for recombinant protein production, including monoclonal antibodies, hormones, and enzymes. Several species, including *Chlorella vulgaris, Chlorella sorokiniana*, *Chlorella ellipsoidea*, *D. salina, N. oculata*, and the diatom *P. tricornutum* have been successfully used to express biopharmaceuticals (Table 1.1). In particular, *P. tricornutum* has gained noticeable importance due to the ability to secrete fully functional IgG antibodies [18]. An overview of the recombinant biopharmaceuticals produced in microalgae over the last two decades is presented in Table 1.1.

Table 1.1. Twenty years of recombinant biopharmaceutical production in microalgae.

Organism	**Organelle**	**Protein**	**Reference**
C. reinhardtii	Chloroplast	E7 of HPV-16	Demurtas et al 2013[45]
		D2-CTB	Dreesen et al 2010[46]
		α-galactosidase	Georgianna et al 2013[47]
		Phytase	Georgianna et al 2013[47]
		Xylanase	Georgianna et al 2013[47]
		Pfs25	Gregory et al 2012[48]
		Pfs28	Gregory et al 2012[48]
		Pfs25-CTB	Gregory et al 2013[49]
		E2	He et al 2007[50]
		Pfs48/45	Jones et al 2013[51]
		M-SAA	Manuell et al 2007[52]
		Anti-HSV glycoprotein D Isc	Mayfield et al 2003[53]
		12FN3	Rasala et al 2010[54]
		Erythropoietin	Rasala et al 2010[54]
		HMGB1	Rasala et al 2010[54]

		Interferon β	Rasala et al 2010[54]
		Proinsulin	Rasala et al 2010[54]
		SAA-10FN3	Rasala et al 2010[54]
		VEGF	Rasala et al 2010[54]
		Allophycocyanin	Su et al 2005[55]
		VP1-CTB	Sun et al 2003[56]
		V28	Surzycky et al 2009[57]
		Anti-PA 83 anthrax IgG1	Tran et al 2009[58]
		Anti-CD22-gelonin sc	Tran et al 2013a[59]
		Anti-CD22-ETA sc	Tran et al 2013b[60]
		GAD65	Wang et al 2008[61]
		TRAIL	Yang et al 2006[62]
		Phytase (AppA)	Yoon et al 2011[63]
		Metallothionein-2	Zhang et al 2006[64]
C. reinhardtii	Nucleus	Human Epidermal Growth Factor	Baier et al. 2018[65]
		VEGF-165	Chavez et al 2016[66]
		GBSS-AMA1	Dauvillee et al 2010[67]
		GBSS-MSP1	Dauvillee et al 2010[67]
		Erythropoietin	Eichler-Stahlberg et al 2009[39]
		Sep-15	Hou et al 2013[68]
		Lolium Perenne IBP	Lauersen et al 2013[69]
		β-1,4-endoxylanase	Rasala et al 2012[70]
C. vulgaris *C. sorokiniana*	Nucleus	Human growth hormone	Hawkins and Nakamura 1999[13]
C. ellipsoidea	Nucleus	mNP-1	Bai et al 2013[14]
		NP-1	Chen et al 2001[71]
		Flounder growth hormone	Kim et al 2002[72]
D. salina	Chloroplast	α-galactosidase	Georgianna et al 2013[47]
		Phytase	Georgianna et al 2013[47]
		Xylanase	Georgianna et al 2013[47]
D. salina	Nucleus	V28	Feng et al 2014[73]
		HBsAg	Geng et al 2003[15]
P. tricornutum	Nucleus	Anti-Hepatitis B IgG	Hempel and Maier 2012[18]
		Anti-MARV NP IgG	Hempel et al. 2017[74]
N. oculata	Nucleus	Bovine lactoferricin (LFB)	Li et al. 2009[17]
	Nucleus	Flounder growth hormone	Chen et al. 2008[16]

5. POST-TRANSLATIONAL MODIFICATIONS AND GLYCOSYLATION

Post-translational modifications (PTMs) are chemical modifications of a protein during or after its synthesis within the cell. Amino acids, the building blocks that define the physico-chemical structure and functionality of a protein, can be altered by more than 250 different PTMs [75]. The repertoire of PTMs is very complex, considering that 15 out of the 20 common amino acids can be modified [76]. More than 5% of the total human genome encodes for enzymes involved in PTMs, including those involved in phosphorylation (kinases and phosphatases), acetylation (acetylases and deacetylases), and glycosylation (glycosyltransferases) [26]. Protein PTMs, together with alternative RNA splicing and translation, enhance molecular diversification of gene products and participate in a complex system to regulate the physiology of eukaryotic cells [26,77].

PTMs such as phosphorylation, glycosylation, and nitration are involved in important cellular processes [78]. For example, phosphorylation can operate as a switch to modulate specific catalytic activities of proteins [79]. Another function of PTMs is to mark proteins for degradation by ubiquitination [79]. PTMs can also act in response to external stimuli, for example, when a cell is subjected to biological stress it can activate proteins with specific PTMs to counteract the stress [80]. Of the many PTMs relevant to biopharmaceutical production, glycosylation plays a major role. In fact, glycosylation is found on more than 50% of human proteins [25], and on more than 40% of approved recombinant biopharmaceuticals, highlighting the importance to understand and control the glycosylation mechanism of non-human expression systems [26,27].

Glycosylation refers to a covalent bond between a polysaccharide chain and an amino acid, formed during translation of the protein. The two most frequent types of glycosylation are *N*-linked and *O*-linked glycosylation. *N*-linked glycosylation is characterized by the formation of a covalent bond between the glycan and the amidic group of an asparagine (Asn) residue. *O*-linked glycosylated proteins have the glycan linked to the hydroxyl component of a serine (Ser) or a threonine (Thr) residue. Glycosylated proteins and glycan structures strongly regulate fundamental biological processes within

the cell, such as cell adhesion, self/nonself recognition, molecular trafficking and clearance, receptor activation, and endocytosis [81].

In addition, glycosylation significantly enhances yield, efficacy, and pharmacokinetics of recombinant biopharmaceuticals [29]. During recombinant protein production, non-human host organisms can attach glycan residues (monosaccharides) that would be absent on the human endogenous protein, potentially resulting in lower yields and efficacy [10]. Recombinant therapeutics presenting incorrect or no glycosylation can trigger an immune response in the patient resulting in accelerating clearance during therapy or, in some rare cases, life threatening complications [82]. Glycans can trigger an immunogenic reaction either by direct recognition of the glycan sequence by the patient immune system [83] or by affecting the therapeutic protein folding, solubility (e.g., formation of aggregates) and structural stability [84], breaking the immune tolerance of the patient. At least four non-human residues have been identified as being able to induce an immune response in humans. These residues are: *N*-glycolylneuraminic acid (Neu5Gc), galactose-α(1,3)-galactose (α-Gal), β(1,2)-xylose, and α(1,3)-fucose [84]. The α-Gal and Neu5Gc residues are present in therapeutics produced in mammalian cells such as CHO cells, while β(1,2)-xylose and α(1,3)-fucose are present in plant and microalgal-produced glycoproteins [84]. Therefore, it is essential to understand the glycosylation capabilities of a chosen biofactory to produce safe and effective recombinant therapeutics.

5.1. *N*-glycosylation

N-glycosylation requires a strict consensus sequence, where the amino acid sequence must be Asn-Xxx-Ser/Thr [85], where Xxx can be any amino acid other than Proline (Pro). Further studies expanded this sequence to a less frequent, but still relevant, Asn-Xxx-Cysteine (Cys), Asn-Xxx-Valine (Val), and Asn-Glycine (Gly) [86].

N-glycosylation begins in the endoplasmic reticulum (ER), where a biosynthetic precursor, a dolichol-P-P-linked oligosaccharide (comprised of 3 glucose (Glu), 9 mannose (Man) and 2 *N*-acetylglucosamine (GlcNAc) residues) is transferred to Asn residues in nascent polypeptide chains (Figure 1.1). The glycan

is then subjected to further enzymatic maturation as part of a quality control by chaperones (calnexin and calreticulin). The newly formed glycoprotein is then transferred to the Golgi apparatus and further modified by many different glycosyltransferases, until reaching final maturation. *N*-glycans across different eukaryotic organisms present a common structure called the "pentasaccharide core" (2 *N*-acetylglucosamine (GlcNAc) and 3 mannose (Man) residues) [10,87] (Figure 1.1). The rest of the residues will attach onto the final two Man of the pentasaccharide core, thus creating two polysaccharide antennae. However, unlike the well conserved pentasaccharide core, the rest of the *N*-glycan structure vastly varies amongst eukaryotes. In fact, the final maturation of the glycan in the Golgi apparatus by different glycosyltransferases is a species-specific mechanism and is the source of different glycosylation patterns among eukaryotic organisms [88,89] (Figure 1.1).

In humans, depending on how the two terminal Man residues are elongated, three different specific *N*-glycans structures are possible: i) "oligo-mannose" glycans containing only mannose residues, ii) "hybrid" glycans with mannose residues on one antenna and mixed monosaccharides on the second antenna, and iii) "complex" glycans with mixed monosaccharides on both antennae [90]. As a glycoprotein enters the Golgi, it is *N*-linked to 8/9 Man and 2 GlcNAc. This chain is reduced to 5 Man and 2 GlcNAc by an enzyme called α-mannosidase I (α-Man I) (Figure 1.1). Then, the *N*-acetylglucosaminyltransferase I (GnT I) plays a key role by transferring one *N*-acetylglucosamine residue on the α(1,3)-mannose arm of the 5 Man and 2 GlcNAc *N*-linked protein. This structure (1 GlcNAc, 5 Man and 2 GlcNAc) is the starting point for both "hybrid" and "complex" structures (Figure 1.1). From that point on, many different glycosyltransferases will build the complete pattern of the *N*-glycoprotein [91].

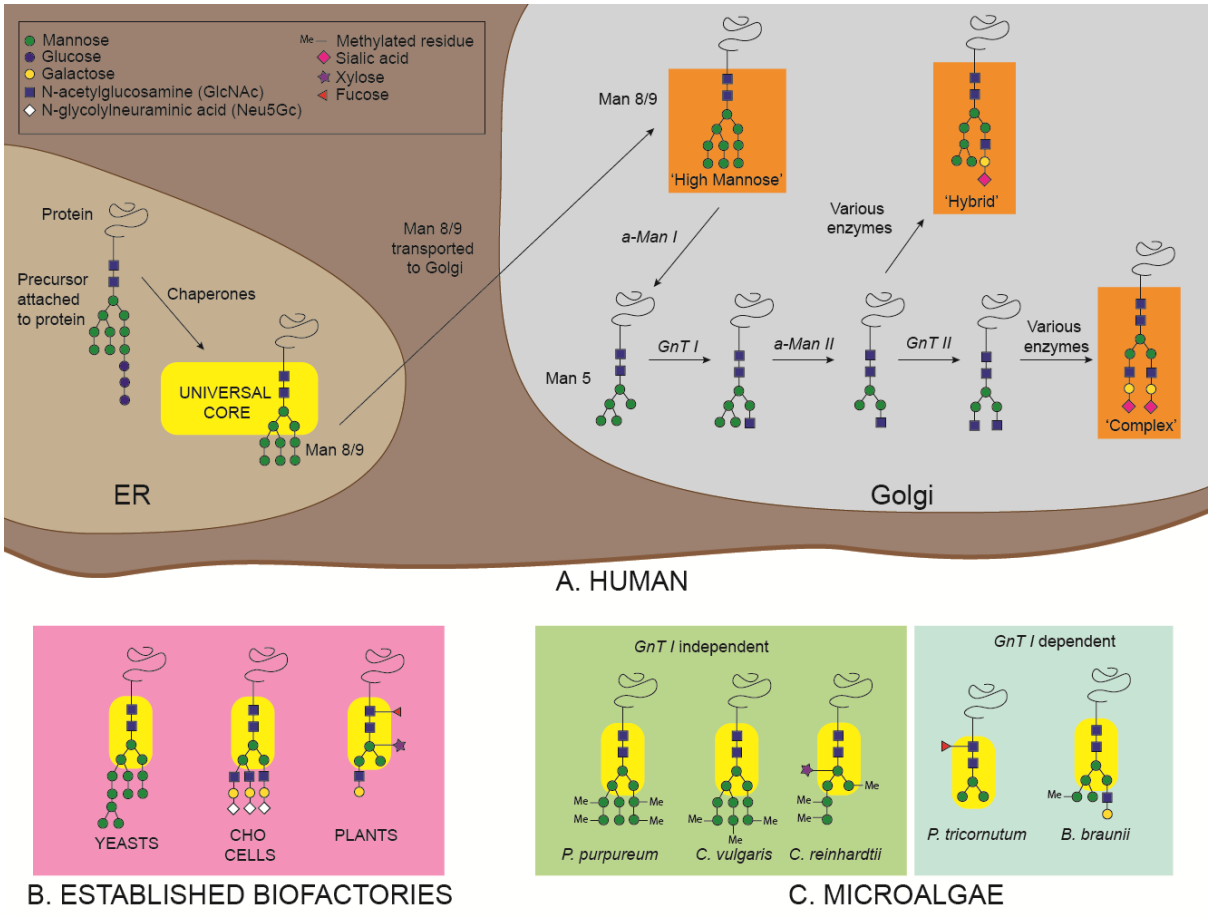


Figure 1.1. The different *N*-glycosylation representative patterns among (A) humans, (B) established biofactories including yeasts, plants, CHO cells, and (C) specific *N*-glycosylation patterns in microalgae. Differences in specificity of yeast, CHO, plant, and microalgal Golgi glycosyltransferases and glycosidases lead to variations in the final glycosylation profiles compared to humans. Consequently, glycans *N*-linked to recombinant proteins produced in these biofactories differ from native human proteins, necessitating glycan engineering to produce efficient and safe biopharmaceuticals in these alternative host systems. Green circle = Mannose. Blue circle = Glucose. Yellow circle = Galactose. Blue square = GlcNAc. White diamond = Neu5Gc. Me = Methylated residue. Fuchsia diamond = Sialic acid. Purple star = Xylose. Red triangle = Fucose. α-Man I = α-mannosidase I. GnT I = *N*-acetylglucosaminyltransferase I. α-Man II = α-mannosidase II. GnT II = *N*-acetylglucosaminyltransferase II.

Bacteria, yeast, CHO, and plant glycoproteins possess species-specific *N*-glycans that differ from human *N*-glycans (Figure 1.1B). Glycosylation and PTMs in general are challenging issues for *E. coli*, as most prokaryotes lack the eukaryotic PTM machinery to perform glycosylation [6]. It is possible to

transfer the basic prokaryotic glycosylation machinery of another bacteria (*Campylobacter jejuni*) into *E. coli* to obtain glycosylated recombinant biopharmaceuticals [92]. However, prokaryotic glycans do not show similarities with human glycosylation patterns, resulting in immunogenic biopharmaceuticals [92,93]. Yeast glycans present an excess of mannose residues assembled together in "hyper-mannosidic" structures [94], which greatly differ from the human patterns [4,95]. Although CHO cells possess human-like glycosylation machinery, some discrepancies still persist [4]. For example, the absence of fundamental human residues like α(2,6)-sialic acid and α(1,4)-fucose, and the production of undesired non-human residues such as Neu5Gc and α-Gal, can result in the production of potentially immunogenic recombinant biopharmaceuticals [4,32]. Similarly, plant cells produce glycans containing immunogenic residues such as β(1,2)-xylose and core α(1,3)-fucose [96].

Compared to other systems, very little is known about *N*-glycosylation in microalgae. The combination of genomic annotation and experimental evidence has revealed some details about *N*-glycans in some species; however, more information is needed. Analysis of five different microalgal species [83,97,98,99,100,101] showed two different glycosylation pathways based on presence or absence of the GnT I enzyme. The green microalgae *C. reinhardtii* and *C. vulgaris*, and the red microalga *Porphyridium purpureum* lack GnT I, which has been validated experimentally [83,97,98,101]. Thus, the *N*-glycosylation pathway in these species is defined as GnT I-independent. In this pathway, the 5 Man and 2 GlcNAc *N*-linked protein is subjected to the action of xylosyltransferases (XyT) and methyltransferases (MeT), leading to unique *N*-linked structures containing methylated mannoses linked to one or two xyloses (Figure 1.1). The structures vary slightly among these microalgae, with different possible locations of the xylose residues [83,97,98,101]. On the other hand, Baïet and colleagues [99] demonstrated that GnT I is present and active in the diatom *P. tricornutum*. In this case, GnT I transfers an *N*-acetylglucosamine residue to the 5 Man and 2 GlcNAc *N*-linked protein in the Golgi apparatus. The structure is then subjected to α-mannosidase II (α-Man II) and fucosyltransferase (FuT), resulting in paucimannosidic ($Man_{3–4}GlcNAc_2$) fucosylated *N*-glycans [99] (Figure 1.1C). The GnT I-dependent pathway is also present in the green microalga *Botryococcus braunii* [100]. *N*-linked glycans

in *B. braunii* present methylation of mannose residues (absent in *P. tricornutum*) and the terminal GlcNAc (linked to the α(1,3)-mannose arm) can be attached to an additional hexose [100] (Figure 1.1C).

Due to the presence of a eukaryotic PTM machinery and several different glycosidases, both GnT I-independent and GnT I-dependent microalgal species show *N*-glycosylation patterns more similar to humans than *E. coli* (glycosylation absent) and yeasts ("hyper-mannosidic" *N*-glycans). Nevertheless, discrepancies between human and microalgal *N*-glycans are still relevant. The absence of GnT I enzyme in GnT I-independent species is a significant issue. GnT I activity serves as starting point to produce "complex" and "hybrid" glycans, and its absence prohibits the construction of two out of three possible human *N*-glycan structures. Moreover, both GnT I-independent species listed here present abundant methylation of residues (absent in humans) and *C. reinhardtii* also presents attachment of Xyl residues, which is another immunogenic trait. Unsurprisingly, GnT I-dependent microalgae resemble human glycosylation more than GnT I-independent species. *B. braunii* shows native *N*-glycans similar to human "hybrid" structures, and *P. tricornutum* presents paucimannosidic glycans, an important pattern found in a specific class of biopharmaceuticals (as explained in Chapter 1, section *6.1.2*) [102,103]. Nonetheless, α(1,3)-fucose residues in *P. tricornutum* [104] and methylation of mannoses in *B. braunii* [100] are both *N*-glycan characteristics absent in humans.

5.2. *O*-glycosylation

O-glycosylation involves an oxygen-carbon bond between the hydroxyl group of a Ser or Thr residue of the protein and the polysaccharide chain (Figure 1.2). There are 7 subclasses of *O*-glycans, based on which monosaccharide is directly attached to the protein [105]. In humans, the most frequently observed *O*-glycoproteins are mucins and proteoglycans [96,105] (Figure 1.2). The first monosaccharide attached to a mucin protein is an *N*-acetylgalactosamine (GalNAc), usually followed by galactose (Gal) or a GlcNAc [96]. On the other hand, xylose is the first monosaccharide attached to a proteoglycan, followed by Gal [106]. Other important *O*-linked monosaccharides that can initiate the polysaccharide chain are Fuc and Man; *O*-fucosylation plays a fundamental role in transmembrane signalling [107,108] and *O*-mannosylation is involved in muscle and brain development [109]. While the *O*-GalNAcylation starts

directly in the Golgi apparatus, all the other structures begin in the ER and the whole polysaccharide is later synthesized in the Golgi, like for *N*-glycans synthesis [105].

In most organisms including humans, *O*-glycosylation does not present a common structure or a consensus sequence, and many different structures are possible. Nevertheless, unique species-specific *O*-glycans are still recognisable in some organisms. *E. coli* lacks the eukaryotic organelles and PTM machinery to perform *O*-glycosylation [6]. *O*-glycans in yeasts present a core structure composed of a Man residue attached to a Ser or a Thr [110]. This structure presents a core Man with multiple mannoses attached, resulting in yeast-specific hyper-mannose *O*-glycans [110] (Figure 1.2). The majority of *O*-glycans in CHO cells present a core structure with a GalNAc residue attached to a Ser or a Thr, the same core structure as human mucins [111]. The GalNAc residue can be further linked to a Gal or a GlcNAc residue [111,112] (Figure 1.2). *O*-glycosylation core structures in plants can present a Gal residue attached to a Ser or a unique arabinose (Ara) residue attached to a hydroxyproline (Hyp) amino acid [113]. Recombinant interferon (IFN) alpha 2b expressed in tobacco cells showed a unique *O*-glycan pattern with a Hyp-*O*-Gal core and several Ara and Gal residues [114].

To date, there is only one study reporting *O*-glycosylation analysis of proteins produced in microalgae. Bollig and colleagues [115] analysed linear hydroxyproline-bound *O*-glycans native proteins of the green alga *C. reinhardtii*, showing similarities and differences with higher plant *O*-glycans. They found the same *O*-glycoprotein core as in plants (Hyp-*O*-Ara-Ara), suggesting conservation within the green lineage. However, they also found a higher heterogeneity of glycans in *C. reinhardtii*, with the presence of galactofuran residues and methylated residues (absent in plants) [115]. Based on the structures experimentally characterized, they speculated about the *O*-glycosylation pathways in *C. reinhardtii*. They proposed that two arabinosyltransferases add the first two Ara residues to Hyp, and that a galactofuranosyltransferase and two methyltransferases specific to *C. reinhardtii* perform the final modifications [115].

While there is minimal experimental information available on *O*-glycosylation in microalgae, existing data suggests key differences with human glycosylation patterns. *C. reinhardtii* native *O*-glycoproteins present methylated residues on *O*-glycans, a trait also found in microalgal *N*-glycans but completely absent in humans. Moreover, the Hyp-*O*-Ara core has not been found in humans, suggesting the possibility of immunogenic activity for these type of *O*-glycans. However, further investigation of microalgal *O*-glycosylation is unquestionably needed; for example, it is not known to what extent the *C. reinhardtii O*-glycan core structure is immunogenic, if other species present similar *O*-glycan core structures, and whether these glycosylation patterns will also be found on recombinant proteins. In conclusion, more research must be conducted to unravel the *O*-glycosylation patterns in microalgae.

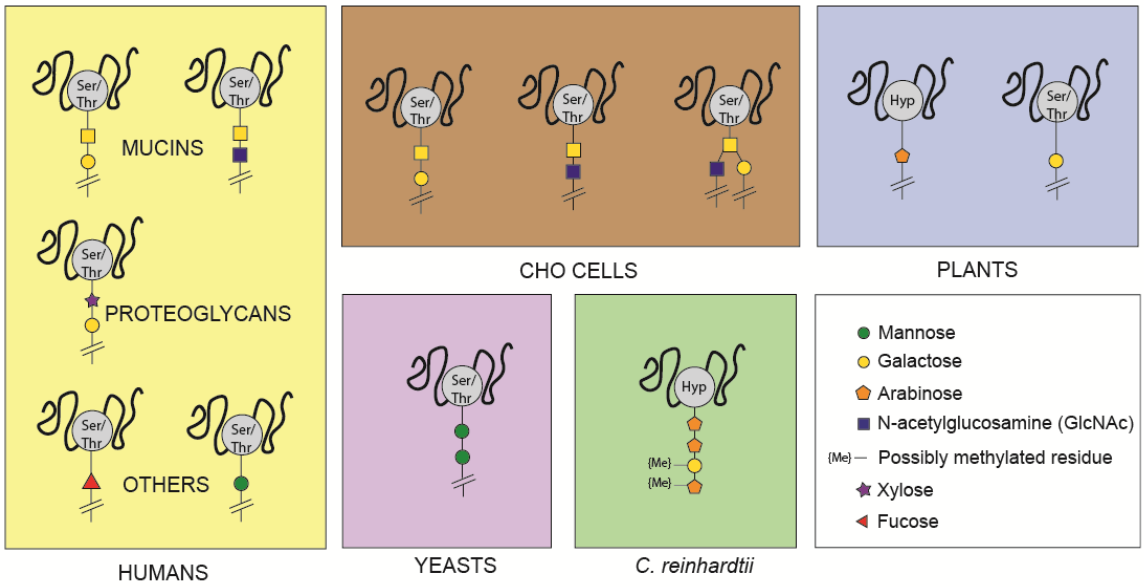
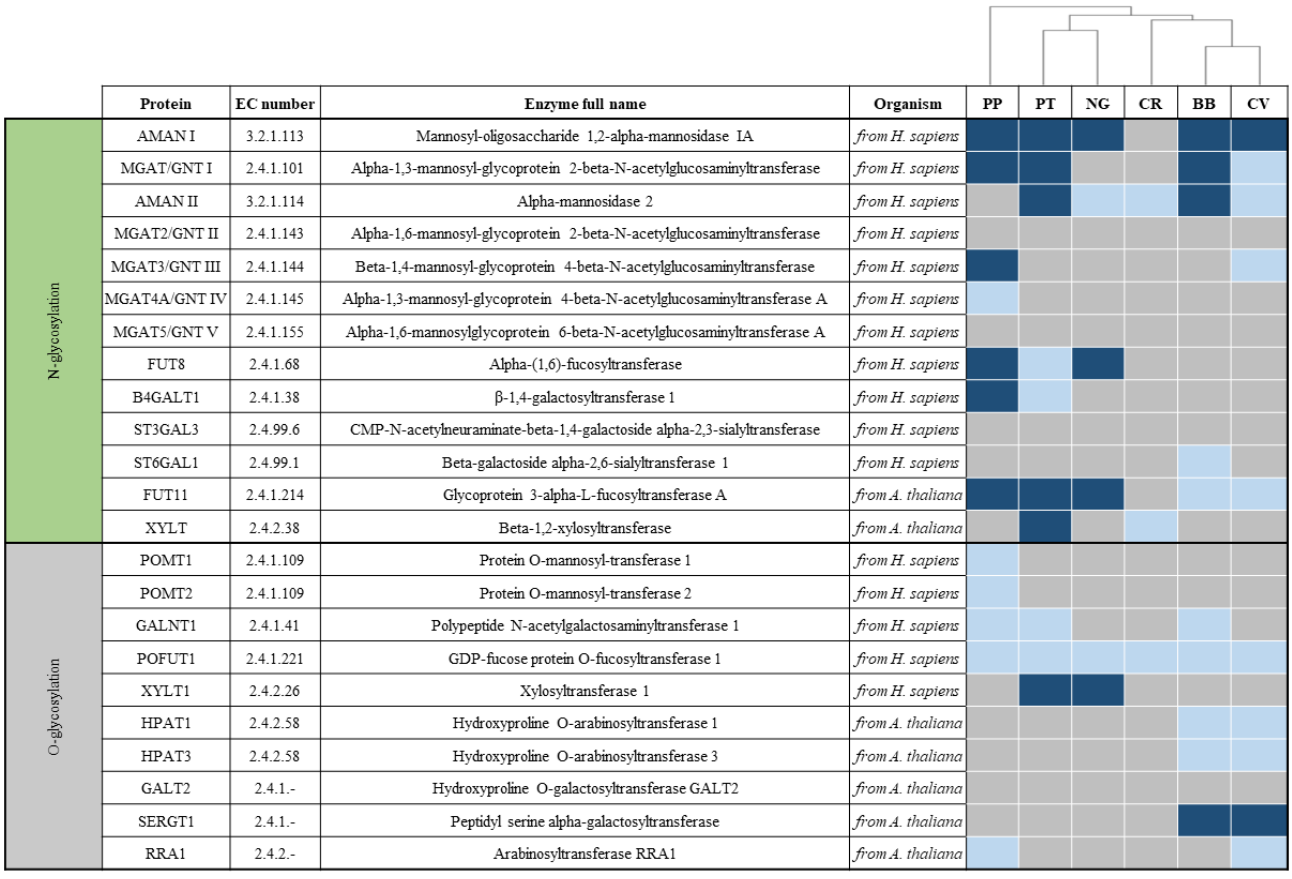


Figure 1.2. A comparison of the different *O*-glycosylation patterns among humans, CHO cells, yeasts, plants, and microalgae. Experimental evidence for *O*-glycosylation in microalgae is limited to *C. reinhardtii* [115]. Native *C. reinhardtii* proteins possess a (Hyp-*O*-Ara-Ara) core and methylated residues, characteristics that differ significantly from human *O*-glycosylation patterns. Ser/Thr = serine or threonine. Hyp = hydroxyproline. Green circle = Mannose. Yellow circle = Galactose. Orange pentagon = Arabinose. Blue square = GlcNAc. Me = Methylated residue. Purple star = Xylose. Red triangle = Fucose.

5.3. Computationally predicted distribution of microalgal *N*- and *O*- glycosylation enzymes

Experimental evidence from a limited number of microalgal species hints at a wide diversity of *N*-glycosylation patterns among different microalgal taxa. For example, the major glycosylation differences in two species (*C. reinhardtii* and *B. braunii*) belonging to the same phylum (Chlorophyta) suggest an extensive level of variation. On the other hand, it is difficult to assess the diversity of *O*-glycoproteins in microalgae, as *O*-glycosylation has not been adequately investigated across multiple groups. Although information from experimental characterization remains limited, computational analysis of available algal genomes permits the hypothetical reconstruction of protein *N*- and *O*-glycosylation pathways (Figure 1.3).



	Protein	EC number	Enzyme full name	Organism	PP	PT	NG	CR	BB	CV
N-glycosylation	AMAN I	3.2.1.113	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	*from H. sapiens*						
	MGAT/GNT I	2.4.1.101	Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	*from H. sapiens*						
	AMAN II	3.2.1.114	Alpha-mannosidase 2	*from H. sapiens*						
	MGAT2/GNT II	2.4.1.143	Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	*from H. sapiens*						
	MGAT3/GNT III	2.4.1.144	Beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase	*from H. sapiens*						
	MGAT4A/GNT IV	2.4.1.145	Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	*from H. sapiens*						
	MGAT5/GNT V	2.4.1.155	Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase A	*from H. sapiens*						
	FUT8	2.4.1.68	Alpha-(1,6)-fucosyltransferase	*from H. sapiens*						
	B4GALT1	2.4.1.38	β-1,4-galactosyltransferase 1	*from H. sapiens*						
	ST3GAL3	2.4.99.6	CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase	*from H. sapiens*						
	ST6GAL1	2.4.99.1	Beta-galactoside alpha-2,6-sialyltransferase 1	*from H. sapiens*						
	FUT11	2.4.1.214	Glycoprotein 3-alpha-L-fucosyltransferase A	*from A. thaliana*						
	XYLT	2.4.2.38	Beta-1,2-xylosyltransferase	*from A. thaliana*						
O-glycosylation	POMT1	2.4.1.109	Protein O-mannosyl-transferase 1	*from H. sapiens*						
	POMT2	2.4.1.109	Protein O-mannosyl-transferase 2	*from H. sapiens*						
	GALNT1	2.4.1.41	Polypeptide N-acetylgalactosaminyltransferase 1	*from H. sapiens*						
	POFUT1	2.4.1.221	GDP-fucose protein O-fucosyltransferase 1	*from H. sapiens*						
	XYLT1	2.4.2.26	Xylosyltransferase 1	*from H. sapiens*						
	HPAT1	2.4.2.58	Hydroxyproline O-arabinosyltransferase 1	*from A. thaliana*						
	HPAT3	2.4.2.58	Hydroxyproline O-arabinosyltransferase 3	*from A. thaliana*						
	GALT2	2.4.1.-	Hydroxyproline O-galactosyltransferase GALT2	*from A. thaliana*						
	SERGT1	2.4.1.-	Peptidyl serine alpha-galactosyltransferase	*from A. thaliana*						
	RRA1	2.4.2.-	Arabinosyltransferase RRA1	*from A. thaliana*						

Figure 1.3. Hypothetical presence of protein *N*-glycosylation and *O*-glycosylation enzymes in the representative microalgal species, *P. purpureum* (PP), *P. tricornutum* (PT), *Nannochloropsis gaditana* (NG), *C. reinhardtii* (CR), *B. braunii* (BB), and *C. vulgaris* (CV), compared with glycosylation enzymes from *H. sapiens* and *A. thaliana*. Enzymes are classified as *present* (dark blue)*, potentially present* (light blue) or *missing* (grey). The genome assemblies of genomes *C. reinhardtii* (GCA_000002595), *P. purpureum* (GCA_000397085), *P. tricornutum* (GCF_000150955)*, B. braunii* (GCA_002005505), *C. vulgaris* (GCA_001021125), and *N. gaditana* (GCA_000569095; GCF_000240725; GCA_001614215), were downloaded from the National Center for Biotechnology Information (NCBI). Potential *Open Reading Frames* (ORF) sequences were predicted on all assemblies using EMBOSS' *getorf* command [116]. Gene candidates were identified using the Basic Local Alignment Tool (BLAST) [117] searching all predicted ORFs for a list of candidate genes. Additionally, template genes were searched for functional domains using PFAM [118] and HMMER v3 [119]. Hit-domain motifs were downloaded and search for in the predicted ORFs. All candidate ORFs determined with BLAST and Pfam domain searches were finally searched and aligned on the NCBI BLASTp webpage. Sequences were classified as present if the candidate ORFs returned good hits (e-value $<1^{-05}$) to proteins of the same or closely related organism annotated with the exact function searched for. Sequences were classified as *potentially present* if: (1) only closely related proteins were identified using BLAST, but specific PFAM domains were present in the genome; (2) BLASTp returned good hits (e-value $<1^{-05}$) to closely related functional proteins in any other organism. All other genes were classified as missing. The glycoprofiles were hierarchically clustered using seaborn cluster map with Euclidean method on Python.

The computational analysis shown in Figure 1.3 supports the observation that glycosylation pathways differ significantly among diverse algal taxa, as all species investigated were predicted to have a unique combination of glycosylation enzymes. However, computational analysis only shows anticipated presence or absence of homologous enzymes and should always be supported by experimental analysis. For example, experimental analysis of glycosylation patterns of *P. purpureum* did not show activity of GnT I enzyme (Figure 1.1). However, GnT I is classified as present in the computational analysis. Similarly, all *O*-arabinosyltransferases from *A. thaliana* (HPAT1 and HPAT3) are classified as absent in *C. reinhardtii* (Figure 1.3), but core arabinose *O*-glycans have been detected in this microalga [115]

(Figure 1.2). Nevertheless, computational analysis can give fundamental insight on microalgal species to be subsequently selected for further experimental analysis.

C. reinhardtii shows a very different enzyme population relative to humans and to *A. thaliana*, except for possible presence of an *O*-fucosyltransferase (POFUT1) that could result in core fucose *O*-glycans for this species. Another interesting result is the possible presence of a plant-like *N*-linked β(1,2)-xylosyltransferase (XYLT). *C. reinhardtii* is known to have xylose residues linked to the core GlcNAc and the antennas, and a previous study has demonstrated the presence of two xylosyltransferases, named XTA and XTB [120]. The absence of HPAT1 and HPAT3 in this microalga was unexpected, given that this microalga was shown to produce *O*-arabinose structures [115]. This suggests the potential presence of alternative enzymes in *C. reinhardtii* capable of linking core *O*-arabinoses.

Unlike *C. reinhardtii,* the green algae *B. braunii* and *C. vulgaris* show the potential presence of GnT I, which is supported by experimental analysis in *B. braunii*. Both species contain candidate genes for core α(1,3)-fucosyltransferase (FUT11), which is immunogenic in humans, and its presence strongly influences the efficacy of biopharmaceuticals [84]. *B. braunii* is the only species to show possible presence of α(2,6)-sialyltransferase (ST6GAL1), which is an important modification found in human proteins. The *O*-glycosylation pathway in *B. braunii* and *C. vulgaris* presents some similarities: the enzymes *O*-fucosyltransferase (POFUT1), and *O*-arabinosyltransferases from *A. thaliana* HPAT1 and HPAT3 are classified as possibly present in both species. However, the two species also show some differences. In fact, only *B. braunii* possibly presents homologues to the *O*-GalNAc transferase (GALNT1), whilst homologous enzymes of plant arabinosyltransferase (RRA1) are possibly present only in *C. vulgaris*.

P. purpureum shows the presence of GnT I, in contradiction with experimental data [98]. It also contains putative homologues to human α(1,6)-fucosyltransferase (FUT8) and FUT11, which can affect biopharmaceutical efficacy [4] and immunogenicity [84], respectively. *P. purpureum* is predicted to have a homologue to human β-1,4-galactosyltransferase 1 (B4GALT1), which is involved in the

attachment of the fundamental human residue sialic acid. However, other enzymes involved in this pathway including α(2,3)-sialyltransferase, α(2,6)-sialyltransferase, GnT I, GnT IV, and GnT V were not detected. For *O*-glycans, *P. purpureum* shows the potential presence of human-like *O*-mannose and *O*-GalNAc cores, and homologous enzymes of plant arabinosyltransferase (RRA1). This enzyme, however, is not involved in core *O*-arabinose linkage; core *O*-arabinose enzymes are reported as missing.

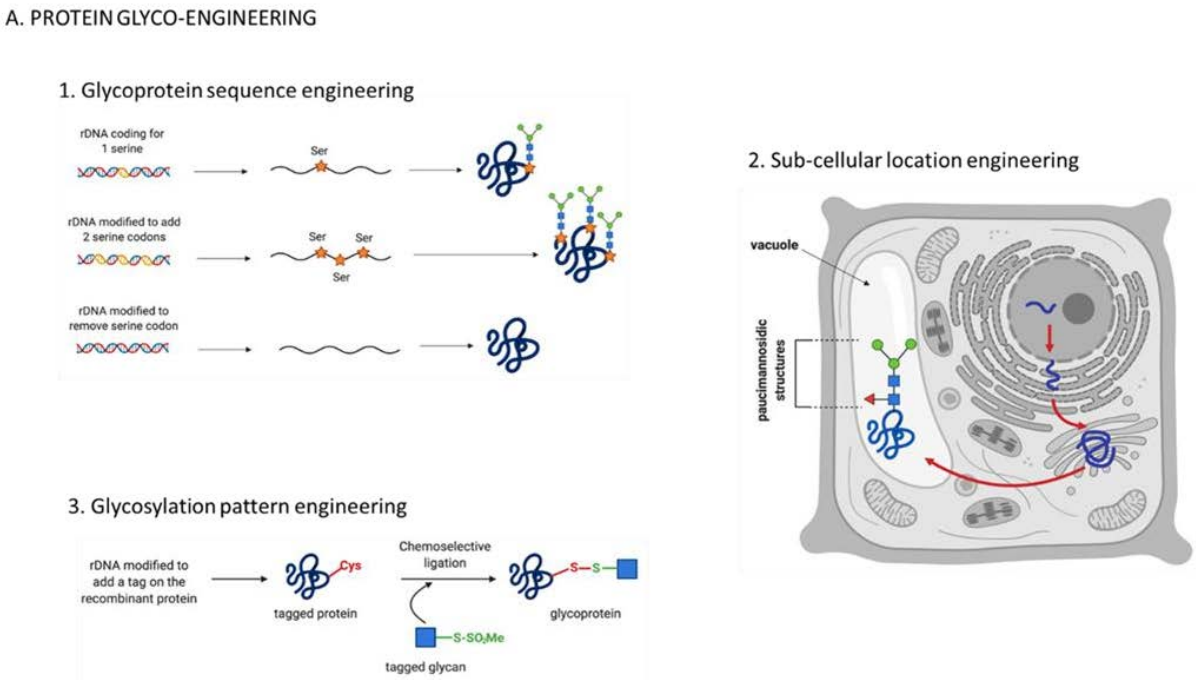
Although both species belong to the stramenopile lineage, *P. tricornutum* and *N. gaditana* vary with respect to the presence of GnT I, which has been predicted in *P. tricornutum* but is not detected in *N. gaditana.* However, both species have putative homologues for human α(1,6)-fucosyltransferase and plant α(1,3)-fucosyltransferase, two enzymes linked to decreased efficacy of biopharmaceuticals. Additionally, *P. tricornutum* is also predicted to harbour a homologue to plant β(1,2)-xylosyltransferase (XYLT) that may add immunogenic residues, although the presence of this enzyme was not detected experimentally [99]. Regarding *O*-glycosylation, both species show possible presence of enzymes involved in the attachment of *O*-xylose and *O*-fucose cores. In addition, *P. tricornutum* may also produce *O*-GalNAc cores; however, experimental analysis of *O*-glycans produced in these species is still needed.

Although the computational analysis suggests that several model algal species possess promising characteristics for biopharmaceutical production, including the presence of GnT I, several potentially problematic enzymes are also present. Given the diversity of microalgal glycans and the predicted differences with human glycosylation profiles, “humanization” of microalgae glycans via glyco-engineering is likely needed to avoid immunogenicity in recombinant biopharmaceuticals produced in microalgae [4].

6. STRATEGIES FOR MANIPULATING PROTEIN GLYCOSYLATION

In biofactories such as *E. coli*, yeasts, CHO cells, and plants, different glyco-engineering techniques have been successfully used to manipulate and “humanize” glycans to produce safe biopharmaceuticals

for patients (*E. coli*: [121,122]. Yeasts: [123,124]. CHO cells: [125,126]. Plants: [103,127]). However, none of these techniques have yet been applied to microalgae. Glyco-engineering strategies can be divided in two main categories: protein engineering and cell engineering [87] (Figure 1.4). Protein glyco-engineering strategies target the recombinant glycoprotein before its translation (by modifying its DNA sequence), during its translation (by modifying its subcellular location), or after its translation (by modifying its glycosylation pattern) [87,128]. Cell glyco-engineering approaches introduce or modify the expression and the activity of target enzymes involved in the glycosylation pathways [87].



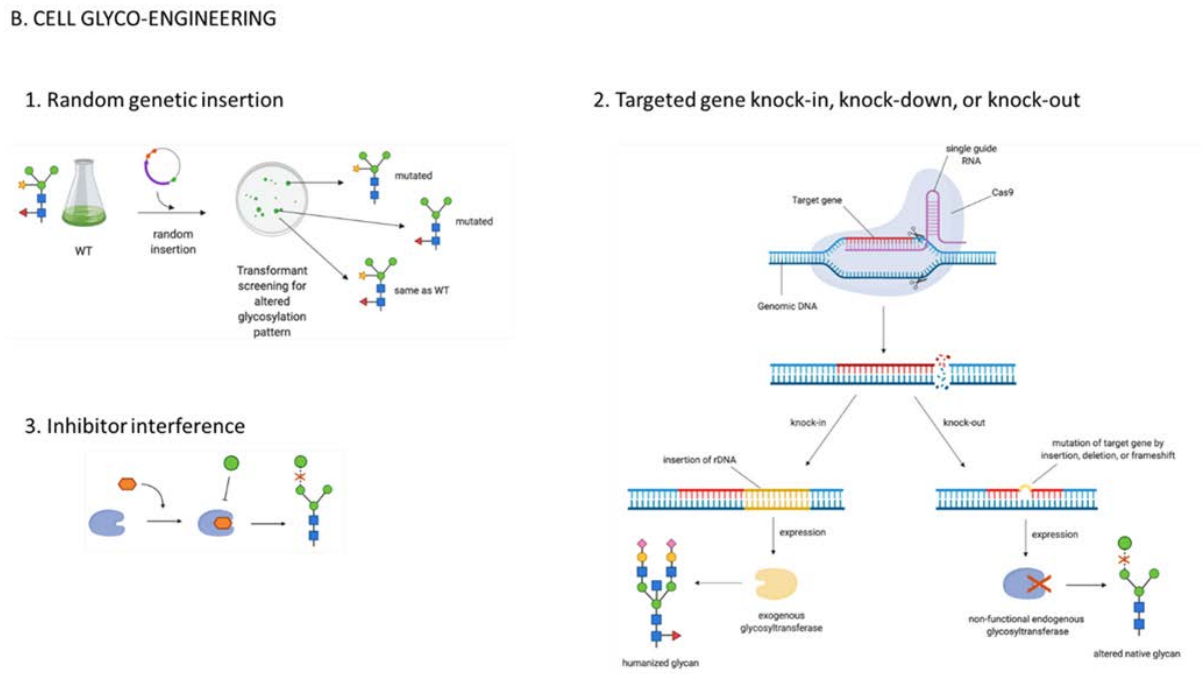


Figure 1.4. A schematic of different glyco-engineering strategies; (A) protein engineering and (B) cell engineering. Protein glyco-engineering approaches (A) can target (1) the recombinant DNA sequence (rDNA), (2) the sub-cellular location of the biopharmaceutical, and (3) the glycosylation pattern of the translated protein. Cell glyco-engineering strategies (B) can modify the activity of glycosylation enzymes by (1) random genetic insertion, (2) targeted gene knock-in or knock-out, and (3) inhibitor interference. Green circle = Mannose. Blue circle = Glucose. Yellow circle = Galactose. Blue square = GlcNAc. Fuchsia diamond = Sialic acid. Yellow star = Xylose. Red triangle = Fucose.

6.1. Protein engineering

6.1.1. Glycoprotein sequence engineering

This strategy is based on changing the amino acid sequence of a recombinant protein without changing its structure or activity, to either i) increase the number of glycans present or ii) remove glycan attachment sites. The first strategy is used in the case of non-immunogenic recombinant glycoproteins in order to enhance their activity, while the second strategy prevents a protein from being glycosylated and therefore reduces its immunogenicity [129]. This approach could be applied to microalgae,

considering the advanced status of microalgal genetic manipulation in many different species [12]. However, these strategies present many limitations, as altering a protein natural glycosylation is likely to cause diminished activity and stability, as demonstrated for IFN-ß [130] and for IgG-like antibody-based therapeutics [131].

6.1.2. Subcellular location engineering

Although major protein modification occurs during ER and Golgi processing, glycans can be additionally modified during transport or within other organelles [102]. Targeting and transporting a recombinant protein to a specific organelle can result in a specific glycan pattern. For example, in plants, glycoproteins targeted and transported to vacuoles present characteristic paucimannosidic *N*-glycans [102]. Recombinant glucocerebrosidase (GCD), a biopharmaceutical used for Gaucher's disease treatment, was found not to be effective unless it presented paucimannosidic *N*-glycans [103]. *In vivo* protein glyco-engineering has been achieved by expressing a recombinant GCD with a C-terminal vacuole-targeting signal in transgenic carrot cells, leading to the successful production of a paucimannosidic biopharmaceutical (ELELYSO®) [103], approved for Gaucher's disease treatment by FDA in 2012 [2]. Microalgae, like plants, possess a vacuole, therefore it might be possible to target proteins to this organelle and obtain paucimannosidic *N*-glycans. Moreover, the microalga *P. tricornutum* naturally expresses paucimannosidic fucosylated *N*-glycans [99]. Although the function of microalgal vacuole is still under-characterized compared to higher plants [132], subcellular localization engineering might result in successful production of paucimannosidic biopharmaceuticals in microalgae. Interestingly, microalgal vacuole signal peptides for *P. tricornutum* have been identified [133]. However, subcellular location engineering presents a major downside: it is not possible to add/remove a targeted residue from the glycan. Therefore, this approach is strictly related to the organelle targeted for protein transportation and the resulting specific glycosylation pattern. Vacuole-targeting glyco-engineering, for example, is specific to proteins like GCD which require a paucimannosidic glycosylation pattern to be active. To remove a selected immunogenic residue, or to add a necessary residue, it is necessary to apply a different glyco-engineering approach.

6.1.3. Glycosylation pattern engineering

Glycosylation pattern engineering targets the glycan after its synthesis by trimming it to its first residue and then reconstructing a new pattern in vitro. This post-translational remodelling of the glycan can be achieved either enzymatically or chemically. These two strategies are briefly described below, but for detailed insight I refer the reader to reviews by Wang and Lomino [134] and Chalker and colleagues [135].

Chemoenzymatic glycan remodelling uses enzymes to both trim the pre-existing glycan and to construct the *in vitro* desired pattern [87]. Studies focusing on the enzymatic single step transfer of a pre-assembled glycan have shown successful results in *E. coli* [121,122], *P. pastoris* [136] and CHO cells [137].

During chemoselective and site-specific glycosylation, a tag is inserted on glycosylation sites on the amino acid backbone by site-directed mutagenesis. A glycan is then added to these tags by bio-orthogonal chemoselective ligation with a modified glycan carrying a compatible functional group. The tag on the protein and the functional group on the glycan are complementary and will selectively recognize each other, enabling a targeted insertion of glycans onto the amino acid backbone [136,137] (Figure 1.4A). Glycosylation pattern engineering is independent of the expression host and therefore could be used in microalgae.

6.2. Cell glyco-engineering

To produce desired human-like glycans or add crucial human residues to existing glycans, it is possible to target the species-specific glycosylation enzymes responsible for the attachment of immunogenic glycans [19,20,21,22,23] using cell glyco-engineering approaches. Cell glyco-engineering can be achieved by using inhibitors or genetic engineering to introduce or modify the expression of enzymes involved in glycosylation pathways [87,136] (Figure 1.4B).

6.2.1. Glyco-engineering by inhibitor interference

An inhibitor interference approach uses small molecules that specifically inhibit or interfere with the activity of specific enzymes in the glycosylation pathway [134] (Figure 1.4B). Small molecules such as *N*-butyl deoxynojirimycin, kifunensine, and swainsonine inhibit the activity of glycosylation enzymes such as the ER α-glucosidases I and II, the ER α-mannosidase-I and the Golgi α-mannosidase II. These molecules are conventionally added to the cell culture before harvesting. Inhibitor interference mostly targets enzymes involved with the formation of the glycan precursor in the ER to block their activity and produce simplified glycans [134]. This technique could be applied to microalgae assuming that non-toxic inhibitors specific to microalgal glycosylation enzymes (such as fucosyltransferase and xylosyltransferase) can be identified. However, successful inhibitors might be difficult to find for these organisms, considering the evolutionary and metabolic biodiversity of microalgae. One example is the insensitivity of the diatom *P. tricornutum* to the activity of terbinafine (well-known inhibitor of the ubiquitous enzyme squalene epoxidase) [138,139]. Moreover, inhibitor interference strategies can only prevent the attachment of specific residues, they cannot enhance or introduce expression of human glycosylation enzymes absent in microalgae (such as the enzyme GnT I in *C. reinhardtii* and *P. purpureum*). To obtain "hybrid" or "complex" glycans it is necessary to genetically engineer the glycosylation machinery of the host organism.

6.2.2. Genetic glyco-engineering

Genetic glyco-engineering can be achieved by either introduction of heterologous glycosylation machinery or inactivation of endogenous enzymes. The integration of one or more genes coding glyco-enzymes can be achieved by random insertion or targeted knock-in (KI) of recombinant DNA into the host genome. The inserted DNA encodes for specific enzymes absent in the wild-type organism that will add the desired residues to the recombinant glycoprotein. One successful example of this strategy is the overexpression of the enzyme GnT III in CHO cells, which is responsible for bisecting GlcNAc containing glycoforms [140]. These bisected glycans are less prone to attach core-linked fucose, resulting in increased efficacy of the recombinant therapeutic glycoprotein [141,142]. Obinutuzumab (GA101 or GAZYVA®) is the example of an FDA approved non-fucosylated drug obtained through

enzyme overexpression in CHO cells [143,144]. Although expression of exogenous enzymes has been proven as an effective strategy, expression must be carefully regulated. Overexpression of glycosylation enzymes can lead to the attachment of extra residues on the glycan, possibly affecting the stability and activity of the recombinant protein.

A major technology advancement across all host systems including microalgae has been the development of targeted genome engineering aided by DNA nucleases, including zinc finger nucleases (ZFNs) [145], transcription activator–like effector nucleases (TALENs) [146], and clustered regularly interspaced short palindromic repeat/targeted Cas9 endonuclease (CRISPR-Cas9) [147,148,149] (Figure 1.4B). DNA nucleases are able to generate double stranded DNA breaks at precise genomic locations of interest, increasing the likelihood of exogenous DNA integration at a desired location, instead of at a random location. Random integration, while much easier to achieve following conventional DNA transformation strategies, is prone to position effect regarding transgene expression and uncharacterised genomic disruptions. Targeted gene integration can improve transgene expression levels when appropriate integration locations are selected; namely, regions that permit insertion of exogenous genes without disrupting the host natural gene expression whilst simultaneously lowering the risk of silencing of the exogenous DNA. Such regions, known as "safe harbours" have been frequently used in human and mouse cell lines [150,151,152]. Identified safe harbour loci knowledge, coupled with efficient endonuclease-mediated targeted integration transformation protocols, can revolutionise genetic engineering strategies to produce recombinant biopharmaceuticals by circumventing reproducibility and stability issues associated with random chromosomal integration [153].

Similar to exogenous enzyme insertion, endogenous enzyme removal strategies can be random or targeted. Gene(s) knock-down (KD) strategies result in reduced activity of the glycosylation enzymes and can be achieved by gene(s) silencing techniques, such as RNA interference [154] or CRISPR interference (CRISPRi) [155]. However, gene(s) KD does not completely inactivate enzyme expression, often resulting in still high levels of the targeted undesired enzyme activity and consequently in the attachment of the unwanted immunogenic residues. Gene(s) knock-out (KO), on the other hand, disrupts

a gene that encodes for a specific enzyme in the glycosylation pathway in order to permanently suppress its function [134]. Gene KO can be obtained by targeted genome editing strategies and/or random insertional mutagenesis [156,157]. This strategy results in loss-of-function mutant cell lines that do not produce the target enzyme, and consequently, lack the undesired residues [87]. Targeted gene(s) KO strategies (ZFNs and CRISPR-Cas9) were successfully used in CHO cells to inactivate the gene responsible for core-linked fucose residues (*FUT8*) and generate non-fucosylated glycans [125,126]. Non-fucosylated biopharmaceuticals expressed in *FUT8*-KO cell lines showed increased efficacy [4]. Mogamulizumab (POTELIGEO®) and Benralizumab (MEDI-563, Fasenra™) are two examples of FDA approved non-fucosylated biopharmaceuticals expressed in *FUT8*-KO CHO cell lines [158]. However, inactivating gene(s) might have secondary detrimental effects on native proteins and consequently on the host organism.

In non-mammalian hosts, integrating human glycosylation enzymes is not sufficient to avoid the presence of immunogenic residues. Exogenous gene integration needs to be coupled with KD or KO of endogenous gene(s) coding for species-specific enzymes [87,134]. A combined approach of exogenous gene insertion and inactivation of endogenous genes was successfully used in yeasts [123,124,159] and plants species such as *Lemna minor* [160], *A. thaliana* [127,161] and *N. benthamiana* [162,163,164].

7. FUTURE PERSPECTIVE FOR GLYCO-ENGINEERING IN MICROALGAE

The effectiveness of glyco-engineering approaches has been demonstrated in many different organisms, both mammalian and non-mammalian, setting the stage for glyco-engineering in microalgae. However, to progress glyco-engineering strategies in microalgae, it is still necessary to improve our knowledge of microalgal glycosylation status (by genomic and experimental evidence) and thus generate a comprehensive and detailed overview of the glycosylation pathways. Unlike in common expression systems, the *N*- and *O*-glycosylation pathways in microalgae have yet to be fully characterized. In this introduction, I provided an *in silico* analysis of glycosylation enzymes in five microalgal species, based on homology with higher organisms (Figure 1.3). These reconstructed pathways, combined with previous glycosylation analysis in other microalgal species (both empirical and putative), support the

presence of a vast diversity of glycosylation patterns and can be used to identify future targets for glyco-engineering in microalgae.

An important consideration for recombinant biopharmaceutical production in microalgae is the choice of host species, as glycosylation status among microalgae has been shown to be very diverse (Chapter 1, section *5.1* and *5.2*). For *N*-glycans, a first selection criteria might be the presence or absence of the GnT I enzyme. GnT I-independent species show complete absence of "hybrid" and "complex" human-like glycans, whereas they still present undesired enzymes and their products (namely Xyl, Fuc, and methylation of residues). However, performing a KI of the GnT I enzyme in a GnT I-independent species lacking FuT, XyT, and MeT (based on computational analysis) might result in non-immunogenic "hybrid" and "complex" glycans. A recent study pursued this strategy in *C. vulgaris* and characterised "hybrid" *N*-glycans lacking fucose or xylose residues after overexpression of recombinant GnT I [101]. Alternatively, GnT I-dependent species already present "hybrid" and "complex" glycans. However, FuT, XyT, and MeT enzymes may still be present and active. Selection of a species that does not possess FuT, XyT, and MeT, or targeted KD or KO of these enzymes would prevent the attachment of the immunogenic residues. Species such as *B. braunii* and *P. tricornutum* are GnT I-dependent species lacking FuT, XyT, and MeT that also have a putative α-mannosidase II (computational analysis). These are particularly interesting candidates to produce paucimannosidic non-immunogenic glycans. However, this strategy is limited by native methylation of residues in *B. braunii*, a possibly immunogenic trait [100]. *P. tricornutum*, on the other hand, does not present methylated residues, but has a core α(1,3)-fucosyltransferase. It might be effective to KD or KO the fucosyltransferase in *P. tricornutum* and obtain paucimannosidic non-immunogenic glycans in an already well-known and well-characterised microalgal species.

Both GnT I-independent and GnT I-dependent species are missing a very important residue present in humans: sialic acid. Enzymes such as α(2,3)-sialyltransferase (ST3GAL3), α(2,6)-sialyltransferase (ST6GAL1), β(1,4)-galactosyltransferase (B4GALT1), GnT I, GnT IV, and GnT V are all involved in the sialic acid pathway, and completely (or almost completely) absent in microalgae. It has been shown

how overexpression of this pathway in CHO cells resulted in improved recombinant biopharmaceuticals production [165,165,167]. Figure 1.3 shows putative presence of some of these enzymes in some microalgal species. However, the complete pathway is absent in all the species that were computationally analysed. Furthermore, sialic acid has never been reported in any of the species experimentally analysed. Therefore, an extensive amount of glyco-engineering might be required to obtain this residue in any of these species.

The absence of a consensus sequence, a common core, and recurring patterns has made a comprehensive *O*-glycosylation immunogenicity analysis more challenging. The only glycosylation analysis available in microalgae (for *C. reinhardtii*) showed similarity with higher plants, with the addition of methylated residues. Immunogenicity of recombinant *O*-glycoproteins from plants has not been tested, however the absence of a similar *O*-glycan core in humans inspires little confidence. The computational analysis shown in Figure 1.3 offers interesting insight. Human *O*-xylose and *O*-fucose cores might be present in more than one microalgal species. Interestingly, microalgae and plants show limited homology (Figure 1.3). However, this should be considered only as a starting point for further experimental analysis of microalgal *O*-glycosylation.

In conclusion, there are still major hurdles limiting glyco-engineering in microalgae. Obtaining experimental information for a single species is challenging and may very well expose immunogenic traits in the selected microalgal glycans. Some immunogenic traits may be compatible with protein or cell engineering, while others may be too complex or widespread to warrant further investigation. There is also no clear evidence to guide which microalgal species might be more appropriate in this regard over others. Moreover, considering the diversity of glycosylation among microalgae, the information obtained cannot be extended to other species, and each microalga must be experimentally investigated. However, glycosylation status of microalgae is still closer to human patterns when compared to glycosylation in bacteria and yeasts. Given the recent advancements in various model microalgal genetic toolkits, glyco-engineering approaches hold great potential to produce non-immunogenic biopharmaceuticals in microalgae.

8. REFERENCES

1. Ghaderi, D.; Zhang, M.; Hurtado-Ziola, N.; Varki, A. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnology and Genetic Engineering Reviews* **2012**, *28*, 147-176, doi:10.5661/bger-28-147.

2. Walsh, G. Biopharmaceutical benchmarks 2014. *Nature Biotechnology* **2014**, *32*, 992, doi:10.1038/nbt.3040.

3. Jayapal, K.P.; Wlaschin, K.F.; Hu, W.; Yap, M.G. Recombinant protein therapeutics from CHO cells-20 years and counting. *Chemical Engineering Progress* **2007**, *103*, 40.

4. Lalonde, M.E.; Durocher, Y. Therapeutic glycoprotein production in mammalian cells. *J Biotechnol* **2017**, *251*, 128-140, doi:10.1016/j.jbiotec.2017.04.028.

5. Hippler, M. *Chlamydomonas: Biotechnology and Biomedicine*; Springer: 2017; Vol. 31.

6. Baeshen, M.N.; Al-Hejin, A.M.; Bora, R.S.; Ahmed, M.M.; Ramadan, H.A.; Saini, K.S.; Baeshen, N.A.; Redwan, E.M. Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives. *J Microbiol Biotechnol* **2015**, *25*, 953-962, doi:10.4014/jmb.1412.12079.

7. Mattanovich, D.; Branduardi, P.; Dato, L.; Gasser, B.; Sauer, M.; Porro, D. Recombinant Protein Production in Yeasts. In *Recombinant Gene Expression*, Lorence, A., Ed. Humana Press: Totowa, NJ, 2012; 10.1007/978-1-61779-433-9_17pp. 329-358.

8. Berlec, A.; Strukelj, B. Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *J Ind Microbiol Biotechnol* **2013**, *40*, 257-274, doi:10.1007/s10295-013-1235-0.

9. Gomord, V.; Faye, L. Posttranslational modification of therapeutic proteins in plants. *Curr Opin Plant Biol* **2004**, *7*, 171-181, doi:10.1016/j.pbi.2004.01.015.

10. Mathieu-Rivet, E.; Kiefer-Meyer, M.C.; Vanier, G.; Ovide, C.; Burel, C.; Lerouge, P.; Bardor, M. Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front Plant Sci* **2014**, *5*, 359, doi:10.3389/fpls.2014.00359.

11. Rasala, B.A.; Mayfield, S.P. Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. *Photosynth Res* **2015**, *123*, 227-239, doi:10.1007/s11120-014-9994-7.

12. Specht, E.; Miyake-Stoner, S.; Mayfield, S. Micro-algae come of age as a platform for recombinant protein production. *Biotechnol Lett* **2010**, *32*, 1373-1383, doi:10.1007/s10529-010-0326-5.

13. Hawkins, R.L.; Nakamura, M. Expression of human growth hormone by the eukaryotic alga, *Chlorella*. *Curr Microbiol* **1999**, *38*, 335-341.

14. Bai, L.L.; Yin, W.B.; Chen, Y.H.; Niu, L.L.; Sun, Y.R.; Zhao, S.M.; Yang, F.Q.; Wang, R.R.; Wu, Q.; Zhang, X.Q., et al. A new strategy to produce a defensin: stable production of mutated NP-1 in nitrate reductase-deficient *Chlorella ellipsoidea*. *PLoS One* **2013**, *8*, e54966, doi:10.1371/journal.pone.0054966.

15. Geng, D.; Wang, Y.; Wang, P.; Li, W.; Sun, Y. Stable expression of hepatitis B surface antigen gene in *Dunaliella salina* (Chlorophyta). *Journal of Applied Phycology* **2003**, *15*, 451-456, doi:10.1023/B:JAPH.0000004298.89183.e5.

16. Chen, H.L.; Li, S.S.; Huang, R.; Tsai, H.-J. Conditional production of a functional fish growth hormone in the transgenic line of *Nannochloropsis oculata* (eustigmatophyceae). *Journal of Phycology* **2008**, *44*, 768-776, doi:10.1111/j.1529-8817.2008.00508.x.

17. Li, S.-S.; Tsai, H.-J. Transgenic microalgae as a non-antibiotic bactericide producer to defend against bacterial pathogen infection in the fish digestive tract. *Fish & Shellfish Immunology* **2009**, *26*, 316-325.

18. Hempel, F.; Maier, U.G. An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microbial Cell Factories* **2012**, *11*, 126, doi:10.1186/1475-2859-11-126.

19. Beltrao, P.; Albanese, V.; Kenner, L.R.; Swaney, D.L.; Burlingame, A.; Villen, J.; Lim, W.A.; Fraser, J.S.; Frydman, J.; Krogan, N.J. Systematic functional prioritization of protein posttranslational modifications. *Cell* **2012**, *150*, 413-425, doi:10.1016/j.cell.2012.05.036.

20. Beltrao, P.; Trinidad, J.C.; Fiedler, D.; Roguev, A.; Lim, W.A.; Shokat, K.M.; Burlingame, A.L.; Krogan, N.J. Evolution of phosphoregulation: comparison of phosphorylation patterns across yeast species. *PLoS Biol* **2009**, *7*, e1000134, doi:10.1371/journal.pbio.1000134.

21. Weinert, B.T.; Wagner, S.A.; Horn, H.; Henriksen, P.; Liu, W.R.; Olsen, J.V.; Jensen, L.J.; Choudhary, C. Proteome-Wide Mapping of the *Drosophila* Acetylome Demonstrates a High Degree of Conservation of Lysine Acetylation. *Science Signaling* **2011**, *4*, ra48-ra48, doi:10.1126/scisignal.2001902.

22. Boekhorst, J.; van Breukelen, B.; Heck, A., Jr.; Snel, B. Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. *Genome Biol* **2008**, *9*, R144, doi:10.1186/gb-2008-9-10-r144.

23. Li, Z.; Wang, Y.; Yao, Q.; Justice, N.B.; Ahn, T.H.; Xu, D.; Hettich, R.L.; Banfield, J.F.; Pan, C. Diverse and divergent protein post-translational modifications in two growth stages of a natural microbial community. *Nat Commun* **2014**, *5*, 4405, doi:10.1038/ncomms5405.

24. Bürkle, A. Posttranslational Modification. In *Encyclopedia of Genetics*, Brenner, S., Miller, J.H., Eds. Academic Press: New York, 2001. 1533.

25. An, H.J.; Froehlich, J.W.; Lebrilla, C.B. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr Opin Chem Biol* **2009**, *13*, 421-426, doi:10.1016/j.cbpa.2009.07.022.

26. Walsh, G.; Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* **2006**, *24*, 1241-1252, doi:10.1038/nbt1252.

27. Higgins, E. Carbohydrate analysis throughout the development of a protein therapeutic. *Glycoconjugate Journal* **2010**, *27*, 211-225, doi:10.1007/s10719-009-9261-x.

28. Butler, M.; Spearman, M. The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr Opin Biotechnol* **2014**, *30*, 107-112, doi:10.1016/j.copbio.2014.06.010.

29. Mizukami, A.; Caron, A.L.; Picanço-Castro, V.; Swiech, K. Platforms for Recombinant Therapeutic Glycoprotein Production. In *Recombinant Glycoprotein Production: Methods and Protocols*, Picanço-Castro, V., Swiech, K., Eds. Springer New York: New York, NY, 2018; 10.1007/978-1-4939-7312-5_1pp. 1-14.

30. Hajba, L.; Szekrenyes, A.; Borza, B.; Guttman, A. On the glycosylation aspects of biosimilarity. *Drug Discov Today* **2018**, *23*, 616-625, doi:10.1016/j.drudis.2018.01.009.

31. Lagasse, H.A.; Alexaki, A.; Simhadri, V.L.; Katagiri, N.H.; Jankowski, W.; Sauna, Z.E.; Kimchi-Sarfaty, C. Recent advances in (therapeutic protein) drug development. *F1000Res* **2017**, *6*, 113, doi:10.12688/f1000research.9970.1.

32. Dumont, J.; Euwart, D.; Mei, B.; Estes, S.; Kshirsagar, R. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit Rev Biotechnol* **2016**, *36*, 1110-1122, doi:10.3109/07388551.2015.1084266.

33. Ahmad, Z.A.; Yeap, S.K.; Ali, A.M.; Ho, W.Y.; Alitheen, N.B.; Hamid, M. scFv antibody: principles and clinical application. *Clin Dev Immunol* **2012**, *2012*, 980250, doi:10.1155/2012/980250.

34. Junutula, J.R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D.D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S.P.; Dennis, M.S., et al. Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat Biotechnol* **2008**, *26*, 925-932, doi:10.1038/nbt.1480.

35. Qiu, X.; Wong, G.; Audet, J.; Bello, A.; Fernando, L.; Alimonti, J.B.; Fausther-Bovendo, H.; Wei, H.; Aviles, J.; Hiatt, E., et al. Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. *Nature* **2014**, *514*, 47-53, doi:10.1038/nature13777.

36. Cadoret, J.P.; Bernard, O. Lipid biofuel production with microalgae: potential and challenges. *J Soc Biol* **2008**, *202*, 201-211, doi:10.1051/jbio:2008022.

37. Sasso, S.; Pohnert, G.; Lohr, M.; Mittag, M.; Hertweck, C. Microalgae in the postgenomic era: a blooming reservoir for new natural products. *FEMS Microbiol Rev* **2012**, *36*, 761-785, doi:10.1111/j.1574-6976.2011.00304.x.

38. Merlin, M.; Gecchele, E.; Capaldi, S.; Pezzotti, M.; Avesani, L. Comparative evaluation of recombinant protein production in different biofactories: the green perspective. *Biomed Res Int* **2014**, *2014*, 136419, doi:10.1155/2014/136419.

39. Eichler-Stahlberg, A.; Weisheit, W.; Ruecker, O.; Heitzer, M. Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. *Planta* **2009**, *229*, 873-883, doi:10.1007/s00425-008-0879-x.

40. Ramos-Martinez, E.M.; Fimognari, L.; Sakuragi, Y. High-yield secretion of recombinant proteins from the microalga *Chlamydomonas reinhardtii*. *Plant Biotechnol J* **2017**, *15*, 1214-1224, doi:10.1111/pbi.12710.

41. Murbach, T.S.; Glavits, R.; Endres, J.R.; Hirka, G.; Vertesi, A.; Beres, E.; Szakonyine, I.P. A Toxicological Evaluation of *Chlamydomonas reinhardtii*, a Green Algae. *Int J Toxicol* **2018**, *37*, 53-62, doi:10.1177/1091581817746109.

42. Fields, F.J.; Ostrand, J.T.; Mayfield, S.P. Fed-batch mixotrophic cultivation of *Chlamydomonas reinhardtii* for high-density cultures. *Algal Research* **2018**, *33*, 109-117, doi:10.1016/j.algal.2018.05.006.

43. Fahad, S.; Khan, F.A.; Pandupuspitasari, N.S.; Ahmed, M.M.; Liao, Y.C.; Waheed, M.T.; Sameeullah, M.; Darkhshan; Hussain, S.; Saud, S., et al. Recent developments in therapeutic protein expression technologies in plants. *Biotechnol Lett* **2015**, *37*, 265-279, doi:10.1007/s10529-014-1699-7.

44. Scaife, M.A.; Nguyen, G.T.; Rico, J.; Lambert, D.; Helliwell, K.E.; Smith, A.G. Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host. *Plant J* **2015**, *82*, 532-546, doi:10.1111/tpj.12781.

45. Demurtas, O.C.; Massa, S.; Ferrante, P.; Venuti, A.; Franconi, R.; Giuliano, G. A *Chlamydomonas*-derived Human Papillomavirus 16 E7 vaccine induces specific tumor protection. *PLoS One* **2013**, *8*, e61473, doi:10.1371/journal.pone.0061473.

46. Dreesen, I.A.; Charpin-El Hamri, G.; Fussenegger, M. Heat-stable oral alga-based vaccine protects mice from *Staphylococcus aureus* infection. *J Biotechnol* **2010**, *145*, 273-280, doi:10.1016/j.jbiotec.2009.12.006.

47. Georgianna, D.R. Production of recombinant enzymes in the marine alga *Dunaliella tertiolecta*. *Algal research* **2013**, *v. 2*, pp. 2-9-2013 v.2012 no.2011, doi:10.1016/j.algal.2012.10.004.

48. Gregory, J.A.; Li, F.; Tomosada, L.M.; Cox, C.J.; Topol, A.B.; Vinetz, J.M.; Mayfield, S. Algae-produced Pfs25 elicits antibodies that inhibit malaria transmission. *PLoS One* **2012**, *7*, e37179, doi:10.1371/journal.pone.0037179.

49. Gregory, J.A.; Topol, A.B.; Doerner, D.Z.; Mayfield, S. Alga-produced cholera toxin-Pfs25 fusion proteins as oral vaccines. *Appl Environ Microbiol* **2013**, *79*, 3917-3925, doi:10.1128/AEM.00714-13.

50. He, D.M.; Qian, K.X.; Shen, G.F.; Zhang, Z.F.; Li, Y.N.; Su, Z.L.; Shao, H.B. Recombination and expression of classical swine fever virus (CSFV) structural protein E2 gene in *Chlamydomonas reinhardtii* chroloplasts. *Colloids Surf B Biointerfaces* **2007**, *55*, 26-30, doi:10.1016/j.colsurfb.2006.10.042.

51. Jones, C.S.; Luong, T.; Hannon, M.; Tran, M.; Gregory, J.A.; Shen, Z.; Briggs, S.P.; Mayfield, S.P. Heterologous expression of the C-terminal antigenic domain of the malaria vaccine candidate Pfs48/45 in the green algae *Chlamydomonas reinhardtii*. *Appl Microbiol Biotechnol* **2013**, *97*, 1987-1995, doi:10.1007/s00253-012-4071-7.

52. Manuell, A.L.; Beligni, M.V.; Elder, J.H.; Siefker, D.T.; Tran, M.; Weber, A.; McDonald, T.L.; Mayfield, S.P. Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast. *Plant Biotechnol J* **2007**, *5*, 402-412, doi:10.1111/j.1467-7652.2007.00249.x.

53. Mayfield, S.P.; Franklin, S.E.; Lerner, R.A. Expression and assembly of a fully active antibody in algae. *Proceedings of the National Academy of Sciences* **2003**, *100*, 438-442, doi:10.1073/pnas.0237108100.

54. Rasala, B.A.; Muto, M.; Lee, P.A.; Jager, M.; Cardoso, R.M.; Behnke, C.A.; Kirk, P.; Hokanson, C.A.; Crea, R.; Mendez, M., et al. Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. *Plant Biotechnol J* **2010**, *8*, 719-733, doi:10.1111/j.1467-7652.2010.00503.x.

55. Su, Z.L.; Qian, K.X.; Tan, C.P.; Meng, C.X.; Qin, S. Recombination and heterologous expression of allophycocyanin gene in the chloroplast of *Chlamydomonas reinhardtii*. *Acta Biochim Biophys Sin (Shanghai)* **2005**, *37*, 709-712.

56. Sun, M.; Qian, K.; Su, N.; Chang, H.; Liu, J.; Shen, G. Foot-and-mouth disease virus VP1 protein fused with cholera toxin B subunit expressed in *Chlamydomonas reinhardtii* chloroplast. *Biotechnol Lett* **2003**, *25*, 1087-1092, doi:10.1023/a:1024140114505.

57. Surzycki, R.; Greenham, K.; Kitayama, K.; Dibal, F.; Wagner, R.; Rochaix, J.D.; Ajam, T.; Surzycki, S. Factors effecting expression of vaccines in microalgae. *Biologicals* **2009**, *37*, 133-138, doi:10.1016/j.biologicals.2009.02.005.

58. Tran, M.; Zhou, B.; Pettersson, P.L.; Gonzalez, M.J.; Mayfield, S.P. Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnology and Bioengineering* **2009**, *104*, 663-673, doi:10.1002/bit.22446.

59. Tran, M.; Henry, R.E.; Siefker, D.; Van, C.; Newkirk, G.; Kim, J.; Bui, J.; Mayfield, S.P. Production of anti-cancer immunotoxins in algae: Ribosome inactivating proteins as fusion partners. *Biotechnology and Bioengineering* **2013**, *110*, 2826-2835, doi:10.1002/bit.24966.

60. Tran, M.; Van, C.; Barrera, D.J.; Pettersson, P.L.; Peinado, C.D.; Bui, J.; Mayfield, S.P. Production of unique immunotoxin cancer therapeutics in algal chloroplasts. *Proc Natl Acad Sci U S A* **2013**, *110*, E15-22, doi:10.1073/pnas.1214638110.

61. Wang, X.; Brandsma, M.; Tremblay, R.; Maxwell, D.; Jevnikar, A.M.; Huner, N.; Ma, S. A novel expression platform for the production of diabetes-associated autoantigen human glutamic acid decarboxylase (hGAD65). *BMC Biotechnol* **2008**, *8*, 87, doi:10.1186/1472-6750-8-87.

62. Yang, Z.; Li, y.; Chen, F.; Li, D.; Zhang, Z.; Liu, Y.; Zheng, D.; Wang, Y.; Shen, G. Expression of human soluble TRAIL in *Chlamydomonas reinhardtii* chloroplast. *Chinese Science Bulletin* **2006**, *51*, 1703-1709, doi:10.1007/s11434-006-2041-0.

63. Yoon, S.M.; Kim, S.Y.; Li, K.F.; Yoon, B.H.; Choe, S.; Kuo, M.M. Transgenic microalgae expressing Escherichia coli AppA phytase as feed additive to reduce phytate excretion in the manure of young broiler chicks. *Appl Microbiol Biotechnol* **2011**, *91*, 553-563, doi:10.1007/s00253-011-3279-2.

64. Zhang, Y.K.; Shen, G.F.; Ru, B.G. Survival of human metallothionein-2 transplastomic *Chlamydomonas reinhardtii* to ultraviolet B exposure. *Acta Biochim Biophys Sin (Shanghai)* **2006**, *38*, 187-193, doi:10.1111/j.1745-7270.2006.00148.x.

65. Baier, T.; Kros, D.; Feiner, R.C.; Lauersen, K.J.; Muller, K.M.; Kruse, O. Engineered Fusion Proteins for Efficient Protein Secretion and Purification of a Human Growth Factor from the Green Microalga *Chlamydomonas reinhardtii*. *ACS Synth Biol* **2018**, 10.1021/acssynbio.8b00226, doi:10.1021/acssynbio.8b00226.

66. Chavez, M.N.; Schenck, T.L.; Hopfner, U.; Centeno-Cerdas, C.; Somlai-Schweiger, I.; Schwarz, C.; Machens, H.G.; Heikenwalder, M.; Bono, M.R.; Allende, M.L., et al. Towards autotrophic tissue engineering: Photosynthetic gene therapy for regeneration. *Biomaterials* **2016**, *75*, 25-36, doi:10.1016/j.biomaterials.2015.10.014.

67. Dauvillee, D.; Delhaye, S.; Gruyer, S.; Slomianny, C.; Moretz, S.E.; d'Hulst, C.; Long, C.A.; Ball, S.G.; Tomavo, S. Engineering the chloroplast targeted malarial vaccine antigens in *Chlamydomonas* starch granules. *PLoS One* **2010**, *5*, e15424, doi:10.1371/journal.pone.0015424.

68. Hou, Q.; Qiu, S.; Liu, Q.; Tian, J.; Hu, Z.; Ni, J. Selenoprotein-transgenic *Chlamydomonas reinhardtii*. *Nutrients* **2013**, *5*, 624-636, doi:10.3390/nu5030624.

69. Lauersen, K.J.; Vanderveer, T.L.; Berger, H.; Kaluza, I.; Mussgnug, J.H.; Walker, V.K.; Kruse, O. Ice recrystallization inhibition mediated by a nuclear-expressed and -secreted recombinant ice-binding protein in the microalga *Chlamydomonas reinhardtii*. *Appl Microbiol Biotechnol* **2013**, *97*, 9763-9772, doi:10.1007/s00253-013-5226-x.

70. Rasala, B.A.; Lee, P.A.; Shen, Z.; Briggs, S.P.; Mendez, M.; Mayfield, S.P. Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS One* **2012**, *7*, e43349, doi:10.1371/journal.pone.0043349.

71. Chen, Y.; Wang, Y.; Sun, Y.; Zhang, L.; Li, W. Highly efficient expression of rabbit neutrophil peptide-1 gene in *Chlorella ellipsoidea* cells. *Curr Genet* **2001**, *39*, 365-370.

72. Kim, D.H.; Kim, Y.T.; Cho, J.J.; Bae, J.H.; Hur, S.B.; Hwang, I.; Choi, T.J. Stable integration and functional expression of flounder growth hormone gene in transformed microalga, *Chlorella ellipsoidea*. *Mar Biotechnol (NY)* **2002**, *4*, 63-73, doi:10.1007/s1012601-0070-x.

73. Feng, S.; Feng, W.; Zhao, L.; Gu, H.; Li, Q.; Shi, K.; Guo, S.; Zhang, N. Preparation of transgenic *Dunaliella salina* for immunization against white spot syndrome virus in crayfish. *Arch Virol* **2014**, *159*, 519-525, doi:10.1007/s00705-013-1856-7.

74. Hempel, F.; Maurer, M.; Brockmann, B.; Mayer, C.; Biedenkopf, N.; Kelterbaum, A.; Becker, S.; Maier, U.G. From hybridomas to a robust microalgal-based production platform: molecular design of a diatom secreting monoclonal antibodies directed against the Marburg virus nucleoprotein. *Microb Cell Fact* **2017**, *16*, 131, doi:10.1186/s12934-017-0745-2.

75. Walsh, C.T.; Garneau-Tsodikova, S.; Gatto, G.J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **2005**, *44*, 7342-7372, doi:10.1002/anie.200501023.

76. Amoresano, A.; Carpentieri, A.; Giangrande, C.; Palmese, A.; Chiappetta, G.; Marino, G.; Pucci, P. Technical advances in proteomics mass spectrometry: identification of post-translational modifications. *Clin Chem Lab Med* **2009**, *47*, 647-665, doi:10.1515/CCLM.2009.154.

77. Deribe, Y.L.; Pawson, T.; Dikic, I. Post-translational modifications in signal integration. *Nat Struct Mol Biol* **2010**, *17*, 666-672, doi:10.1038/nsmb.1842.

78. Nikov, G.; Bhat, V.; Wishnok, J.S.; Tannenbaum, S.R. Analysis of nitrated proteins by nitrotyrosine-specific affinity probes and mass spectrometry. *Analytical Biochemistry* **2003**, *320*, 214-222, doi:10.1016/s0003-2697(03)00359-2.

79. Swaney, D.L.; Beltrao, P.; Starita, L.; Guo, A.; Rush, J.; Fields, S.; Krogan, N.J.; Villen, J. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat Methods* **2013**, *10*, 676-682, doi:10.1038/nmeth.2519.

80. Uversky, V.N. Wrecked regulation of intrinsically disordered proteins in diseases: pathogenicity of deregulated regulators. *Front Mol Biosci* **2014**, *1*, 6, doi:10.3389/fmolb.2014.00006.

81. Ohtsubo, K.; Marth, J.D. Glycosylation in cellular mechanisms of health and disease. *Cell* **2006**, *126*, 855-867, doi:10.1016/j.cell.2006.08.019.

82. Kessler, M.; Goldsmith, D.; Schellekens, H. Immunogenicity of biopharmaceuticals. *Nephrol Dial Transplant* **2006**, *21 Suppl 5*, v9-12, doi:10.1093/ndt/gfl476.

83. Vanier, G.; Lucas, P.L.; Loutelier-Bourhis, C.; Vanier, J.; Plasson, C.; Walet-Balieu, M.L.; Tchi-Song, P.C.; Remy-Jouet, I.; Richard, V.; Bernard, S., et al. Heterologous expression of the N-acetylglucosaminyltransferase I dictates a reinvestigation of the N-glycosylation pathway in Chlamydomonas reinhardtii. *Sci Rep* **2017**, *7*, 10156, doi:10.1038/s41598-017-10698-z.

84. van Beers, M.M.C.; Bardor, M. Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins. *Biotechnology Journal* **2012**, *7*, 1473-1484, doi:10.1002/biot.201200065.

85. Marshall, R.D. The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp* **1974**, 17-26.

86. Zielinska, D.F.; Gnad, F.; Wisniewski, J.R.; Mann, M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **2010**, *141*, 897-907, doi:10.1016/j.cell.2010.04.012.

87. Costa, A.R.; Rodrigues, M.E.; Henriques, M.; Oliveira, R.; Azeredo, J. Glycosylation: impact, control and improvement during therapeutic protein production. *Crit Rev Biotechnol* **2014**, *34*, 281-299, doi:10.3109/07388551.2013.793649.

88. Parodi, A.J. Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochemical Journal* **2000**, *348*, 1.

89. Dell, A.; Morris, H.R. Glycoprotein Structure Determination by Mass Spectrometry. *Science* **2001**, *291*, 2351.

90. Harvey, D.J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2011-2012. *Mass Spectrom Rev* **2015**, 10.1002/mas.21471, doi:10.1002/mas.21471.

91. Varelas, X.; Bouchie, M.P.; Kukuruzinska, M.A. Protein N-glycosylation in oral cancer: dysregulated cellular networks among DPAGT1, E-cadherin adhesion and canonical Wnt signaling. *Glycobiology* **2014**, *24*, 579-591, doi:10.1093/glycob/cwu031.

92. Wacker, M.; Linton, D.; Hitchen, P.G.; Nita-Lazar, M.; Haslam, S.M.; North, S.J.; Panico, M.; Morris, H.R.; Dell, A.; Wren, B.W., et al. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into E. coli. (Reports). *Science* **2002**, *298*, 1790+.

93. Linton, D.; Dorrell, N.; Hitchen, P.G.; Amber, S.; Karlyshev, A.V.; Morris, H.R.; Dell, A.; Valvano, M.A.; Aebi, M.; Wren, B.W. Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol Microbiol* **2005**, *55*, 1695-1703, doi:10.1111/j.1365-2958.2005.04519.x.

94. Tang, H.; Wang, S.; Wang, J.; Song, M.; Xu, M.; Zhang, M.; Shen, Y.; Hou, J.; Bao, X. N-hypermannose glycosylation disruption enhances recombinant protein production by regulating

secretory pathway and cell wall integrity in *Saccharomyces cerevisiae*. *Sci Rep* **2016**, *6*, 25654, doi:10.1038/srep25654.

95. Parsaie Nasab, F.; Aebi, M.; Bernhard, G.; Frey, A.D. A combined system for engineering glycosylation efficiency and glycan structure in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **2013**, *79*, 997-1007, doi:10.1128/AEM.02817-12.

96. Gomord, V.; Fitchette, A.C.; Menu-Bouaouiche, L.; Saint-Jore-Dupas, C.; Plasson, C.; Michaud, D.; Faye, L. Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol J* **2010**, *8*, 564-587, doi:10.1111/j.1467-7652.2009.00497.x.

97. Mathieu-Rivet, E.; Scholz, M.; Arias, C.; Dardelle, F.; Schulze, S.; Le Mauff, F.; Teo, G.; Hochmal, A.K.; Blanco-Rivero, A.; Loutelier-Bourhis, C., et al. Exploring the N-glycosylation pathway in *Chlamydomonas reinhardtii* unravels novel complex structures. *Mol Cell Proteomics* **2013**, *12*, 3160-3183, doi:10.1074/mcp.M113.028191.

98. Levy-Ontman, O.; Arad, S.M.; Harvey, D.J.; Parsons, T.B.; Fairbanks, A.; Tekoah, Y. Unique N-glycan moieties of the 66-kDa cell wall glycoprotein from the red microalga *Porphyridium* sp. *J Biol Chem* **2011**, *286*, 21340-21352, doi:10.1074/jbc.M110.175042.

99. Baiet, B.; Burel, C.; Saint-Jean, B.; Louvet, R.; Menu-Bouaouiche, L.; Kiefer-Meyer, M.C.; Mathieu-Rivet, E.; Lefebvre, T.; Castel, H.; Carlier, A., et al. N-glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. *J Biol Chem* **2011**, *286*, 6152-6164, doi:10.1074/jbc.M110.175711.

100. Schulze, S.; Urzica, E.; Reijnders, M.; van de Geest, H.; Warris, S.; Bakker, L.V.; Fufezan, C.; Martins Dos Santos, V.A.P.; Schaap, P.J.; Peters, S.A., et al. Identification of methylated GnTI-dependent N-glycans in *Botryococcus brauni*. *New Phytol* **2017**, *215*, 1361-1369, doi:10.1111/nph.14713.

101. Mocsai, R.; Figl, R.; Troschl, C.; Strasser, R.; Svehla, E.; Windwarder, M.; Thader, A.; Altmann, F. N-glycans of the microalga *Chlorella vulgaris* are of the oligomannosidic type but highly methylated. *Sci Rep* **2019**, *9*, 331, doi:10.1038/s41598-018-36884-1.

102. Lerouge, P.; Cabanes-Macheteau, M.; Rayon, C.; Fischette-Lainé, A.-C.; Gomord, V.; Faye, L. N-Glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Molecular Biology* **1998**, *38*, 31-48, doi:10.1023/a:1006012005654.

103. Shaaltiel, Y.; Bartfeld, D.; Hashmueli, S.; Baum, G.; Brill-Almon, E.; Galili, G.; Dym, O.; Boldin-Adamsky, S.A.; Silman, I.; Sussman, J.L., et al. Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. *Plant Biotechnol J* **2007**, *5*, 579-590, doi:10.1111/j.1467-7652.2007.00263.x.

104. Zhang, P.; Burel, C.; Plasson, C.; Kiefer-Meyer, M.C.; Ovide, C.; Gugi, B.; Wan, C.; Teo, G.; Mak, A.; Song, Z., et al. Characterization of a GDP-Fucose Transporter and a Fucosyltransferase Involved in the Fucosylation of Glycoproteins in the Diatom *Phaeodactylum tricornutum*. *Front Plant Sci* **2019**, *10*, 610, doi:10.3389/fpls.2019.00610.

105. Nakayama, Y.; Nakamura, N.; Tsuji, D.; Itoh, K.; Kurosak, A. Genetic Diseases Associated with Protein Glycosylation Disorders in Mammals. **2013**, 10.5772/54097, doi:10.5772/54097.

106. Yusibov, V.; Kushnir, N.; Streatfield, S.J. Antibody Production in Plants and Green Algae. *Annu Rev Plant Biol* **2016**, *67*, 669-701, doi:10.1146/annurev-arplant-043015-111812.

107. Mumm, J.S.; Kopan, R. Notch signaling: from the outside in. *Dev Biol* **2000**, *228*, 151-165, doi:10.1006/dbio.2000.9960.

108. Rana, N.A.; Haltiwanger, R.S. Fringe benefits: functional and structural impacts of O-glycosylation on the extracellular domain of Notch receptors. *Curr Opin Struct Biol* **2011**, *21*, 583-589, doi:10.1016/j.sbi.2011.08.008.

109. Dobson, C.M.; Hempel, S.J.; Stalnaker, S.H.; Stuart, R.; Wells, L. O-Mannosylation and human disease. *Cell Mol Life Sci* **2013**, *70*, 2849-2857, doi:10.1007/s00018-012-1193-0.

110. De Pourcq, K.; De Schutter, K.; Callewaert, N. Engineering of glycosylation in yeast and other fungi: current state and perspectives. *Appl Microbiol Biotechnol* **2010**, *87*, 1617-1631, doi:10.1007/s00253-010-2721-1.

111. Liu, J.; Jin, C.; Cherian, R.M.; Karlsson, N.G.; Holgersson, J. O-glycan repertoires on a mucin-type reporter protein expressed in CHO cell pools transiently transfected with O-glycan core enzyme cDNAs. *J Biotechnol* **2015**, *199*, 77-89, doi:10.1016/j.jbiotec.2015.02.017.

112. Mitoma, J.; Petryniak, B.; Hiraoka, N.; Yeh, J.C.; Lowe, J.B.; Fukuda, M. Extended core 1 and core 2 branched O-glycans differentially modulate sialyl Lewis X-type L-selectin ligand activity. *J Biol Chem* **2003**, *278*, 9953-9961, doi:10.1074/jbc.M212756200.

113. Nguema-Ona, E.; Vicre-Gibouin, M.; Gotte, M.; Plancot, B.; Lerouge, P.; Bardor, M.; Driouich, A. Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis and function. *Front Plant Sci* **2014**, *5*, 499, doi:10.3389/fpls.2014.00499.

114. Xu, J.; Tan, L.; Goodrum, K.J.; Kieliszewski, M.J. High-yields and extended serum half-life of human interferon alpha2b expressed in tobacco cells as arabinogalactan-protein fusions. *Biotechnol Bioeng* **2007**, *97*, 997-1008, doi:10.1002/bit.21407.

115. Bollig, K.; Lamshoft, M.; Schweimer, K.; Marner, F.J.; Budzikiewicz, H.; Waffenschmidt, S. Structural analysis of linear hydroxyproline-bound O-glycans of *Chlamydomonas reinhardtii*--conservation of the inner core in *Chlamydomonas* and land plants. *Carbohydr Res* **2007**, *342*, 2557-2566, doi:10.1016/j.carres.2007.08.008.

116. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **2000**, *16*, 276-277, doi.

117. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *Journal of Molecular Biology* **1990**, *215*, 403-410, doi.

118. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A., et al. The Pfam protein families database in 2019. *Nucleic acids research* **2019**, *47*, D427-D432, doi:10.1093/nar/gky995.

119. Johnson, L.S.; Eddy, S.R.; Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **2010**, *11*, 431-431, doi:10.1186/1471-2105-11-431.

120. Lucas, P.-L.; Mathieu-Rivet, E.; Chan Tchi Song, P.; Oltmanns, A.; Loutelier-Bourhis, C.; Plasson, C.; Afonso, C.; Hippler, M.; Lerouge, P.; Mati-Baouche, N., et al. Multiple xylosyltransferases heterogeneously xylosylate protein N-linked glycans in Chlamydomonas reinhardtii. *The Plant Journal* **2019**, *n/a*, doi:10.1111/tpj.14620.

121. Kowarik, M.; Young, N.M.; Numao, S.; Schulz, B.L.; Hug, I.; Callewaert, N.; Mills, D.C.; Watson, D.C.; Hernandez, M.; Kelly, J.F., et al. Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J* **2006**, *25*, 1957-1966, doi:10.1038/sj.emboj.7601087.

122. Schwarz, D.; Daley, D.; Beckhaus, T.; Dötsch, V.; Bernhard, F. Cell-free expression profiling of *E. coli* inner membrane proteins. *PROTEOMICS* **2010**, *10*, 1762-1779, doi:10.1002/pmic.200900485.

123. Hamilton, S.R.; Davidson, R.C.; Sethuraman, N.; Nett, J.H.; Jiang, Y.; Rios, S.; Bobrowicz, P.; Stadheim, T.A.; Li, H.; Choi, B.K., et al. Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* **2006**, *313*, 1441-1443, doi:10.1126/science.1130256.

124. Li, L.; Huang, Y.; Xia, X.; Sun, Z. Preferential duplication in the sparse part of yeast protein interaction network. *Mol Biol Evol* **2006**, *23*, 2467-2473, doi:10.1093/molbev/msl121.

125. Malphettes, L.; Freyvert, Y.; Chang, J.; Liu, P.Q.; Chan, E.; Miller, J.C.; Zhou, Z.; Nguyen, T.; Tsai, C.; Snowden, A.W., et al. Highly efficient deletion of FUT8 in CHO cell lines using zinc-finger nucleases yields cells that produce completely nonfucosylated antibodies. *Biotechnol Bioeng* **2010**, *106*, 774-783, doi:10.1002/bit.22751.

126. Ronda, C.; Pedersen, L.E.; Hansen, H.G.; Kallehauge, T.B.; Betenbaugh, M.J.; Nielsen, A.T.; Kildegaard, H.F. Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol Bioeng* **2014**, *111*, 1604-1616, doi:10.1002/bit.25233.

127. Piron, R.; Santens, F.; De Paepe, A.; Depicker, A.; Callewaert, N. Using GlycoDelete to produce proteins lacking plant-specific N-glycan modification in seeds. *Nature Biotechnology* **2015**, *33*, 1135, doi:10.1038/nbt.3359.

128. Castilho, A.; Steinkellner, H. Glyco-engineering in plants to produce human-like N-glycan structures. *Biotechnol J* **2012**, *7*, 1088-1098, doi:10.1002/biot.201200032.

129. Zhou, Q.; Shankara, S.; Roy, A.; Qiu, H.; Estes, S.; McVie-Wylie, A.; Culm-Merdek, K.; Park, A.; Pan, C.; Edmunds, T. Development of a simple and rapid method for producing non-fucosylated oligomannose containing antibodies with increased effector function. *Biotechnol Bioeng* **2008**, *99*, 652-665, doi:10.1002/bit.21598.

130. Runkel, L.; Meier, W.; Pepinsky, R.B.; Karpusas, M.; Whitty, A.; Kimball, K.; Brickelmaier, M.; Muldowney, C.; Jones, W.; Goelz, S.E. Structural and Functional Differences Between Glycosylated and Non-glycosylated Forms of Human Interferon-β (IFN-β). *Pharmaceutical Research* **1998**, *15*, 641-649, doi:10.1023/a:1011974512425.

131. Sola, R.J.; Griebenow, K. Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci* **2009**, *98*, 1223-1245, doi:10.1002/jps.21504.

132. Shebanova, A.; Ismagulova, T.; Solovchenko, A.; Baulina, O.; Lobakova, E.; Ivanova, A.; Moiseenko, A.; Shaitan, K.; Polshakov, V.; Nedbal, L., et al. Versatility of the green microalga cell vacuole function as revealed by analytical transmission electron microscopy. *Protoplasma* **2017**, *254*, 1323-1340, doi:10.1007/s00709-016-1024-5.

133. Schreiber, V.; Dersch, J.; Puzik, K.; Backer, O.; Liu, X.; Stork, S.; Schulz, J.; Heimerl, T.; Klingl, A.; Zauner, S., et al. The Central Vacuole of the Diatom *Phaeodactylum tricornutum*: Identification of New Vacuolar Membrane Proteins and of a Functional Di-leucine-based Targeting Motif. *Protist* **2017**, *168*, 271-282, doi:10.1016/j.protis.2017.03.001.

134. Wang, L.X.; Lomino, J.V. Emerging technologies for making glycan-defined glycoproteins. *ACS Chem Biol* **2012**, *7*, 110-122, doi:10.1021/cb200429n.

135. Chalker, J.M.; Bernardes, G.J.L.; Davis, B.G. A "Tag-and-Modify" Approach to Site-Selective Protein Modification. *Accounts of Chemical Research* **2011**, *44*, 730-741, doi:10.1021/ar200056q.

136. Wei, C.; Zhou, X.; Zhang, Y. Improving intracellular production of recombinant protein in *Pichia pastoris* using an optimized preinduction glycerol-feeding scheme. *Appl Microbiol Biotechnol* **2008**, *78*, 257-264, doi:10.1007/s00253-007-1315-z.

137. Zou, G.; Ochiai, H.; Huang, W.; Yang, Q.; Li, C.; Wang, L.-X. Chemoenzymatic Synthesis and Fcγ Receptor Binding of Homogeneous Glycoforms of Antibody Fc Domain. Presence of a Bisecting Sugar Moiety Enhances the Affinity of Fc to FcγIIIa Receptor. *Journal of the American Chemical Society* **2011**, *133*, 18975-18991, doi:10.1021/ja208390n.

138. Fabris, M.; Matthijs, M.; Carbonelle, S.; Moses, T.; Pollier, J.; Dasseville, R.; Baart, G.J.; Vyverman, W.; Goossens, A. Tracking the sterol biosynthesis pathway of the diatom *Phaeodactylum tricornutum*. *New Phytol* **2014**, *204*, 521-535, doi:10.1111/nph.12917.

139. Pollier, J.; Vancaester, E.; Kuzhiumparambil, U.; Vickers, C.E.; Vandepoele, K.; Goossens, A.; Fabris, M. A widespread alternative squalene epoxidase participates in eukaryote steroid biosynthesis. *Nat Microbiol* **2019**, *4*, 226-233, doi:10.1038/s41564-018-0305-5.

140. Narasimhan, S. Control of glycoprotein synthesis. UDP-GlcNAc:glycopeptide beta 4-N-acetylglucosaminyltransferase III, an enzyme in hen oviduct which adds GlcNAc in beta 1-4 linkage to the beta-linked mannose of the trimannosyl core of N-glycosyl oligosaccharides. *Journal of Biological Chemistry* **1982**, *257*, 10235-10242.

141. Schuster, M.; Umana, P.; Ferrara, C.; Brunker, P.; Gerdes, C.; Waxenecker, G.; Wiederkum, S.; Schwager, C.; Loibner, H.; Himmler, G., et al. Improved effector functions of a therapeutic monoclonal Lewis Y-specific antibody by glycoform engineering. *Cancer Res* **2005**, *65*, 7934-7941, doi:10.1158/0008-5472.CAN-04-4212.

142. Ferrara, C.; Brünker, P.; Suter, T.; Moser, S.; Püntener, U.; Umaña, P. Modulation of therapeutic antibody effector functions by glycosylation engineering: Influence of Golgi enzyme localization domain and co-expression of heterologous β1, 4-N-acetylglucosaminyltransferase III and Golgi α-mannosidase II. *Biotechnology and Bioengineering* **2006**, *93*, 851-861, doi:10.1002/bit.20777.

143. Mössner, E.; Brünker, P.; Moser, S.; Püntener, U.; Schmidt, C.; Herter, S.; Grau, R.; Gerdes, C.; Nopora, A.; van Puijenbroek, E., et al. Increasing the efficacy of CD20 antibody therapy through the engineering of a new type II anti-CD20 antibody with enhanced direct and immune effector cell-mediated B-cell cytotoxicity. *Blood* **2010**, *115*, 4393-4402, doi:10.1182/blood-2009-06-225979.

144. Cameron, F.; McCormack, P.L. Obinutuzumab: First Global Approval. *Drugs* **2014**, *74*, 147-154, doi:10.1007/s40265-013-0167-3.

145. Sizova, I.; Greiner, A.; Awasthi, M.; Kateriya, S.; Hegemann, P. Nuclear gene targeting in *Chlamydomonas* using engineered zinc-finger nucleases. *Plant J* **2013**, *73*, 873-882, doi:10.1111/tpj.12066.

146. Daboussi, F.; Leduc, S.; Marechal, A.; Dubois, G.; Guyot, V.; Perez-Michaut, C.; Amato, A.; Falciatore, A.; Juillerat, A.; Beurdeley, M., et al. Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. *Nat Commun* **2014**, *5*, 3831, doi:10.1038/ncomms4831.

147. Jiang, W.; Bikard, D.; Cox, D.; Zhang, F.; Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **2013**, *31*, 233-239, doi:10.1038/nbt.2508.

148. Nymark, M.; Sharma, A.K.; Sparstad, T.; Bones, A.M.; Winge, P. A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci Rep* **2016**, *6*, 24951, doi:10.1038/srep24951.

149. Wang, Q.; Lu, Y.; Xin, Y.; Wei, L.; Huang, S.; Xu, J. Genome editing of model oleaginous microalgae *Nannochloropsis* spp. by CRISPR/Cas9. *Plant J* **2016**, *88*, 1071-1081, doi:10.1111/tpj.13307.

150. Cheng, J.K.; Lewis, A.M.; Kim, D.S.; Dyess, T.; Alper, H.S. Identifying and retargeting transcriptional hot spots in the human genome. *Biotechnology Journal* **2016**, *11*, 1100-1109, doi:10.1002/biot.201600015.

151. Lee, J.S.; Kallehauge, T.B.; Pedersen, L.E.; Kildegaard, H.F. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci Rep* **2015**, *5*, 8572, doi:10.1038/srep08572.

152. Papapetrou, E.P.; Schambach, A. Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Mol Ther* **2016**, *24*, 678-684, doi:10.1038/mt.2016.38.

153. Clausen, H.; Wandall, H.H.; Steentoft, C.; Stanley, P.; Schnaar, R.L. Glycosylation engineering. In *Essentials of Glycobiology [Internet]. 3rd edition*, Cold Spring Harbor Laboratory Press: 2017.

154. Omasa, T.; Tanaka, R.; Doi, T.; Ando, M.; Kitamoto, Y.; Honda, K.; Kishimoto, M.; Ohtake, H. Decrease in antithrombin III fucosylation by expressing GDP-fucose transporter siRNA in Chinese hamster ovary cells. *J Biosci Bioeng* **2008**, *106*, 168-173, doi:10.1263/jbb.106.168.

155. Myrbraten, I.S.; Wiull, K.; Salehian, Z.; Havarstein, L.S.; Straume, D.; Mathiesen, G.; Kjos, M. CRISPR Interference for Rapid Knockdown of Essential Cell Cycle Genes in *Lactobacillus plantarum*. *mSphere* **2019**, *4*, doi:10.1128/mSphere.00007-19.

156. Cheng, X.; Liu, G.; Ke, W.; Zhao, L.; Lv, B.; Ma, X.; Xu, N.; Xia, X.; Deng, X.; Zheng, C., et al. Building a multipurpose insertional mutant library for forward and reverse genetics in *Chlamydomonas*. *Plant Methods* **2017**, *13*, 36, doi:10.1186/s13007-017-0183-5.

157. Schulze, S.; Oltmanns, A.; Machnik, N.; Liu, G.; Xu, N.; Jarmatz, N.; Scholz, M.; Sugimoto, K.; Fufezan, C.; Huang, K., et al. N-Glycoproteomic Characterization of Mannosidase and Xylosyltransferase Mutant Strains of *Chlamydomonas reinhardtii*. *Plant Physiol* **2018**, *176*, 1952-1964, doi:10.1104/pp.17.01450.

158. Pereira, N.A.; Chan, K.F.; Lin, P.C.; Song, Z. The "less-is-more" in therapeutic antibodies: Afucosylated anti-cancer antibodies with enhanced antibody-dependent cellular cytotoxicity. *MAbs* **2018**, *10*, 693-711, doi:10.1080/19420862.2018.1466767.

159. Bobrowicz, P.; Davidson, R.C.; Li, H.; Potgieter, T.I.; Nett, J.H.; Hamilton, S.R.; Stadheim, T.A.; Miele, R.G.; Bobrowicz, B.; Mitchell, T., et al. Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose. *Glycobiology* **2004**, *14*, 757-766, doi:10.1093/glycob/cwh104.

160. Cox, K.M.; Sterling, J.D.; Regan, J.T.; Gasdaska, J.R.; Frantz, K.K.; Peele, C.G.; Black, A.; Passmore, D.; Moldovan-Loomis, C.; Srinivasan, M., et al. Glycan optimization of a human monoclonal antibody in the aquatic plant *Lemna minor*. *Nat Biotechnol* **2006**, *24*, 1591-1597, doi:10.1038/nbt1260.

161. Strasser, R.; Altmann, F.; Mach, L.; Glössl, J.; Steinkellner, H. Generation of *Arabidopsis thaliana* plants with complex N-glycans lacking β1,2-linked xylose and core α1,3-linked fucose. *FEBS Letters* **2004**, *561*, 132-136, doi:10.1016/s0014-5793(04)00150-4.

162. Strasser, R.; Stadlmann, J.; Schahs, M.; Stiegler, G.; Quendler, H.; Mach, L.; Glossl, J.; Weterings, K.; Pabst, M.; Steinkellner, H. Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. *Plant Biotechnol J* **2008**, *6*, 392-402, doi:10.1111/j.1467-7652.2008.00330.x.

163. Castilho, A.; Strasser, R.; Stadlmann, J.; Grass, J.; Jez, J.; Gattinger, P.; Kunert, R.; Quendler, H.; Pabst, M.; Leonard, R., et al. In planta protein sialylation through overexpression of the respective mammalian pathway. *J Biol Chem* **2010**, *285*, 15923-15930, doi:10.1074/jbc.M109.088401.

164. Castilho, A.; Gattinger, P.; Grass, J.; Jez, J.; Pabst, M.; Altmann, F.; Gorfer, M.; Strasser, R.; Steinkellner, H. N-glycosylation engineering of plants for the biosynthesis of glycoproteins with

bisected and branched complex N-glycans. *Glycobiology* **2011**, *21*, 813-823, doi:10.1093/glycob/cwr009.

165. Weikert, S.; Papac, D.; Briggs, J.; Cowfer, D.; Tom, S.; Gawlitzek, M.; Lofgren, J.; Mehta, S.; Chisholm, V.; Modi, N., et al. Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins. *Nature Biotechnology* **1999**, *17*, 1116-1121, doi:10.1038/15104.

166. Bork, K.; Reutter, W.; Weidemann, W.; Horstkorte, R. Enhanced sialylation of EPO by overexpression of UDP-GlcNAc 2-epimerase/ManAc kinase containing a sialuria mutation in CHO cells. *FEBS Lett* **2007**, *581*, 4195-4198, doi:10.1016/j.febslet.2007.07.060.

167. Goh, J.S.Y.; Liu, Y.; Liu, H.; Chan, K.F.; Wan, C.; Teo, G.; Zhou, X.; Xie, F.; Zhang, P.; Zhang, Y., et al. Highly sialylated recombinant human erythropoietin production in large-scale perfusion bioreactor utilizing CHO-gmt4 (JW152) with restored GnT I function. *Biotechnology Journal* **2014**, *9*, 100-109, doi:10.1002/biot.201300301.

CHAPTER 2: Proteomic analysis of *Chlamydomonas reinhardtii* strain "UVM4" reveals molecular reprogramming related to enhanced transgene expression

Lorenzo Barolo[1], Audrey S. Commault[1], Raffaela M. Abbriano[1], Matthew P. Padula[2], Unnikrishnan Kuzhiumparambil[1], Mikael Kim[1], Peter J. Ralph[1] and Mathieu Pernice[1]

AFFILIATIONS

[1] University of Technology Sydney, Climate Change Cluster, Broadway Campus, Ultimo NSW 2007, Sydney, Australia

[2]School of Life Sciences and Proteomics Core Facility, Faculty of Science, University of Technology Sydney, Ultimo NSW 2007, Sydney, Australia

CORRESPONDING AUTHORS

Lorenzo Barolo

University of Technology Sydney, Climate Change Cluster, Broadway Campus, Ultimo NSW 2007, Sydney, Australia

Email: lorenzo.barolo@student.uts.edu.au

KEYWORDS: *C. reinhardtii*, UVM4, proteomics, recombinant protein, histone, lysosome.

CONTRIBUTIONS: LB, ASC, RMA, and MP designed the experiment; LB carried the experiment; LB and MPP performed the mass spectrometry analysis; LB, ASC, RMA, and MP performed the data analysis; UK performed the acetate analysis; MK and MP performed the microscope analysis; LB wrote the manuscript with contributions from all authors.

This article was submitted on the 20th of February 2020 to the journal "Journal of Proteome Research".

1. ABSTRACT

The green microalga *Chlamydomonas reinhardtii* has recently emerged as a promising cell biofactory for recombinant protein production. However, major disadvantages such as low nuclear transgene expression still have to be overcome to establish *C. reinhardtii* as a commercially viable platform. One of the reasons for low expression levels is the defence mechanisms activated by the alga to counteract the genetic modifications. To better understand these mechanisms, I compared the proteomes of the wild type strain 137c and the broadly used cell wall-deficient UV mutated strain UVM4 known for its higher nuclear transgene expression efficiency. When grown under the same conditions, strain UVM4 showed different intracellular and extracellular proteomic profiles than the wild type, with particularly higher abundance of proteins involved in chromatin remodelling and translation pathways, while proteolysis proteins showed lower abundance. These results offer important new insights about the modified pathways in strain UVM4 and possible reasons for its higher secreted recombinant protein production. This study also identifies potential protein targets for future genetic engineering to generate new highly expressing strains or to improve the existing ones, and to consequently increase the yields of recombinant biopharmaceuticals in *Chlamydomonas reinhardtii*.

2. INTRODUCTION

Over the past twenty years, the chlorophyte *Chlamydomonas reinhardtii* has emerged as a promising alternative biofactory for the production of recombinant proteins (RPs) (Table 1.1, Chapter 1) [1,2]. *C. reinhardtii* possesses economically interesting features, such as high growth rates and minimal culture media requirements [3]. Unlike mammalian cells, *C. reinhardtii* cannot be contaminated by human pathogens and is generally recognized as safe (GRAS) [4]. *C. reinhardtii* also shares a similar post-translational machinery with higher eukaryotes, facilitating production of complex RPs that require post-translational processing and modification to enhance their stability and functional activity [3,5,6,7]. These advantages, combined with the availability of its fully sequenced nuclear and organellar genomes as well as an advanced genetic toolkit [4], make *C. reinhardtii* the most comprehensive microalgal platform for expression of RPs.

RP production in *C. reinhardtii* can be achieved in both the chloroplast and the nucleus. The level of chloroplastic transgene expression can reach 20% of total soluble protein [3], however the chloroplast lacks the required machinery to perform a fundamental post-translational modification (PTM) such as glycosylation [4,7]. Nuclear-encoded proteins, on the other hand, can be glycosylated through the secretory pathway. By adding a secretion signal peptide, the nuclear translated protein can be directed to the endoplasmic reticulum (ER) and subsequently to the Golgi apparatus [8]. During the transition through the secretory pathway, the recombinant protein is modified to reach its final stable and active form, although the glycosylation pattern may not resemble the original one from the source organism (Figure 1.1 and 1.2, Chapter 1) [9].

Unfortunately, low nuclear transgene expression in *C. reinhardtii* still represents a major issue for industrial RP production [10]. Many different strategies have been examined over time to achieve efficient nuclear expression in *C. reinhardtii*, from the development of new promoters to the addition of functional peptides to improve expression and allow secretion [8,11,12]. A major advance was the generation by Neupert and colleagues [10] of the UVM4 strain, a cell wall-deficient UV mutated *C.*

reinhardtii cell line with improved RP yields. The work from Neupert and colleagues was based on the hypothesis that low nuclear expression levels in *C. reinhardtii* is caused by epigenetic transgene suppression mechanisms or by unconventionally tight chromatin structure. Therefore, the authors aimed at knocking-out the gene(s) allegedly responsible for low nuclear transgenic expression through UV mutagenesis [10]. To select the higher expressive cell lines, an arginine-auxotrophic cell wall-deficient *C. reinhardtii* strain (cw15-302 cwd *mt*+ arg7) was genetically modified by nuclear co-transformation of the *ARG7* gene (to restore arginine prototrophy) with the *CRY1-1* gene (to allow transgene expression measurement). A positive strain (arginine-prototrophic strain showing overexpression of the *CRY1-1* gene) was subjected to UV mutagenesis. Two of these mutant strains (UMV4 and UVM11) were found to have increased expression of recombinant GFP and YFP [10]. UMV4 has since been widely exploited for RP production, achieving remarkable yields of secreted RP (12 – 15 mg/L) [13,14]. However, despite this significant advance towards more efficient nuclear transgene expression in *C. reinhardtii*, little is known about the mechanism(s) that convey the high RP phenotype in UVM4.

Ten years after the publication by Neupert and colleagues [10], the gene(s) and pathway(s) that might be associated with the enhance RPs expression in UVM4, such as enhanced transgene expression, diminished recombinant protein degradation, or others, remain unexplored [10,15]. Moreover, random UV mutagenesis could have mutated other pathways unrelated to transgene expression but still relevant for altered RP yield. For example, a mutant with faster energy metabolism might result in a higher growth rate and consequently higher biomass, contributing to higher RP yields. Another important consideration is glycosylation of the RP. In fact, glycosylation directly and indirectly influences yield, efficacy, pharmacokinetics, and immunogenicity of RPs (as described in Chapter 1).

Here, I conduct the first proteome analysis and comparison of both intracellular and extracellular fractions of the cell wall-deficient UV mutated *C. reinhardtii* strain UVM4, compared to the basic wild type strain cc-125 *mt*+ (also named strain 137c), to investigate abundance of the products of Open Reading Frames to infer changes in gene expression. This study provides the first proteomic insights into this biotechnologically useful and widespread strain with the ultimate goal to better understand the cellular and molecular mechanisms possibly driving higher RP production in strain UVM4.

3. MATERIALS AND METHODS

C. reinhardtii cultivation conditions and harvesting

C. reinhardtii wild type strain cc-125 *mt+* (also named strain 137c) was purchased from Invitrogen (GeneArt® Chlamydomonas Protein Expression Kit). *C. reinhardtii* strain UVM4 was graciously provided by Prof. Ralph Bock. Strain UVM4 was obtained after genetic engineering and UV mutagenesis of strain CC-4350. Strain CC-4350 was isolated by Matagne et al. from a cross between a cell wall-deficient strain obtained from Davies and Plaskitt [28] (from the cw15 family) and an arg7-8 nit-1 nit-2 strain from the laboratory of Prof. Matagne. Both strain cw15 and arg7-8 nit-1 nit-2 were derived from strain CC-125 *mt+* (also named strain 137c) [28] [Prof. Matagne dir. comm.]. *C. reinhardtii* strains 137c and UVM4 were up-scaled from a single colony and grown under mixotrophic conditions at 25°C and 100 rpm in 4 mL Tris Acetate Phosphate (TAP) medium [17] with ~50 µmol photons m^{-2} s^{-1} of continuous light until mid-exponential phase (~1.5 OD at 750 nm). The cells were then transferred (1:250 inoculum ratio) to 500 mL of TAP medium and grown under the same conditions. Samples (cells and medium) were harvested when the optical density of the cultures reached ~1.5 OD at 750 nm (mid-exponential phase; 72 hours). *C. reinhardtii* cells and medium were separated by tangential flow microfiltration with a 0.2 µm polyethersulfone (PES) membrane (Vivaflow 200, Sartorius). The filtrate (growth medium) was subsequently concentrated 250-fold by crossflow ultrafiltration with a 3 kDa cut-off (3,000 MWCO PES membrane, Vivaflow 200 and Vivaspin Turbo 15, Sartorius). Samples were stored at -80°C degrees. The experiment was performed with three independent biological replicates for each strain. To validate the harvesting method, and to prove that the tangential flow microfiltration was not disrupting living cells (consequently altering the intracellular and extracellular proteomes), dead cells were monitored before and after the filtration, using the LIVE/DEAD® Fixable Violet Dead Cell Stain (ThermoFisher Scientific) and analysed by flow cytometry. The results showed no significant difference in dead cell amount before and after the filtration (data not shown), demonstrating that the ultrafiltration did not damage the cells.

Cell morphology and motility

Cell pellets of strains 137c and UVM4 were obtained by centrifugation at 1000 g for 5 minutes and placed in a chemical fixative solution containing 1% formaldehyde and 2.5% glutaraldehyde in phosphate buffered-saline (PBS: 0.1M phosphate, 0.6M sucrose, pH 7.5) for 24h at 4°C. Samples were then stored in 1x PBS at 4°C before being transported to the Australian Centre for Microscopy and Microanalysis at the University of Sydney for embedding in SPURR resin. Briefly, cell pellets were dehydrated in an increasing gradient of ethanol (50%, 70%, 90%, 100%) followed by 100% acetone. Infiltration with SPURR resin was achieved via an increasing gradient of SPURR resin (25%, 50%, 75%, 90% and 100%) mixed with acetone, including an overnight infiltration session with fresh 100% SPURR resin. The samples were then placed in a 65°C oven for 24h to allow for polymerisation. Semi-thin sections (500nm thick) were obtained using a glass blade on an ultramicrotome (Leica EM UC6, Leica Microsystems, Wetzlar Germany) and stained using a 5% toluidine blue solution (5% toluidine blue, 10% $Na_2B_4O_7$, ultrapure water). Imaging was performed to identify regions of interest. Image analysis was carried out using the Eclipse Ci-L microscope (Nikon) and the images were captured and calibrated of scale bar using the Infinity Analyse software.

Protein extraction and sample preparation

The harvested cells were resuspended (1:5 cell pellet to buffer ratio) in cell lysing buffer (Tris-HCl 50 mM, NaCl 0.4 M, Tween20 0.5%, pH 8.0) and sonicated on ice for 5 minutes (30 seconds on / 30 seconds off) at 30% amplitude (Sonicator Q500, QSonica). The cell debris were separated from the lysate by centrifugation at 10,000 g for 10 minutes at 4°C.

To remove non-protein contaminants, the proteins in both intracellular and extracellular fractions were precipitated using a chloroform/methanol precipitation protocol adapted from Wessel and Flugge [18] and subsequently resuspended in 100 mM of triethylammonium bicarbonate buffer (TEAB) with different concentrations of urea (i.e. 8M for intracellular samples, 1M for extracellular samples). The concentration of proteins in each fraction was quantified with a bicinchoninic (BCA) protein assay (Pierce™ BCA Protein Assay Kit, ThermoFisher Scientific). To remove any possibility of further contamination, 40 µg of total protein from both intracellular and extracellular fractions were subjected

to a single-pot, solid-phase-enhanced sample preparation (SP3) as described in Hughes et al. [19]. The protein preparation step (reduction and alkylation) was performed using 5 mM of tris(2-carboxyethyl)phosphine (TCEP), 20 mM of acrylamide monomers (AM), and 20 mM of dithiothreitol (DTT). After SP3 preparation, samples were resuspended in 100 µL of 200 mM ammonium bicarbonate (AMBIC), obtaining a final concentration of protein of 0.4 µg/µL. Finally, protein digestion was performed overnight with proteomic grade trypsin (Trypsin Gold, Promega) in a 1:50 w/w ratio at 37°C. The peptides obtained from digestion were quantified using Pierce™ Quantitative Colorimetric Peptide Assay (ThermoFisher Scientific). 2 µg of peptides were concentrated and resuspended in 5 µL of mass spectrometry loading buffer (2% acetonitrile (ACN), 0.2% trifluoroacetic acid (TFA) and analysed by mass spectrometry.

Mass spectrometry proteome analysis

Using an Acquity M-class nanoLC system (Waters, USA), 5 µL of the sample was loaded at 15µL/min for 2 minutes onto a nanoEase Symmetry C18 trapping column (180 µm x 20 mm) before being washed onto a PicoFrit column (75 µm ID x 300 mm; New Objective, Woburn, MA) packed with Magic C18AQ resin (3 µm, Michrom Bioresources). Peptides were eluted from the column and into the source of a Q Exactive Plus mass spectrometer (Thermo Scientific) using the following program: 5-30% MS buffer B (98% Acetonitrile + 0.2% Formic Acid) over 90 minutes, 30 – 80% MS buffer B over 3 minutes, 80% MS buffer B for 2 minutes, 80 – 5% for 3 min. The eluting peptides were ionised at 2400V. A Data Dependant MS/MS (dd-MS2) experiment was performed, with a survey scan of 350 – 1,500 Da performed at 70,000 resolution for peptides of charge state 2+ or higher with an AGC target of $3e^6$ and maximum Injection Time of 50 ms. The Top 12 peptides were selected fragmented in the HCD cell using an isolation window of 1.4 m/z, an AGC target of $1e^5$ and maximum injection time of 100ms. Fragments were scanned in the Orbitrap analyser at 17,500 resolution and the product ion fragment masses measured over a mass range of 50 – 2,000 Da. The mass of the precursor peptide was then excluded for 30 seconds.

Data Analysis

The MS/MS data files were searched using Peaks Studio X against the UniProt *Chlamydomonas reinhardtii* proteome (UP000006906, protein count: 18,829) and a database of common contaminants with the following parameter settings: Fixed modifications: none; Variable modifications: oxidised methionine, deamidated asparagine; Enzyme: semi-trypsin; Number of allowed missed cleavages: 3; Peptide mass tolerance: 10 ppm; MS/MS mass tolerance: 0.05 Da. The results of the search were then filtered to include peptides with a $-\log_{10}(P_{value})$ score that was determined by the False Discovery Rate (FDR) of <1%, the score being that where decoy database search matches were <1% of the total matches.

Protein annotation

Intracellular and extracellular proteins were quantified using PEAKS label-free quantification. This quantification method does not use isotope labelling. Instead, label-free quantification is based on the m/z (mass-to-charge ratio), LC retention time, and signal intensities (peak area in MS1 scan). Significant difference was measured by PEAKS Q label-free quantification algorithm [20]. Differentially expressed proteins were selected based on fold change (≥ 2), significance (≥ 20), and unique peptides (≥ 3). The significance parameter is based on the $-\log_{10}(P_{value})$ and it is generated by the PEAKS Q algorithm. All the different datasets were exported as .csv files and uploaded onto the UniProt website to obtain a more complete annotation of the proteins. PEAKS identification includes the protein accession number, which matches the UniProt database number. However, redundancies in the database present a significant issue, as the same amino acidic sequence can often be found under two, three, or even four different accession numbers and/or protein names. These redundant annotations were removed manually to simplify the proteomic analysis; however, we suggest that removing redundancy from the UniProt *C. reinhardtii* database should be done to prevent complications in similar analyses. Annotations from the PEAKS software identification and the UniProt database annotation were combined to manually categorize the significant proteins based on Gene Ontology (GO) biological process, molecular function and cellular component [21].

Statistical analysis

Statistical analysis of the cell density and the acetate uptake data were done in GraphPad Prism 5.0. Kolmogorov-Smirnov and Levene's test were first used to confirm normality and homoscedasticity of the data, respectively. A parametric test (two-way ANOVA) was applied accordingly. Significant effects were then analysed using post-hoc Fisher's LSD test. These analyses tested the null hypothesis that there was no difference in growth and acetate uptake between the 137c and UVM4 strains. The results were considered significant at p-value < 0.05. Throughout the Chapter, values given are mean ± SEM (n=3 biological replicates).

4. RESULTS AND DISCUSSION

4.1. Cell biology

Besides the absence of a cell wall and higher secreted RP yields in strain UVM4, the physiological differences between the UV mutated strain and the wild type strain 137c have not yet been fully investigated. Therefore, I analysed and compared cell morphology, motility, growth rates, and acetate uptake of the two strains.

4.1.1. Cell morphology and motility

Notable morphological differences were detected between strain 137c and UVM4 by optical microscopy after 5% toluidine blue staining. Strain UVM4 cells were much smaller (around 5 μm for UVM4 instead of 10 μm for strain 137c), lacking the usual round shape and cell wall (Figure 2.1).

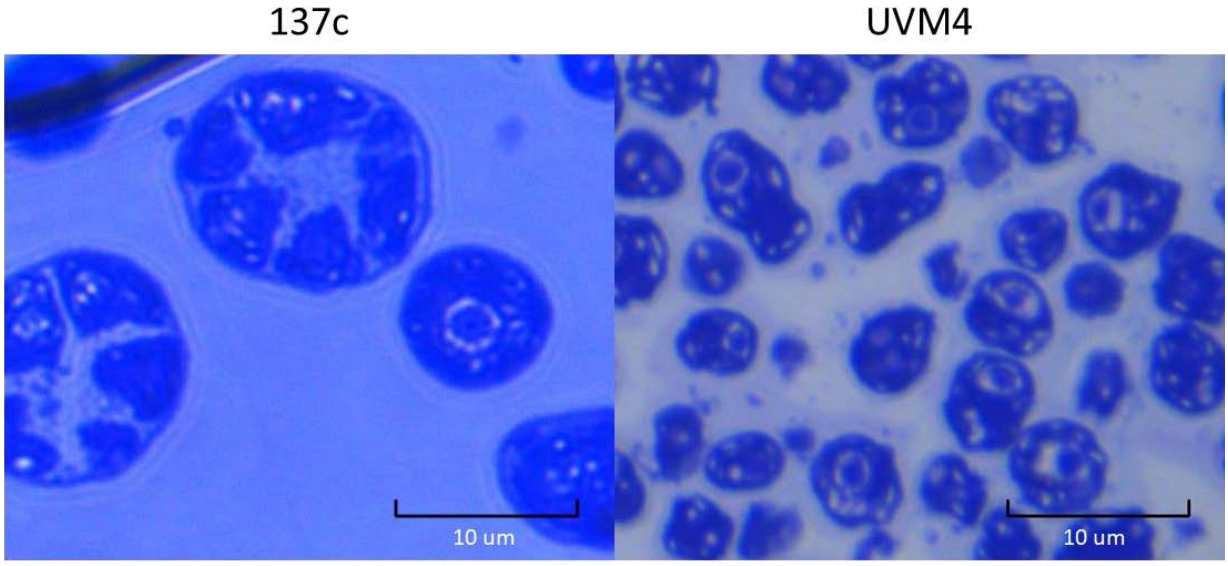


Figure 2.1. Optical microscope observations (magnification 100X) of strain 137c and strain UVM4. Scale bar: 10 μm.

Additionally, unlike strain 137c, strain UVM4 cells were non-motile (data not shown).

4.1.2. Growth and acetate uptake

Considering the differences in size between strains 137c and UVM4, cell density (cells mL^{-1}) and optical density (at 750 nm) were simultaneously measured to assess differences in growth (Figure 2.2A and 2.2B).

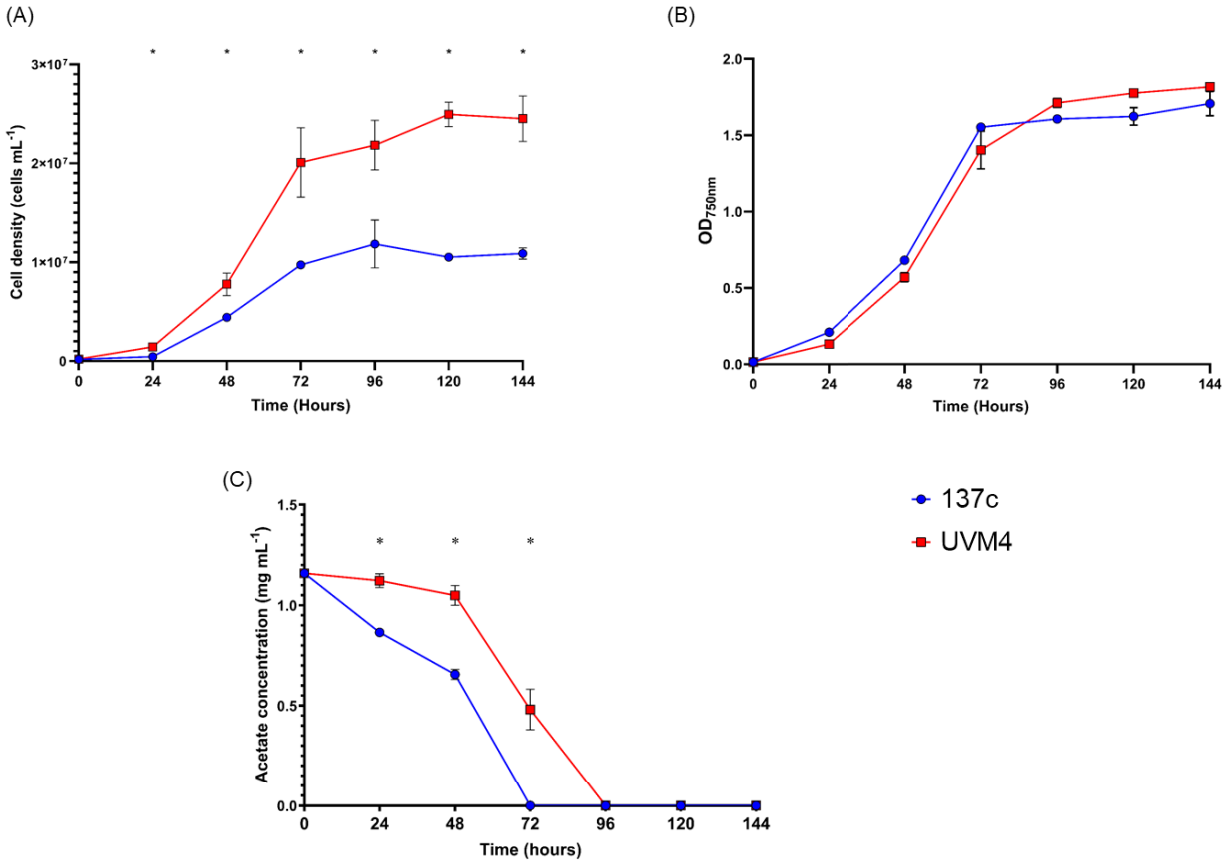


Figure 2.2. Comparison of 137c and UVM4 strains growth (cell density (A) and optical density (B)), and acetate consumption (C). Acetate intake was assessed by measuring the concentration of acetate in spent media over time. Data are mean ± SEM (n=3). *Significant differences between the two strains (Fisher's LSD, $p < 0.05$).

As shown in Figure 2.2A and 2.2B, although the cell density values for strains 137c and UVM4 cells were significantly different, the optical density values were similar. This discrepancy between cell density and optical density measurement is likely to be related to the size difference observed between strain UVM4 and 137c cells. Indeed, the optical density reflects the volume occupied by cells in the culture medium. Solutions of culture medium containing cells of different size will display different

optical densities when grown to the same cell density, and therefore will experience significantly different light and nutrient availability. Mid-exponential phase was reached at ~72 hours (Figure 2.2A and 2.2B) for both strain UVM4 and strain 137c, therefore I selected this time point (72 hours) to harvest both cultures for proteomic analysis. During the following day (96 hours), the two strains showed a slight, although not significant, difference in growth with strain 137c entering stationary phase 24 hours before UVM4 strain (Figure 2.2B).

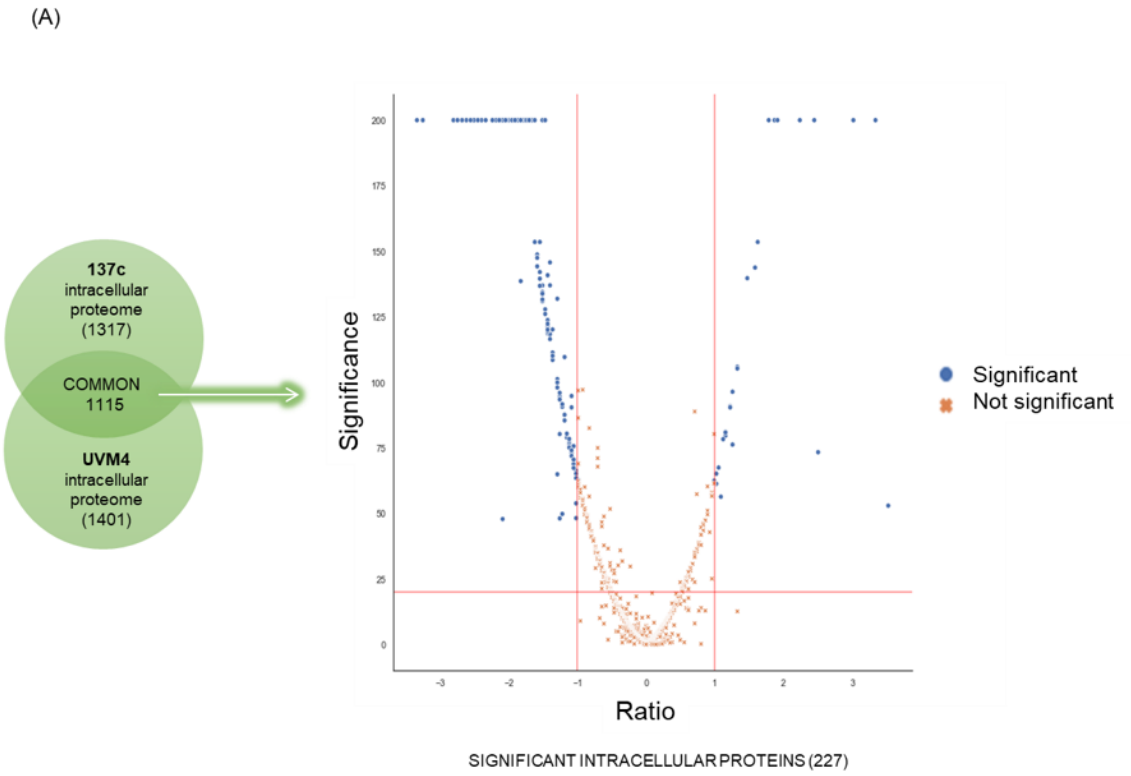
The two strains showed significant differences in acetate consumption as soon as 24 hours after inoculation, with strain 137c consuming acetate more quickly than strain UVM4 when grown under the same conditions. The concentration of acetate in strain 137c growth media started decreasing right after inoculation, to reach a negligible concentration at 72 hours. On the other hand, acetate concentration in strain UVM4 spent medium showed almost no decrease in the first 48 hours of growth, to reach negligible concentration at 96 hours. In each case, the depletion of acetate from the medium was consistent with the growth curve, as both strains entered stationary phase once the acetate was depleted. The reason for delayed acetate consumption in strain UVM4 is unknown, but it could be related to the strain smaller size and/or to the fact that strain UVM4 could rely more on photosynthesis than organic carbon uptake at the beginning of its growth phase.

4.2. Proteomic analysis

4.2.1. Protein selection and annotation

The cells and media were harvested at 72 hours for proteomic analysis. Even so the cell densities showed differences, the extracted total protein was quantified and the same amount of peptides were loaded into the mass spectrometer. The proteins detected in all three biological replicates were identified and the complete proteome profiles of strains 137c and UVM4 were compared, obtaining common proteins (i.e. detected and identified in both strains) and unique proteins (i.e. detected and identified only in one strain). The common intracellular proteins detected were 1115 (Figure 2.3A). These proteins were analysed using label-free quantification to obtain relative abundance of the proteins detected in strain UVM4 (using strain 137c as a control) (Figure 2.3A). Of these, only the significant proteins

(significance value ≥ 20 and -1 ≥ $\log_2$ fold change ≥ 1, parameters generated by the PEAKS Q algorithm [20]) were investigated further. The extracellular proteomes were analysed in the same way (Figure 2.3B). For the unique proteins, it was not possible to obtain relative abundance using label-free quantification; however, these unique proteomic profiles were examined to validate and support the results obtained from the label-free quantification of the common proteins.



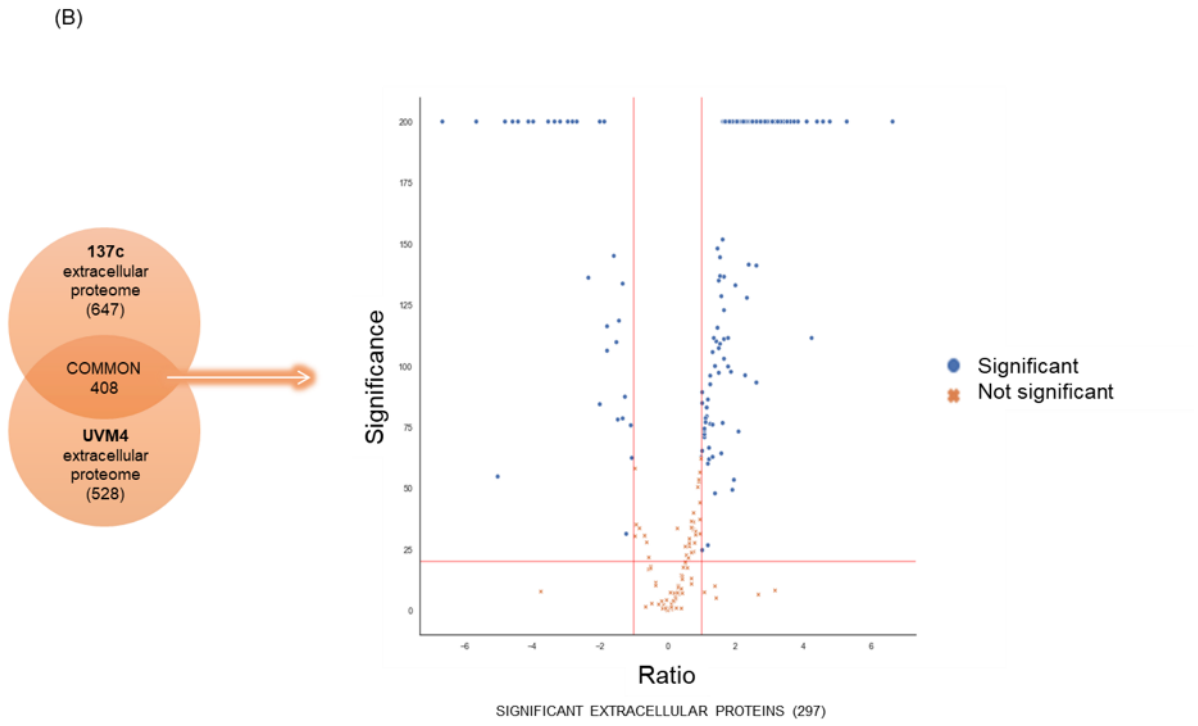


Figure 2.3. Venn diagrams and volcano plots of the common proteins in intracellular (A) and extracellular (B) proteomes comparison. The number of proteins in each category is given in brackets. Out of the common proteins, 227 and 297 proteins were found to be significantly different in the intracellular and extracellular proteomes, respectively. Differentially produced proteins were selected based on significance value ≥ 20 and fold change of $-1 \geq \log_2 \geq 1$.

The proteins were manually annotated based on the Gene Ontology project categories: biological process, molecular function, and cellular component [21], and organised into subgroups based on predicted functional annotation (Supplementary figure 1). Amongst the different groups identified, I chose to focus this study on specific groups directly or indirectly related to higher recombinant protein yields (i.e. growth, cell wall assembly, transcription, translation and proteolysis).

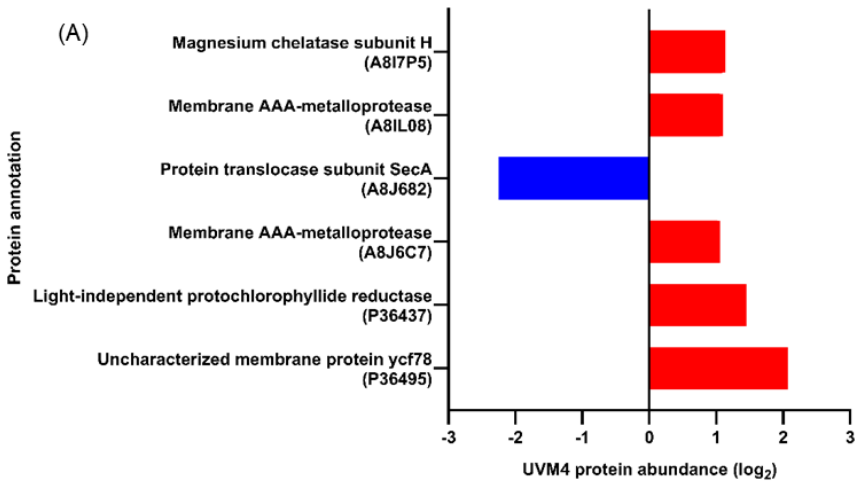
4.2.2. CRY1-1 and ARG7 gene expression

Strain UVM4 was originally obtained by genetic engineering and UV mutation; arginine-auxotrophic cell wall-deficient *C. reinhardtii* cells (from strain cw15-302 cwd *mt*+ arg7) were first subjected to transformation with the *CRY1-1* and the *ARG7* genes, and subsequently subjected to UV mutagenesis

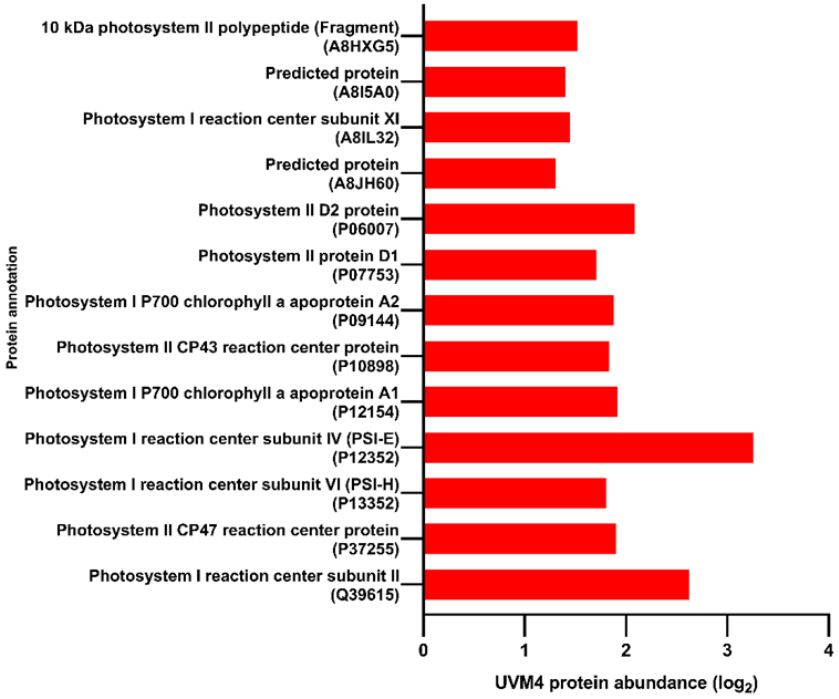
[10]. The *CRY1-1* gene codes for the cytosolic ribosomal protein S14 (UniProt accession number: P46295), which neutralises the action of the translational inhibitor emetine [22]. Therefore, the expression level of *CRY1-1* is directly proportional to the level of emetine resistance [10]. The cells were co-transformed with the *ARG7* gene to restore arginine prototrophy. The *ARG7* codes for the argininosuccinate lyase (UniProt accession number: P22675), which restored the arginine prototrophy of this strain. A positive strain (arginine-prototrophic strain showing resistance to emetine) was subjected to UV mutagenesis and became strain UVM4. Therefore, these two proteins were selected for a preliminary analysis to validate the method used. Unfortunately, the ribosomal protein S14 was below the detection limit, and could not be analysed. On the other hand, the protein argininosuccinate lyase was detected in both strains, showing higher abundance in strain UVM4 (fold change = 4.14 ($2^{2.05}$)) compared to strain 137c (set as control), in congruence with initial genetic transformation by Neupert and colleagues [10].

4.2.3. Growth and photosynthesis

As shown in Figure 2.2, strain UVM4 grows to a higher cell density than strain 137c and reaches stationary phase 24 hours later, when grown in the same conditions and inoculated at non-significantly different cell density. To elucidate the reasons for these differences, the abundance of proteins involved in photosynthesis and growth mechanisms of strain UVM4 and strain 137c (set as control) were compared (Figure 2.4).



(B)


10 kDa photosystem II polypeptide (Fragment) (A8HXG5)
Predicted protein (A8I5A0)
Photosystem I reaction center subunit XI (A8IL32)
Predicted protein (A8JH60)
Photosystem II D2 protein (P06007)
Photosystem II protein D1 (P07753)
Photosystem I P700 chlorophyll a apoprotein A2 (P09144)
Photosystem II CP43 reaction center protein (P10898)
Photosystem I P700 chlorophyll a apoprotein A1 (P12154)
Photosystem I reaction center subunit IV (PSI-E) (P12352)
Photosystem I reaction center subunit VI (PSI-H) (P13352)
Photosystem II CP47 reaction center protein (P37255)
Photosystem I reaction center subunit II (Q39615)
Protein annotation
0
1
2
3
4
UVM4 protein abundance (log2)

(C)

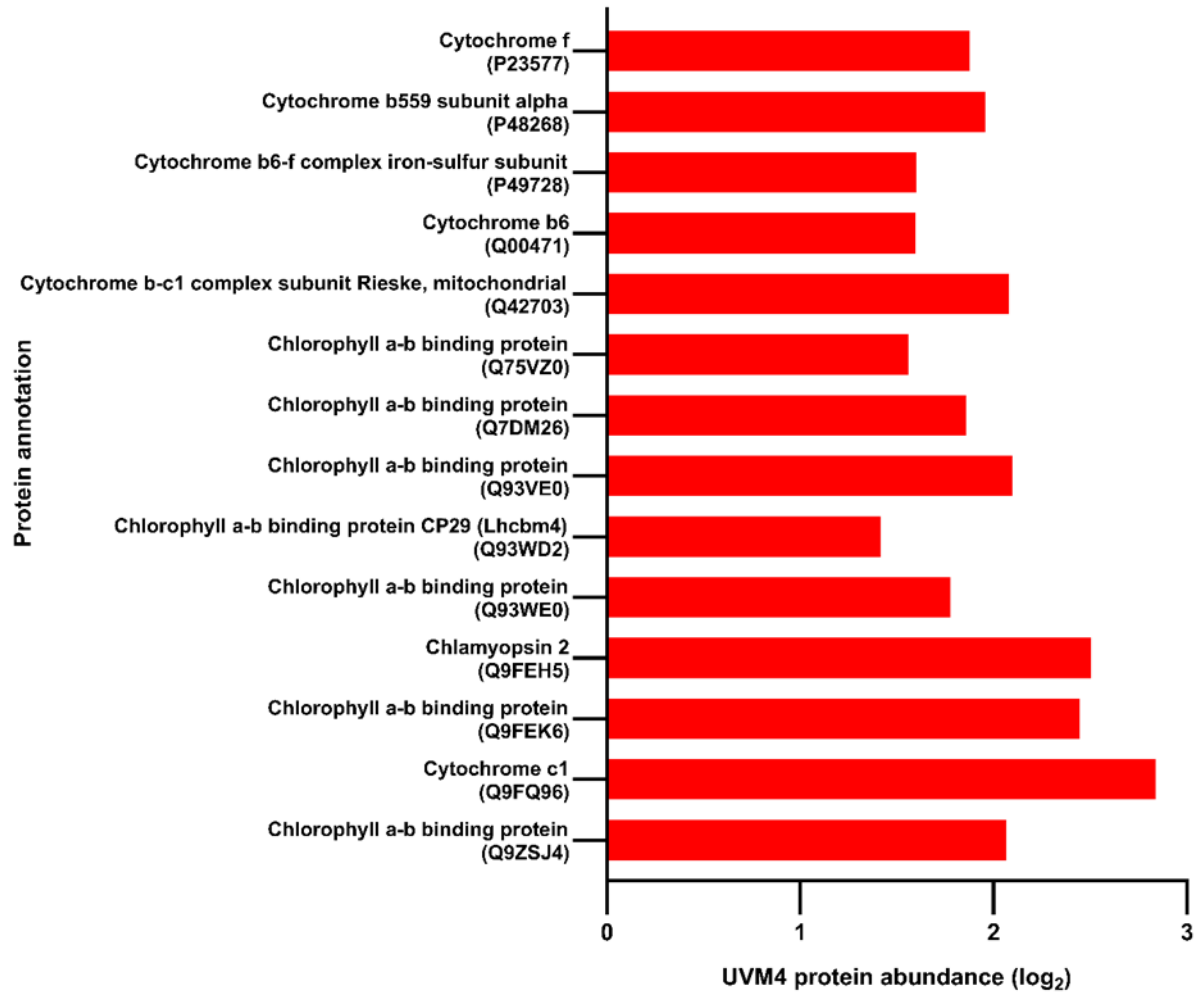

Cytochrome f (P23577)
Cytochrome b559 subunit alpha (P48268)
Cytochrome b6-f complex iron-sulfur subunit (P49728)
Cytochrome b6 (Q00471)
Cytochrome b-c1 complex subunit Rieske, mitochondrial (Q42703)
Chlorophyll a-b binding protein (Q75VZ0)
Chlorophyll a-b binding protein (Q7DM26)
Chlorophyll a-b binding protein (Q93VE0)
Chlorophyll a-b binding protein CP29 (Lhcbm4) (Q93WD2)
Chlorophyll a-b binding protein (Q93WE0)
Chlamyopsin 2 (Q9FEH5)
Chlorophyll a-b binding protein (Q9FEK6)
Cytochrome c1 (Q9FQ96)
Chlorophyll a-b binding protein (Q9ZSJ4)
Protein annotation
0
1
2
3
UVM4 protein abundance (log2)

Figure 2.4. Relative abundance of strain UVM4 intracellular proteins involved in: A) chloroplast membrane regulation and transmembrane transport; B) photosystem I & II regulation; C) chlorophyll-binding proteins and cytochromes, using strain 137c protein abundance as reference. Proteins with a relative abundance > 1 are considered more abundant in strain UVM4 than in strain 137c (red) and vice-versa for relative abundances < -1 (blue), n = 3.

All the differentially abundant proteins involved in photosynthetic processes, except one, were more abundant in strain UVM4 compared to strain 137c (Figure 2.4A, 2.4B, and 2.4C). Most proteins showing higher relative abundance are involved in the activity and regulation of the reaction centre of photosystem I & II (PSI and PSII) (Figure 2.4B), and chlorophyll-binding proteins and cytochromes (Figure 2.4C). The higher abundance of these proteins in strain UVM4 may relate to their different growth stage at time of sampling (Figure 2.2B). Even though there was not a significant difference in optical density at time of harvest, strain UVM4 was in exponential phase, while 137c was entering stationary phase. Therefore, it is possible that the photosynthetic system was more active in strain UVM4 or that these proteins were being produced because strain UVM4 cells were still dividing. Extended growth rates and consequent higher biomass is directly related to higher yield of RPs, and thus can be listed as one of the possible reasons for higher production of secreted RPs in strain UVM4 (reported in [13,14]).

The single less abundant protein in strain UVM4 was the protein translocase subunit SecA (UniProt accession number: A8J682) (Figure 2.4A). SecA is an ATPase-coupled protein with a transmembrane transporter activity. This protein is predicted to be involved in many processes, such as chloroplast organization, protein import, protein targeting, and regulation of photosynthesis. In *Arabidopsis thaliana*, the protein translocase subunit SecA (UniProt accession number: Q9SYI0) is involved in photosynthesis acclimation and chloroplast biogenesis [23,24]. Subunit SecA is one of the components of the chloroplast Sec pathway, which transport proteins synthesised in the nucleus to the thylakoid membrane and lumen during photosynthesis [23,25]. In particular, SecA hydrolyses ATP to generate energy to activate and regulate the transfer of proteins across the thylakoid membrane [24]. Considering

the higher abundance of all other proteins involved in photosynthesis in strain UVM4, including two membrane AAA-metalloproteases involved in thylakoid membrane organisation (UniProt accession numbers: A8IL08 and A8J6C7) (Fig. 2.4A), it seems counter-intuitive to find SecA in lower abundance at the time of harvesting. However, positive and negative regulation of nuclear gene expression via retrograde signalling from damaged or modified chloroplasts has been reported in *C. reinhardtii* [26]. Chlorophyll biosynthetic precursors located in the chloroplast (such as magnesium-protoporphyrin IX) directly regulate the nuclear expression of chlorophyll-binding proteins, which will subsequently coordinate the biosynthesis of chlorophyll [26]. Future studies are therefore needed to investigate retrograde signalling in UVM4 cells and potentially demonstrate similar behaviour for the expression of the protein translocase subunit SecA.

4.2.4. Cell wall

The absence of a cell wall facilitates the transformation process of *C. reinhardtii* to produce genetically modified strains [27], explaining in part the wide use of strain UVM4 for RP production. In cell wall-deficient strains, previous studies have reported that cell wall proteins are still successfully produced; however, they do not assemble in the multi-layered cell wall and they are subsequently released into the extracellular space [28,29,30]. Therefore, cell wall proteins present in the extracellular proteome of strain UVM4 were analysed (Fig. 2.5).

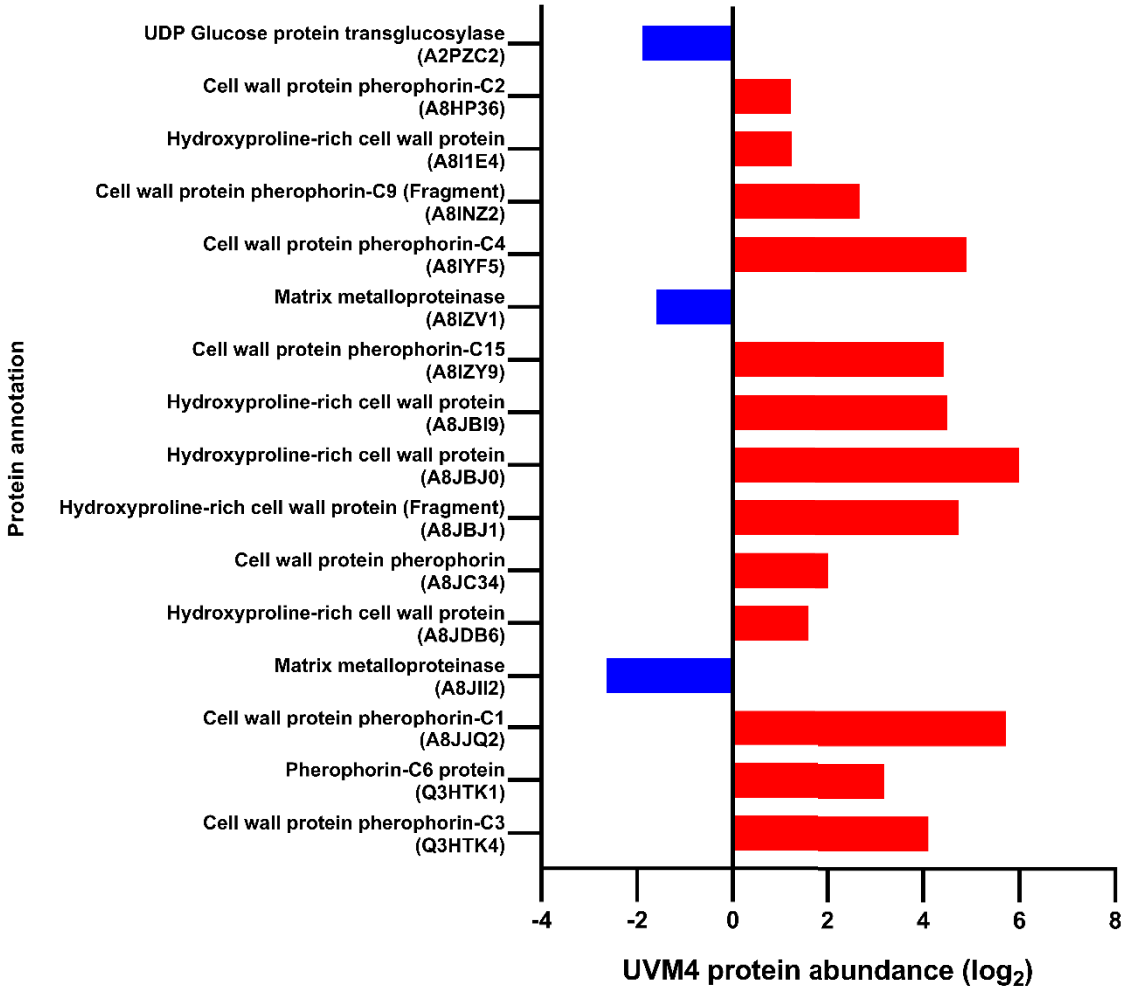


Figure 2.5. Relative abundance of cell wall proteins in strain UVM4 using protein abundance in strain 137c as reference. The proteins showing higher abundance in strain UVM4 than in 137c are in red, while the less abundant proteins are in blue, n = 3.

In congruence with the literature, many of the differentially produced cell wall proteins were found to be more abundant in the UVM4 extracellular fraction. Of these, half were 16 – 64 (i.e. 2^4 – 2^6) times more abundant in strain UVM4 than the same proteins produced in strain 137c, representing a substantial increase in cell wall proteins being secreted into the media of strain UVM4. All the cell wall proteins showing higher abundance in strain UMV4 were either hydroxyproline-rich proteins or pherophorins. Hydroxyproline-rich proteins are fundamental cell wall glycoproteins involved in

structural roles [31], whilst pherophorins are cell wall glycoproteins structurally related to a sex-inducing pheromone [32].

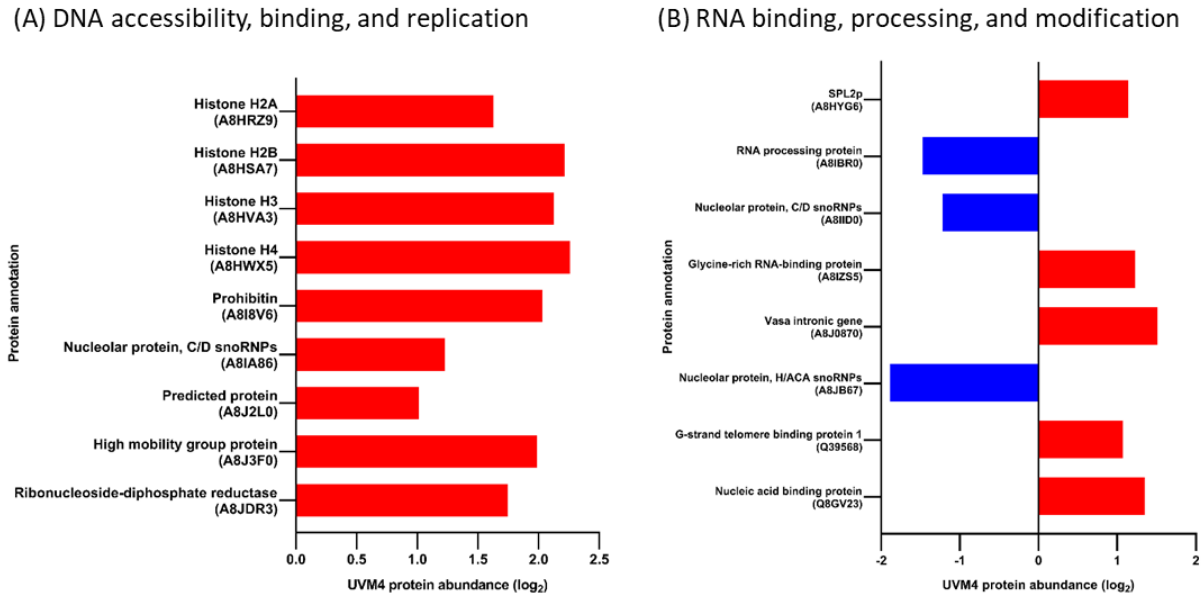
The extracellular cell wall proteins that were less abundant in UVM4 included a UDP-Glucose:protein transglucosylase (UniProt accession number: A2PZC2), which has been characterised in *Solanum tuberosum* [33]. It is involved in the biogenesis of cell wall polysaccharides. The lower abundance of this protein in UVM4 may be due to the absence of a cell wall to anchor the generated polysaccharides. The two other proteins with lower abundance are both metalloproteinases. These proteins (A8IZV1 and A8JII2) are involved in metalloendopeptidase activity, specifically targeting proline- and hydroxyproline-rich cell wall proteins of *C. reinhardtii*. The lower abundance of metalloproteinases in strain UVM4 secretome may be contributing to the observed accumulation of hydroxyproline-rich cell wall proteins in the extracellular fraction through reduced degradation of these types of proteins.

When hydroxyproline-rich cell wall proteins accumulate in the extracellular space (Fig. 2.5), they can form aggregates that negatively affect RP downstream purification processes and final yield [9]. Therefore, albeit cell wall deficiency is beneficial for increased transformation efficiency, it can be linked to impaired RP purification and low final yields [9]. Baier and colleagues circumvented this issue by fusing an amphiphilic hydrophobin protein tag to their secreted RP and using a detergent-based aqueous two-phase protein extraction system as the first step of purification prior to affinity chromatography. However, the final yield of purified RP was still low (15 μg L^{-1}) [9].

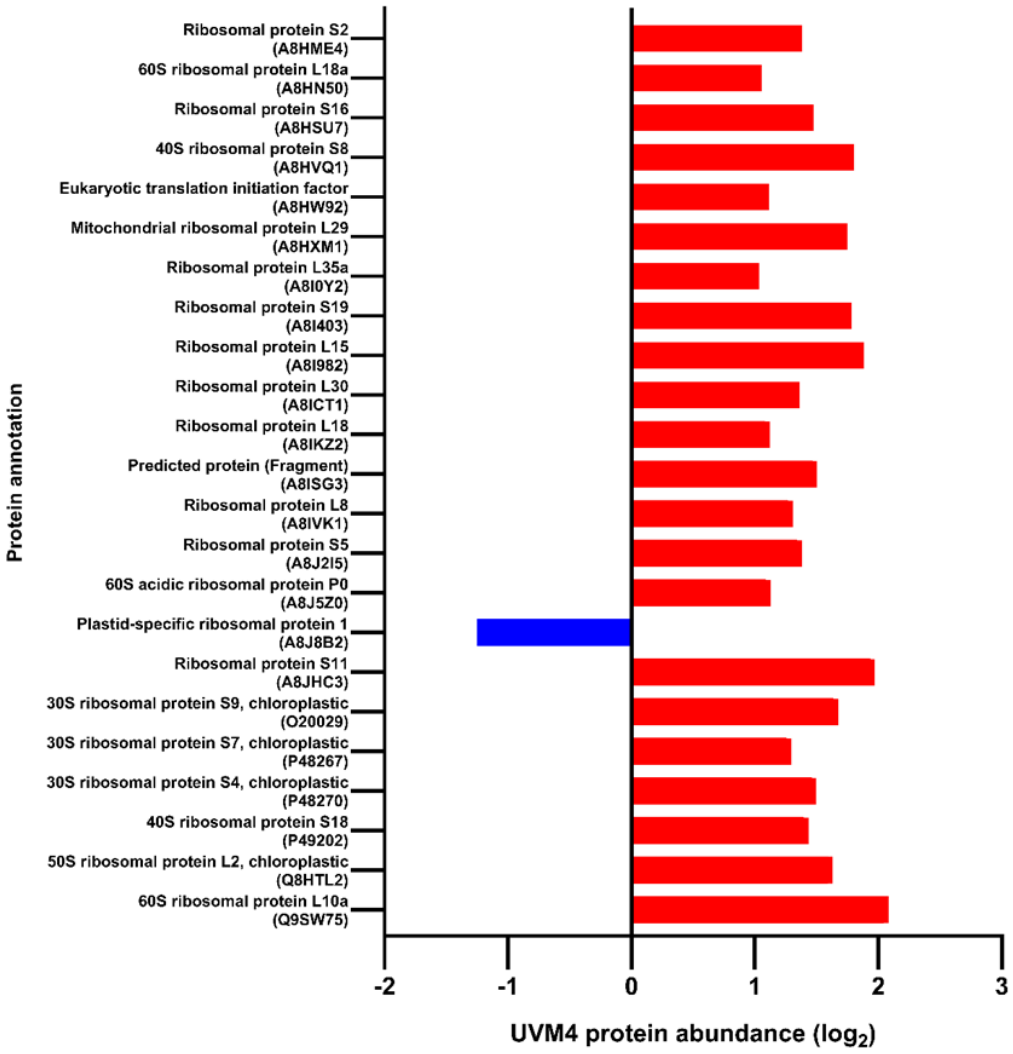
Impaired RP purification due to hydroxyproline-rich cell wall proteins populating the extracellular space is a well-known issue for cell wall deficient strains [9]. However, in this study I identified possible targets for molecular biology approaches that could facilitate RP purification and enhance yields. Possible strategies are knock-down/knock-out of the insoluble hydroxyproline-rich glycoproteins that form aggregates, or knock-in of the metalloproteinases responsible for hydroxyproline-rich proteins degradation.

4.2.5. Nucleic acid processing, transcription, translation, and protein degradation

Strain UVM4 is capable of remarkably high yields of secreted RP produced by nuclear transgenes [10,13,14]. Since RP production and accumulation depend on many different factors, such as DNA accessibility, gene silencing mechanisms, DNA transcription, mRNA translation, and protein degradation, the intracellular and extracellular proteomes were analysed searching for differentially abundant proteins involved in these pathways.



(C) DNA transcription



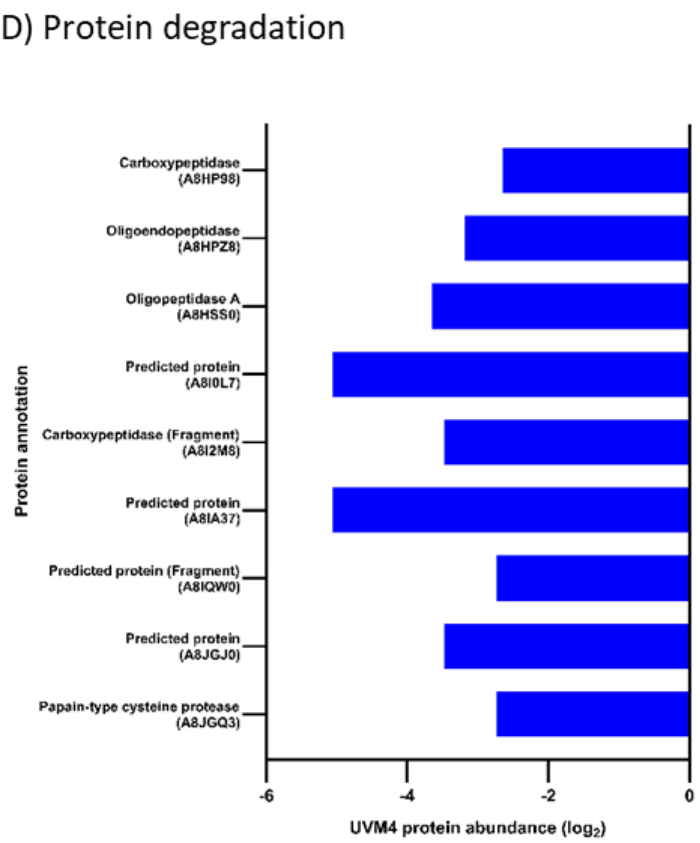


Figure 2.6. Relative abundance of strain UVM4 intracellular (A, B, C) and extracellular (D) proteins in comparison with strain 137c proteins (used as reference) involved in: (A) DNA accessibility, binding, and replication; (B) RNA binding, processing, and modification; (C) DNA transcription; (D) protein degradation. The proteins showing higher abundance in strain UVM4 than in strain 137c are in red, while the less abundant proteins are in blue, n = 3.

4.2.5.1. DNA processing

All the intracellular proteins involved in DNA processing showed higher abundance in strain UVM4 (Figure 2.6A). Histones H2A, H2B, H3, and H4 were among the more abundant proteins in the intracellular proteome of strain UVM4, with a fold change > 3.03 ($2^{1.6}$) (Fig. 2.6A). In the nucleus, core histones H2A, H2B, H3 and H4 are bound together to form an octamer that binds DNA and wraps it into a complex called the nucleosome [34,35,36]. Histones and their linkage to DNA play a major role in transcription regulation by condensing or extending the chromatin structure and making DNA more or less accessible [37,38]. Finding all the four core histones more abundant in strain UVM4 is consistent with the literature, as these four proteins are typically present in the nucleus in stoichiometric amounts

[39,40,41]. It is also coherent to find the high mobility group protein (HMGB) (UniProt accession number: A8J3F0) among the more abundant intracellular proteins in strain UVM4. In fact, in mammalian cells, the role of this protein is to bend DNA and facilitate the insertion of histones during formation of the nucleosome, and the absence of HMGB is directly related to a lower content of core histones [42]. By any means, a higher amount of core histones and consequent tightness of chromatin in a strain renowned for its high secreted RP yields might seem counterintuitive. However, having more nucleosomes does not necessarily translate to lower transcription rates. In fact, it has been proven that nucleosomes are also involved in positive regulation of transcription. Wyrick and colleagues [43] analysed the effect of histone depletion on gene expression in yeast, to find that only 15% of genes showed higher expression rates. Surprisingly, 75% of genes were not significantly altered, whilst 10% of genes even showed lower expression [43]. Wyrick and colleagues hypothesise that nucleosomes probably have a minimal role in transcription, and that its regulation lays upon transcriptional repressors and activators (such as HMGB protein). Multiple studies reinforced these results, showing that nucleosomes can regulate the activity and synergy of transcription factors based on their position on the DNA strand [44,45,46]. Therefore, the absence of the nucleosome can negatively affect the synergism of the multiple transcription co-factors and enzymes, reducing gene expression rates. It is now clear that higher histone amounts do not directly correlate to lower expression of genes, therefore the higher production of secreted RP proteins in UVM4 could still, in part, arise from higher transcription capacity.

The condensation and the distension of the chromatin structure (and the consequent enhancement or diminishment of transcription) is regulated by PTMs of the histones [47]. Reported PTMs of core histones linked to gene regulation are methylation (lysine, arginine, and glutamine), acetylation (lysine), and phosphorylation (serine, threonine, and tyrosine) [47,48,49]. More specifically, some modifications are related to gene silencing, whilst other are involved in gene expression. The thorough review by Schroda [50] has reported modifications of the N-termini of H3 and H4 involved in gene expression and/or silencing in *C. reinhardtii*. Monomethylation of H3 lysine 4 and 9, mono and dimethylation of H3 lysine 27, and phosphorylation of H3 threonine 3, are all histone PTMs related to gene silencing. Trimethylation of H3 lysine 4 and acetylation of H3 and H4 multiple lysine, on the other hand, are

PTMs involved in gene expression enhancement [50]. None of the enzymes involved in these listed histone modifications were found to be significantly different in strain UVM4. Another study hypothesized that in yeasts and humans, methylation of H2A glutamine is directly related to enhanced transcription [48,49]. Interestingly, the enzyme responsible for the methylation of yeast H2A glutamine 105 is called NOP1, which is reported here in this study as more abundant in strain UVM4 (C/D box snoRNP, *NOP1*, UniProt accession number: A8IA86) (Fig. 2.6A). A higher abundance of the NOP1 protein might modify the structure of chromatin to help a cell cope with high demands for transcription of ribosomal components. Hence, considering that core histones H2A, H2B, H3, and H4 and their PTMs are well preserved in eukaryotes [51], it is possible that a higher abundance of this protein in strain UVM4 is related to enhanced transcription rates. However, core histones have been characterised in plants (*A. thaliana*, [52]; *Saccharum sp.*, [47]) and in *C. reinhardtii* [53,54], and interestingly, none of these studies found any methylated glutamine on histone H2A. Nonetheless, considering that the annotation of NOP1 protein in *C. reinhardtii* is based on homology [21,55], and that histone modifications in *C. reinhardtii* show unique variants [54], further analysis is needed to fully comprehend the role of the NOP1 protein in *C. reinhardtii*.

Before drawing any conclusion on more abundant core histones and their role in strain UVM4, two more factors need to be considered. Firstly, the linker histone H1 did not show higher abundance in strain UVM4. To form the final chromatin structure, multiple nucleosomes need to bind together to form a 30 nm fibre. This DNA-condensation step is catalysed by the linker histone H1 [52]. The ratio of the histone H1 and the other four is not stoichiometric, nevertheless a huge disproportion between linker and octamer could still cause modification in the chromatin structure, possibly leading to more accessible DNA and higher transcription. The linker histone H1 was detected only in the intracellular proteomic profile of strain 137c (unique intracellular protein of strain 137c, Figure 2.3), therefore it was not possible to obtain a relative abundance of this protein. Nonetheless, the absence of the linker histone H1 in the proteomic profile of strain UVM4 (especially considering the higher abundance of the core histones) is noteworthy. However, further analysis is needed to fully understand the role of histone H1 in strain UVM4. Secondly, it is possible that the differences in growth rates of the two strains favoured

the higher abundance of core histones in strain UVM4. In fact, histone production is strongly related to cell growth and widely varies during the cell cycle. Many studies have linked cell senescence and histone depletion [56,57,58]. Considering that strain UVM4 was still in mid-exponential phase while 137c was entering stationary phase, it is possible that the difference in histone amount was related to the strain growth stage. The data obtained in this study do not allow me to draw unambiguous conclusions on higher DNA accessibility and possible up-regulation of transcription in UVM4; however, RP yield depends on a combination of multiple factors and DNA availability is only one of them. The reason for higher yields of secreted recombinant protein reported in strain UVM4 might be a reliance on lower gene silencing, higher translation, and/or lower proteolysis.

4.2.5.2. RNA processing

Figure 2.6B shows 5 more abundant and 3 less abundant intracellular proteins, all involved in RNA processing in strain UVM4. The first less abundant protein is a messenger RNA processing protein (UniProt accession number: A8IBR0). Information regarding this protein is limited. Its predicted biological process is to bind mRNA, however it is unknown if this protein is involved in positive or negative regulation of transcription or translation. The other two less abundant proteins are both involved in small nucleolar ribonucleoprotein (snoRNP) complexes. A snoRNP is a complex formed of small nucleolar RNAs (snoRNAs) and four or more proteins that form part of the spliceosome. The spliceosome is located in the nucleus and it is responsible for the splicing (removal of introns) of precursor mRNAs, generating mRNAs that are subsequently translated into proteins [59]. The snoRNP complex is essential to the removal of introns from pre-mRNA, and is present in two configurations: box C/D and box H/ACA. Box C/D snoRNPs are involved in methylation mechanisms, while box H/ACA snoRNPs are involved in pseudouridylation [60]. The precise effect of methylation and pseudouridylation on the function of the mature RNAs is not yet elucidated, but these modifications appear to reduce RNA hydrolysis, subtly enhancing the RNA folding and interaction with ribosomal proteins [61]. One of the snoRNP detected (NOP58, UniProt accession number: A8IID0) is a nucleolar protein component of C/D snoRNPs, while the second one (NHP2, UniProt accession number: A8JB67) is a constituent of the small subunit of H/ACA snoRNPs. The lower abundance of proteins involved in

mRNA processing in strain UVM4 is counterintuitive, as mRNA are essential for protein synthesis and strain UVM4 has been reported to produced higher amount of secreted recombinant proteins than wild type [13,14]. However, one protein involved in the assembly of the spliceosome complex (SPL2, UniProt accession number: A8HYG6) and all the other proteins involved in RNA processing were found to be more abundant in strain UVM4 (Fig 2.6B).

4.2.5.3. Translation

Except for one protein, all the intracellular proteins showing significant different abundance were ribosomal and all showed higher abundance in strain UVM4 (Fig. 2.6C). Interestingly, the only protein showing lower abundance was a plastid-specific ribosomal protein (accession number: A8J8B2), and its role is to down-regulate translation in the chloroplast. Translation of chloroplast proteins is strictly related to photosynthesis [62], therefore a lower abundance of a negative regulator of chloroplast translation in strain UVM4 is consistent with the results shown in section *4.2.3*, with higher abundance of proteins involved in photosynthesis. Moreover, chloroplastic proteins involved in translation were all more abundant in strain UVM4 than strain 137c (Fig. 2.6C), again following the trend of section *4.2.3* (Fig 2.4). The other proteins more abundant in strain UVM4 are cytoplasmic ribosomal proteins (Fig. 2.6C), suggesting that a possible reason for higher secreted RP yields reported in strain UVM4 might be a general increase in translation activity [63].

4.2.5.4. Proteolysis

Along with protein translation, another major cellular mechanism that affects RP yields is protein degradation. RPs are produced by the host system and accumulate in the intracellular or the extracellular space, where they can be degraded by proteases. The intracellular proteome did not show differentially produced proteins involved in degradation, whereas the extracellular proteome showed interesting results with nine proteins involved in proteolysis mechanisms, all showing significantly lower abundance in strain UVM4 (Fig. 2.6D). Among these proteins, two serine-type carboxypeptidases (accession numbers: A8HP98 and A8I2M8) involved in hydrolysis of C-terminal peptide bond in polypeptide chains were detected (Fig. 2.6D). In addition, some of the cysteine proteases (UniProt

accession numbers: A8I0L7 and A8JGQ3) predicted to be present in the lysosome and/or in the extracellular space of *C. reinhardtii* based on bioinformatics analysis [21] were detected in this study. Lastly, the protein Oligoendopeptidase (UniProt accession number: A8HPZ8), the protein Oligopeptidase A (UniProt accession number: A8HSS0), and predicted protein (UniProt accession number: A8IA37), are all metalloproteinases. The metalloproteinases shown in Figure 2.5 (section 4.2.4) were targeting proline- and hydroxyproline-rich cell wall proteins. The metalloproteinases shown in Figure 2.6D, on the other hand, are involved in proteolysis of any class of protein. Irrespective of target, it is important to highlight that all the proteins involved in proteolysis were less abundant in UVM4. Thus, these results suggest that the lower abundance of proteins involved in degradation mechanisms might be linked to the higher yields of secreted RP reported in strain UVM4 transgenic cell lines.

4.2.6. Post-translational modifications (PTMs)

As mentioned above, PTMs play a significant role in stability, functionality, and activity of proteins [64]. Phosphorylation, acetylation, nitration, disulphide bonds, and glycosylation are some examples of PTMs involved in fundamental cellular processes [65]. Therefore, the intracellular and extracellular proteomes were analysed searching for significant differentially produced PTM enzymes, to investigate the PTM machinery status of strain UMV4. Both intracellular and extracellular proteomes did not show any significant differences in PTM enzyme abundance. However, this result does not necessarily indicate an absence of PTM differences between the two strains. In fact, PTMs enzyme activity is also strictly dependent on chemical and biological modifications (e.g. pH, temperature, hormonal regulation of enzyme activity) [66,67]. Therefore, further PTM analyses are still needed to postulate about differences in PTMs between strains UVM4 and 137c (performed in Chapter 4).

5. CONCLUSIONS

This study provides the first proteomic analysis of *C. reinhardtii* strain UVM4 and its comparison to the wild type strain 137c, giving further insight into the molecular reprogramming occurring in the cell wall-deficient strain UVM4 (summarized in Figure 2.7). Strain UVM4 is a smaller non-motile strain that reaches stationary phase 24 hours after the wild type strain 137c. Cell wall deficiency results in production and subsequent secretion of a substantial amount of cell wall proteins into the extracellular space. However, the most interesting findings were related to proteins involved in nucleic acid processing, transcription, translation, and protein degradation. All four core histones were surprisingly more abundant in strain UVM4. Moreover, the majority of proteins involved in nucleic acid processing were also found in higher abundance in UVM4. Of the 23 detected proteins involved in transcription, 22 were more abundant in strain UVM4, and the single less abundant protein was involved in negative regulation of transcription, supporting the hypothesis that higher secreted RP yields reported in strain UVM4 might be related to a general increase in translation activity. And lastly, all the extracellular proteins involved in proteolysis were less abundant in strain UVM4 suggesting that the higher yields of secreted RP obtained in this strain could be linked to the prevention or inhibition of proteolysis.

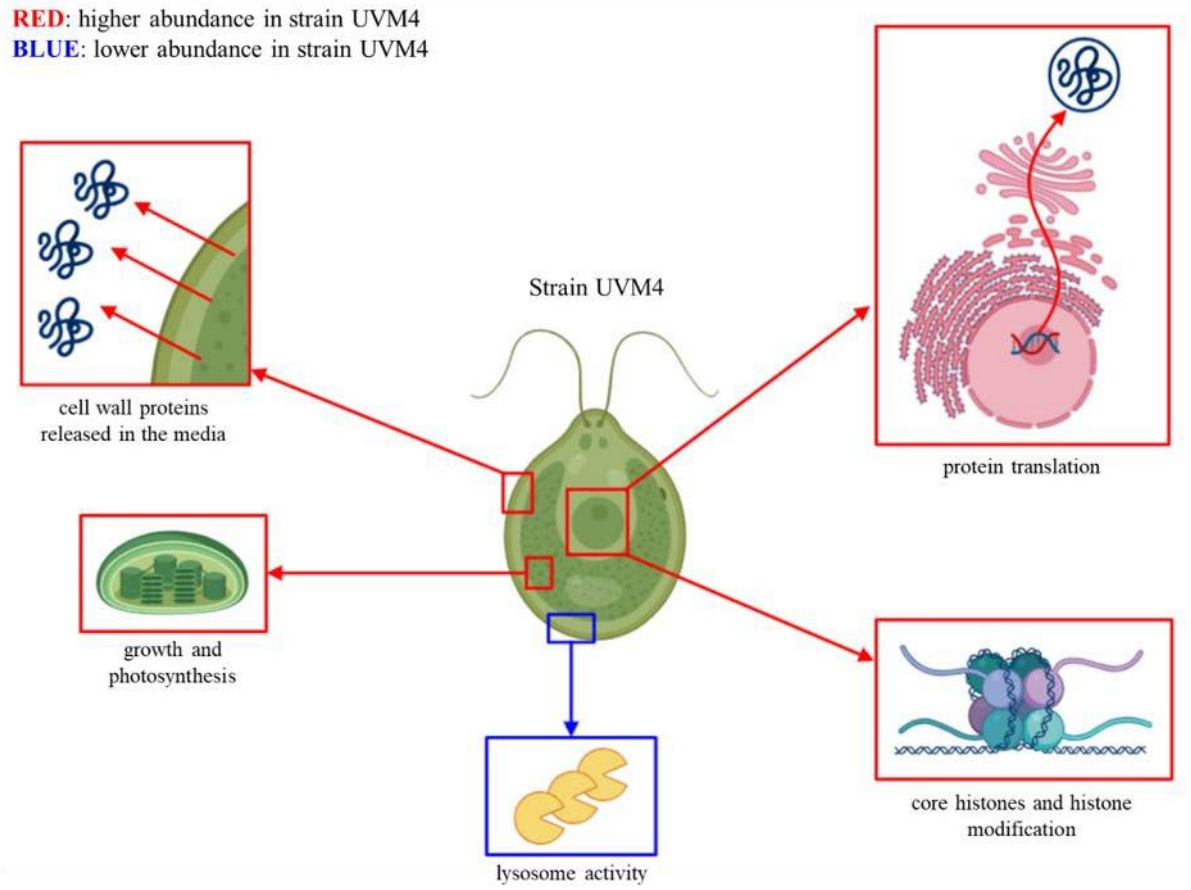
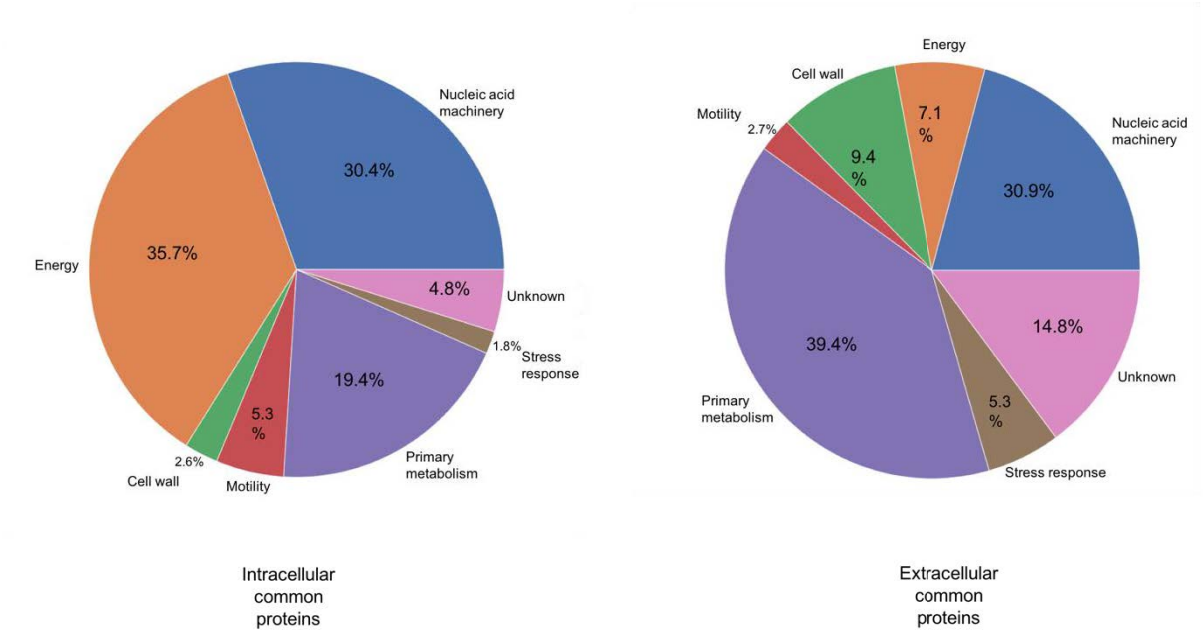


Figure 2.7. Schematic diagram summarizing key regulations observed by comparative proteomics in this study, indicating molecular mechanisms potentially driving phenotypic differences and improved recombinant protein production in strain UVM4.

C. reinhardtii strain UMV4 combines the advantages of photosynthetic cell hosts with remarkable RP yields [13,14], and over the past decade it has shown to be a powerful biotechnological tool. Until now, the molecular pathways altered in strain UVM4 remained unexplored. Here I report the first evidence of higher abundance of ribosomal proteins and lower abundance of proteases, two processes involved in RP production and possibly related to the higher yields reported in strain UVM4 [13,14]. The proteins involved in these pathways could be the target for knock-in, knock-down, or knock-out approaches in future studies aimed at further enhancing RP yields in other *C. reinhardtii* strains or even other microalgal species. Strain UVM4 is a derivative of strain 137c, however this is not the direct parental strain. The best strain to compare the intracellular proteome of the mutant to would be strain Elow47, derivative of CC-4350 and immediate predecessor of strain UVM4. Future studies should perform a

three-way comparison between 137c, Elow47, and UMV4, to fully comprehend the effect of the UV mutation on strain UVM4. In this study, I also report extensive presence of cell wall proteins in the extracellular space, a well-known issue for RP purification in cell wall deficient strains. These extracellular proteins could be the targets for knock-down or knock-out approaches to improve downstream purification processes. However, it is possible that only a portion of these proteins form the aggregates responsible for RP purification complications. Therefore, to facilitate genetic engineering work, future studies should investigate exactly which cell wall proteins are involved in the formation of the extracellular aggregates. Lastly, the analysis shown in this Chapter involves only wild type strain 137c and UV mutated strain UVM4. In Chapter 3, I performed the same proteomic analysis on genetically modified strains of 137c and UVM4, to investigate pathways activated or deactivated by the transformation process.

SUPPLEMENTARY FIGURES



Supplementary Figure 2.1: Common proteins present in intracellular and extracellular fractions categorised based on Gene Ontology (GO): biological process, molecular function, and cellular component.

6. REFERENCES

1. Eichler-Stahlberg, A.; Weisheit, W.; Ruecker, O.; Heitzer, M. Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. *Planta* **2009**, *229*, 873-883, doi:10.1007/s00425-008-0879-x.

2. Rasala, B.A.; Mayfield, S.P. The microalga *Chlamydomonas reinhardtii* as a platform for the production of human protein therapeutics. *Bioeng Bugs* **2011**, *2*, 50-54, doi:10.4161/bbug.2.1.13423.

3. Mathieu-Rivet, E.; Kiefer-Meyer, M.C.; Vanier, G.; Ovide, C.; Burel, C.; Lerouge, P.; Bardor, M. Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front Plant Sci* **2014**, *5*, 359, doi:10.3389/fpls.2014.00359.

4. Specht, E.; Miyake-Stoner, S.; Mayfield, S. Micro-algae come of age as a platform for recombinant protein production. *Biotechnol Lett* **2010**, *32*, 1373-1383, doi:10.1007/s10529-010-0326-5.

5. De Muynck, B.; Navarre, C.; Boutry, M. Production of antibodies in plants: status after twenty years. *Plant Biotechnol J* **2010**, *8*, 529-563, doi:10.1111/j.1467-7652.2009.00494.x.

6. Rasala, B.A.; Mayfield, S.P. Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. *Photosynth Res* **2015**, *123*, 227-239, doi:10.1007/s11120-014-9994-7.

7. Fahad, S.; Khan, F.A.; Pandupuspitasari, N.S.; Ahmed, M.M.; Liao, Y.C.; Waheed, M.T.; Sameeullah, M.; Darkhshan; Hussain, S.; Saud, S., et al. Recent developments in therapeutic protein expression technologies in plants. *Biotechnol Lett* **2015**, *37*, 265-279, doi:10.1007/s10529-014-1699-7.

8. Molino, J.V.D.; de Carvalho, J.C.M.; Mayfield, S.P. Comparison of secretory signal peptides for heterologous protein expression in microalgae: Expanding the secretion portfolio for Chlamydomonas reinhardtii. *PLoS One* **2018**, *13*, e0192433, doi:10.1371/journal.pone.0192433.

9. Baier, T.; Kros, D.; Feiner, R.C.; Lauersen, K.J.; Muller, K.M.; Kruse, O. Engineered Fusion Proteins for Efficient Protein Secretion and Purification of a Human Growth Factor from the Green Microalga *Chlamydomonas reinhardtii*. *ACS Synth Biol* **2018**, 10.1021/acssynbio.8b00226, doi:10.1021/acssynbio.8b00226.

10. Neupert, J.; Karcher, D.; Bock, R. Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J* **2009**, *57*, 1140-1150, doi:10.1111/j.1365-313X.2008.03746.x.

11. Schroda, M.; Blöcker, D.; Beck, C.F. The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *The Plant Journal* **2000**, *21*, 121-131, doi:10.1046/j.1365-313x.2000.00652.x.

12. Rasala, B.A.; Lee, P.A.; Shen, Z.; Briggs, S.P.; Mendez, M.; Mayfield, S.P. Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS One* **2012**, *7*, e43349, doi:10.1371/journal.pone.0043349.

13. Lauersen, K.J.; Huber, I.; Wichmann, J.; Baier, T.; Leiter, A.; Gaukel, V.; Kartushin, V.; Rattenholl, A.; Steinweg, C.; von Riesen, L., et al. Investigating the dynamics of recombinant protein secretion from a microalgal host. *J Biotechnol* **2015**, *215*, 62-71, doi:10.1016/j.jbiotec.2015.05.001.

14. Ramos-Martinez, E.M.; Fimognari, L.; Sakuragi, Y. High-yield secretion of recombinant proteins from the microalga *Chlamydomonas reinhardtii*. *Plant Biotechnol J* **2017**, *15*, 1214-1224, doi:10.1111/pbi.12710.

15. Ikehata, H.; Ono, T. The Mechanisms of UV Mutagenesis. *Journal of Radiation Research* **2011**, *52*, 115-125, doi:10.1269/jrr.10175.

16. Mizukami, A.; Caron, A.L.; Picanço-Castro, V.; Swiech, K. Platforms for Recombinant Therapeutic Glycoprotein Production. In *Recombinant Glycoprotein Production: Methods and Protocols*, Picanço-Castro, V., Swiech, K., Eds. Springer New York: New York, NY, 2018; 10.1007/978-1-4939-7312-5_1pp. 1-14.

17. Gorman, D.S.; Levine, R.P. Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardi*. *Proc Natl Acad Sci U S A* **1965**, *54*, 1665-1669, doi:10.1073/pnas.54.6.1665.

18. Wessel, D.; Flugge, U.I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **1984**, *138*, 141-143, doi:10.1016/0003-2697(84)90782-6.

19. Hughes, C.S.; Moggridge, S.; Muller, T.; Sorensen, P.H.; Morin, G.B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* **2019**, *14*, 68-85, doi:10.1038/s41596-018-0082-x.

20. Lin, H.; He, L.; Ma, B. A combinatorial approach to the peptide feature matching problem for label-free quantification. *Bioinformatics* **2013**, *29*, 1768-1775, doi:10.1093/bioinformatics/btt274.

21. Gaudet, P.; Livstone, M.S.; Lewis, S.E.; Thomas, P.D. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief Bioinform* **2011**, *12*, 449-462, doi:10.1093/bib/bbr042.

22. Nelson, J.; Savereide, P.B.; Lefebvre, P.A. The CRY1 gene in *Chlamydomonas reinhardtii*: structure and use as a dominant selectable marker for nuclear transformation. *Molecular and cellular biology* **1994**, *14*, 4011-4019.

23. Liu, D.; Gong, Q.; Ma, Y.; Li, P.; Li, J.; Yang, S.; Yuan, L.; Yu, Y.; Pan, D.; Xu, F., et al. cpSecA, a thylakoid protein translocase subunit, is essential for photosynthetic development in *Arabidopsis*. *J Exp Bot* **2010**, *61*, 1655-1669, doi:10.1093/jxb/erq033.

24. Skalitzky, C.A.; Martin, J.R.; Harwood, J.H.; Beirne, J.J.; Adamczyk, B.J.; Heck, G.R.; Cline, K.; Fernandez, D.E. Plastids contain a second sec translocase system with essential functions. *Plant Physiol* **2011**, *155*, 354-369, doi:10.1104/pp.110.166546.

25. Dietzel, L.; Brautigam, K.; Pfannschmidt, T. Photosynthetic acclimation: state transitions and adjustment of photosystem stoichiometry--functional relationships between short-term and long-term light quality acclimation in plants. *FEBS J* **2008**, *275*, 1080-1088, doi:10.1111/j.1742-4658.2008.06264.x.

26. Nott, A.; Jung, H.S.; Koussevitzky, S.; Chory, J. Plastid-to-nucleus retrograde signaling. *Annu Rev Plant Biol* **2006**, *57*, 739-759, doi:10.1146/annurev.arplant.57.032905.105310.

27. Kindle, K.L. High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci U S A* **1990**, *87*, 1228-1232, doi:10.1073/pnas.87.3.1228.

28. Davies, D.R.; Plaskitt, A. Genetical and structural analyses of cell-wall formation in *Chlamydomonas reinhardi*. *Genetical Research* **2009**, *17*, 33-43, doi:10.1017/s0016672300012015.

29. Voigt, J.; Hinkelmann, B.; Harris, E.H. Production of cell wall polypeptides by different cell wall mutants of the unicellular green alga *Chlamydomonas reinhardtii*. *Microbiological Research* **1997**, *152*, 189-198, doi:10.1016/s0944-5013(97)80012-2.

30. Cronmiller, E.; Toor, D.; Shao, N.C.; Kariyawasam, T.; Wang, M.H.; Lee, J.H. Cell wall integrity signaling regulates cell wall-related gene expression in *Chlamydomonas reinhardtii*. *Sci Rep* **2019**, *9*, 12204, doi:10.1038/s41598-019-48523-4.

31. Sommer-Knudsen, J.; Bacic, A. A micro-scale method for determining relative metal-binding affinities of proteins. *Mol Biotechnol* **1997**, *8*, 215-218, doi:10.1007/BF02760774.

32. Godl, K.; Hallmann, A.; Rappel, A.; Sumper, M. Pherophorins: a family of extracellular matrix glycoproteins from Volvox structurally related to the sex-inducing pheromone. *Planta* **1995**, *196*, 781-787.

33. Wald, F.A.; Kissen, R.; Du Jardin, P.; Moreno, S. Characterization of UDP-glucose: protein transglucosylase genes from potato. *Plant molecular biology* **2003**, *52*, 705-714.

34. Arents, G.; Burlingame, R.W.; Wang, B.C.; Love, W.E.; Moudrianakis, E.N. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences of the United States of America* **1991**, *88*, 10148-10152, doi:10.1073/pnas.88.22.10148.

35. Arents, G.; Moudrianakis, E.N. Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 10489-10493, doi:10.1073/pnas.90.22.10489.

36. Richmond, T.J.; Rechsteiner, T.; Luger, K. Studies of nucleosome structure. In Proceedings of Cold Spring Harbor symposia on quantitative biology; pp. 265-272.

37. Turner, B.M. Cellular Memory and the Histone Code. *Cell* **2002**, *111*, 285-291, doi.

38. Berger, S.L. Histone modifications in transcriptional regulation. *Current Opinion in Genetics & Development* **2002**, *12*, 142-148, doi.

39. Feser, J.; Truong, D.; Das, C.; Carson, J.J.; Kieft, J.; Harkness, T.; Tyler, J.K. Elevated histone expression promotes life span extension. *Molecular cell* **2010**, *39*, 724-735.

40. O'sullivan, R.J.; Kubicek, S.; Schreiber, S.L.; Karlseder, J. Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nature structural & molecular biology* **2010**, *17*, 1218.

41. Celona, B.; Weiner, A.; Di Felice, F.; Mancuso, F.M.; Cesarini, E.; Rossi, R.L.; Gregory, L.; Baban, D.; Rossetti, G.; Grianti, P. Substantial histone reduction modulates genomewide nucleosomal occupancy and global transcriptional output. *PLoS biology* **2011**, *9*.

42. Karnavas, T.; Pintonello, L.; Agresti, A.; Bianchi, M.E. Histone content increases in differentiating embryonic stem cells. *Front Physiol* **2014**, *5*, 330, doi:10.3389/fphys.2014.00330.

43. Wyrick, J.J.; Holstege, F.C.; Jennings, E.G.; Causton, H.C.; Shore, D.; Grunstein, M.; Lander, E.S.; Young, R.A. Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature* **1999**, *402*, 418-421, doi:10.1038/46567.

44. Li, B.; Carey, M.; Workman, J.L. The role of chromatin during transcription. *Cell* **2007**, *128*, 707-719, doi:10.1016/j.cell.2007.01.015.

45. Chávez, S.; Beato, M. Nucleosome-mediated synergism between transcription factors on the mouse mammary tumor virus promoter. *Proceedings of the National Academy of Sciences* **1997**, *94*, 2885-2890.

46. Prado, F.; Koop, R.; Beato, M. Accurate chromatin organization of the mouse mammary tumor virus promoter determines the nature of the synergism between transcription factors. *J Biol Chem* **2002**, *277*, 4911-4917, doi:10.1074/jbc.M110094200.

47. Moraes, I.; Yuan, Z.F.; Liu, S.; Souza, G.M.; Garcia, B.A.; Casas-Mollano, J.A. Analysis of Histones H3 and H4 Reveals Novel and Conserved Post-Translational Modifications in Sugarcane. *PLoS One* **2015**, *10*, e0134586, doi:10.1371/journal.pone.0134586.

48. Tessarz, P.; Santos-Rosa, H.; Robson, S.C.; Sylvestersen, K.B.; Nelson, C.J.; Nielsen, M.L.; Kouzarides, T. Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* **2014**, *505*, 564-568, doi:10.1038/nature12819.

49. Leonhardt, H.; Hake, S.B. Histone glutamine methylation afFACTing rDNA transcription. *Cell Res* **2014**, *24*, 261-262, doi:10.1038/cr.2014.22.

50. Schroda, M. Good News for Nuclear Transgene Expression in *Chlamydomonas*. *Cells* **2019**, *8*, doi:10.3390/cells8121534.

51. Freeman, L.; Kurumizaka, H.; Wolffe, A.P. Functional domains for assembly of histones H3 and H4 into the chromatin of Xenopus embryos. *Proceedings of the National Academy of Sciences* **1996**, *93*, 12780-12785.

52. Zhang, K.; Sridhar, V.V.; Zhu, J.; Kapoor, A.; Zhu, J.K. Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. *PLoS One* **2007**, *2*, e1210, doi:10.1371/journal.pone.0001210.

53. Waterborg, J.H.; Robertson, A.J.; Tatar, D.L.; Borza, C.M.; Davie, J.R. Histones of *Chlamydomonas reinhardtii* (Synthesis, Acetylation, and Methylation). *Plant physiology* **1995**, *109*, 393-407.

54. Khan, A.; Eikani, C.K.; Khan, H.; Iavarone, A.T.; Pesavento, J.J. Characterization of *Chlamydomonas reinhardtii* core histones by top-down mass spectrometry reveals unique algae-specific variants and post-translational modifications. *Journal of proteome research* **2018**, *17*, 23-32.

55. Merchant, S.S.; Prochnik, S.E.; Vallon, O.; Harris, E.H.; Karpowicz, S.J.; Witman, G.B.; Terry, A.; Salamov, A.; Fritz-Laylin, L.K.; Marechal-Drouard, L., et al. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **2007**, *318*, 245-250, doi:10.1126/science.1143609.

56. Prado, F.; Jimeno-Gonzalez, S.; Reyes, J.C. Histone availability as a strategy to control gene expression. *RNA Biol* **2017**, *14*, 281-286, doi:10.1080/15476286.2016.1189071.

57. Maya Miles, D.; Penate, X.; Sanmartin Olmo, T.; Jourquin, F.; Munoz Centeno, M.C.; Mendoza, M.; Simon, M.N.; Chavez, S.; Geli, V. High levels of histones promote whole-genome-duplications and trigger a Swe1(WEE1)-dependent phosphorylation of Cdc28(CDK1). *Elife* **2018**, *7*, doi:10.7554/eLife.35337.

58. Strenkert, D.; Schmollinger, S.; Gallaher, S.D.; Salome, P.A.; Purvine, S.O.; Nicora, C.D.; Mettler-Altmann, T.; Soubeyrand, E.; Weber, A.P.M.; Lipton, M.S., et al. Multiomics resolution of molecular events during a day in the life of *Chlamydomonas*. *Proc Natl Acad Sci U S A* **2019**, *116*, 2374-2383, doi:10.1073/pnas.1815238116.

59. Reiter, N.J.; Chan, C.W.; Mondragon, A. Emerging structural themes in large RNA molecules. *Curr Opin Struct Biol* **2011**, *21*, 319-326, doi:10.1016/j.sbi.2011.03.003.

60. De Zoysa, M.D.; Yu, Y.-T. Posttranscriptional RNA pseudouridylation. In *The Enzymes*, Elsevier: 2017; Vol. 41, pp. 151-167.

61. Bachellerie, J.-P.; Cavaillé, J.; Hüttenhofer, A. The expanding snoRNA world. *Biochimie* **2002**, *84*, 775-790.

62. Mayfield, S.P.; Cohen, A.; Danon, A.; Yohn, C.B. Translation of the psbA mRNA of *Chlamydomonas reinhardtii* requires a structured RNA element contained within the 5' untranslated region. *J Cell Biol* **1994**, *127*, 1537-1545, doi:10.1083/jcb.127.6.1537.

63. Zhou, X.; Liao, W.J.; Liao, J.M.; Liao, P.; Lu, H. Ribosomal proteins: functions beyond the ribosome. *J Mol Cell Biol* **2015**, *7*, 92-104, doi:10.1093/jmcb/mjv014.

64. Bürkle, A. Posttranslational Modification. In *Encyclopedia of Genetics*, Brenner, S., Miller, J.H., Eds. Academic Press: New York, 2001. 1533.

65. Nikov, G.; Bhat, V.; Wishnok, J.S.; Tannenbaum, S.R. Analysis of nitrated proteins by nitrotyrosine-specific affinity probes and mass spectrometry. *Analytical Biochemistry* **2003**, *320*, 214-222, doi:10.1016/s0003-2697(03)00359-2.

66. Zoldos, V.; Novokmet, M.; Beceheli, I.; Lauc, G. Genomics and epigenomics of the human glycome. *Glycoconj J* **2013**, *30*, 41-50, doi:10.1007/s10719-012-9397-y.

67. Butler, M.; Spearman, M. The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr Opin Biotechnol* **2014**, *30*, 107-112, doi:10.1016/j.copbio.2014.06.010.

CHAPTER 3: Transgene expression elicits widespread proteomic reprogramming in *Chlamydomonas reinhardtii*

Lorenzo Barolo[1], Raffaela M. Abbriano[1], Audrey S. Commault[1], Matthew P. Padula[2], and Mathieu Pernice[1]

AFFILIATIONS

[1] University of Technology Sydney, Climate Change Cluster, Broadway Campus, Ultimo NSW 2007, Sydney, Australia

[2]School of Life Sciences and Proteomics Core Facility, Faculty of Science, University of Technology Sydney, Ultimo NSW 2007, Sydney, Australia

CONTRIBUTIONS: LB, RMA, ASC, and MP designed the experiment; LB carried the experiment; LB and MPP performed the mass spectrometry analysis; LB, ASC, RMA, and MP performed the data analysis; LB wrote the manuscript with contributions from all authors.

This chapter has been formatted for submission to the journal: Algal Research.

1. ABSTRACT

Gene silencing mechanisms, position effect, and low nuclear transgene expression are major drawbacks for recombinant protein production in *Chlamydomonas reinhardtii*. While the generation of the highly expressive and cell wall-deficient strain UVM4 intended to overcome these issues, its performances as a biofactory remain not suitable for industrial application so far. To unveil the mechanisms counteracting transgene insertion and expression in this species, I performed genetic engineering on two *C. reinhardtii* strains, wild type strain 137c and the broadly used cell wall-deficient strain UVM4, and subsequently analysed their intracellular proteomes. The obtained transgenic cell lines of both strains showed major differences in protein abundance and multiple altered pathways after transformation, with principally lower abundance of proteins involved in chromatin remodelling, translation initiation and elongation, and protein quality control and transport. These results provide new insights into the response of *C. reinhardtii* to transgene insertion and expression, highlighting possible pathways involved in gene silencing and position effect mechanisms. Moreover, this study identifies multiple protein targets for future genetic engineering approaches, to improve *C. reinhardtii* performances as cell biofactory and consequently validate this species for industrial applications.

2. INTRODUCTION

Chlamydomonas reinhardtii (Chlorophyta) is a unicellular green microalga that has historically been used as a model system to investigate fundamental aspects of algal biology, including photosynthesis, cell wall biogenesis, and nuclear/chloroplast interactions [1]. *C. reinhardtii* biomass is also currently produced for use in various industries, such as aquaculture, cosmetics, and nutraceuticals [2]. In addition, recent studies demonstrated the ability of *C. reinhardtii* to produce valuable secondary metabolites [3,4] and recombinant biopharmaceuticals (RBs) [5,6,Commault et al. 2020 submitted].

C. reinhardtii presents many unique features suitable for industrial applications: i) it can be cultivated in large volumes and requires a lower up-front investment compared to bacterial or mammalian cell systems [7]; ii) it does not produce endotoxins, does not harbor human pathogens, and is generally recognized as safe (GRAS) [8]; and iii) high growth rates can be achieved in inexpensive media, resulting in potentially lower production costs when compared to other organisms [8,9]. In addition, the chloroplast, nuclear, and mitochondrial genomes have been fully sequenced [10], and transformation protocols for chloroplast and nuclear DNA are well established [1,11,12]. Chloroplast transgene expression can reach up to 20% of total soluble protein in *C. reinhardtii*; however, the plastid lacks the machinery to perform post-translational modifications (PTMs) [13,14]. Given the importance of PTMs for recombinant biopharmaceutical production, stability, and efficacy (as described in Chapter 1), chloroplast expression is not suitable for valuable complex glycoproteins (such as antibodies). RBs produced in the nucleus, on the other hand, can be targeted to the secretion route and can be subjected to PTMs [6].

Given that targeted nuclear transformation is still imprecise [15], nuclear transgene expression in *C. reinhardtii* is currently largely accomplished via random transgene insertion. Unfortunately, this approach is hindered by low transformation efficiency and extensive variability in transgene expression among transgenic cell lines [16,17,18]. These downsides are mainly due to position effects and gene silencing mechanisms [19,20]. It has been reported that the chromosomal integration site plays a major

role in transgene expression [21,22]. This phenomenon, called position effect, is possibly related to the position of the nucleosome on the DNA strand and its interaction with transcription repressors and activators [23,24,25]. Strong position effects result in high variation of transgene expression among transgenic cell lines generated with the same recombinant DNA. Unfortunately, these mechanisms are still not fully described nor understood [20]. Transgene silencing is mainly regulated by histone modification [18,20,22,26,27]. At the transgene insertion locus, enzymes will add PTMs to the histones to prompt a rearrangement of chromatin, condensing its structure [18,22,27,28]. Condensation of the chromatin structure at the site of insertion results in reduced transcription of the transgene [29,30,31]. Another mechanism affecting negatively transgene expression is methylation of recombinant DNA. In *C. reinhardtii*, it has been reported that methylation of transgene cytosine residues actively impede recombinant DNA transcription [26,31]. Transgene silencing is responsible for the reduction over time of cells expressing the transgene in the transgenic culture.

Many different strategies have arisen over time to achieve efficient nuclear expression in *C. reinhardtii* through optimisation of either the transgene design or the microalgal host. Transgene optimisation focused on a new generation of promoters [19], on the insertion of native introns in the exogenous DNA sequence [32,33,34], and on the addition of functional peptides to enhance expression and permit secretion [35,36]. Optimisation of the microalgal host system concentrated on testing different growth conditions [Commault et al. 2020 submitted], and on the generation of modified and/or mutated *C. reinhardtii* strains showing enhanced transgene expression [18,37]. UV mutated strains such as strain UVM4 [18] reported high yields of secreted recombinant protein (12 – 15 mg/L), which are 1.5 to 3-fold higher than *C. reinhardtii* wild type strains [38,39]. In Chapter 2, I performed the first comparative proteomic analysis between strain 137c and strain UVM4, revealing multiple altered pathways including particularly higher abundance of proteins involved in chromatin remodelling and ribosomal proteins. However, in Chapter 2, I did not analyse the potential effect of genetic engineering on the proteomes of strains 137 and UVM4 and further analysis is therefore required to unveil mechanisms counteracting transgene insertion and/or expression in *C. reinhardtii*.

This chapter reports the first proteomic analysis and comparison between transgenic cell lines and wild type of two *Chlamydomonas reinhardtii* strains, strain cc-125 *mt+* (also named strain 137c) and cell wall-deficient strain UVM4, to analyse the alteration in abundance of Open Reading Frame products to infer changes in gene expression caused by transgene insertion and expression. This study provides fundamental information on the pathways activated or deactivated by the organism to hinder the transformation process in both strains, with the final aim of developing strategies to improve *C. reinhardtii* for industrial applications.

3. MATERIALS AND METHODS

C. reinhardtii cultivation, harvesting, and transformation

C. reinhardtii strain 137c was purchased from Invitrogen (GeneArt® Chlamydomonas Protein Expression Kit). *C. reinhardtii* strain UVM4 was graciously provided by Prof. Ralph Bock. The plasmid pOptimized (pOpt) pOpt_mVenus_Paro (NCBI: KM061060, Fig S1) containing the mVenus and the paromomycin selection marker expression cassettes [32] was purchased from the Chlamydomonas Resource Center (https://www.chlamycollection.org/).

C. reinhardtii strains 137c and UVM4 were grown mixotrophically in Tris Acetate Phosphate (TAP) medium [40] under 50 µmol photons $m^{-2}s^{-1}$ of continuous light in a 25°C shaking incubator (100 rpm). Cells were collected at early exponential phase (2-4x10^6 cells mL^{-1}; Fig. S2) by centrifugation at 1,500 g for 5 minutes, washed twice with 10 mL of MAX Efficiency™ Transformation Reagent for Algae (Invitrogen™), and resuspended in MAX Efficiency™ buffer to a final concentration of 2-3x10^8 cells mL^{-1}. To prepare for electroporation, the concentrated cells (250 µL) and the linearized plasmid (2 – 4 µg, linearized with PsiI restriction enzyme) were transferred and mixed into an ice-cold electroporation cuvette (0.4 cm). The following settings on the Gene Pulser Xcell™ Electroporation System (Bio-Rad) were used: 500 V, 50 µF, and 800 Ω (for strain 137c) and 750 V, 25 µF and 0 Ω (for strain UVM4). Electroporation is needed to increase the permeability of the cell membranes, to permit insertion and incorporation of recombinant DNA. The absence of a cell wall in strain UVM4 required different electroporation settings, to allow survival of the cultures [41]. Electroporated cells were left to rest for 10 minutes at room temperature and then transferred into 10 mL of TAP-sucrose (40 mM) for overnight recovery. After recovery, the cells were centrifuged for 5 minutes at 1,000 g. The supernatant was discarded and the cells were resuspended in 200 µL of TAP medium. The resuspended cells were transferred on solid TAP-paromomycin (10 µg mL^{-1}) 1.5% agar plates. After one week, the colonies growing on plates were all transferred to 5 mL of liquid TAP-paromomycin (10 µg mL^{-1}) prior selection by cell sorting.

Selection and screening of positive *C. reinhardtii* transformants

After one week of growth in selective liquid medium, the cultures were subjected to fluorescence-activated cell sorting (FACS) on BD Influx. At first, *C. reinhardtii* cells were detected by their chlorophyll signals, excited by 488nm laser and detected in 692±40nm channel, and followed by single cell gating, area versus pulse width signals on trigger channel, to gate out cell aggregates. *C. reinhardtii* single cells were then displayed in FACS plot as chlorophyll signal versus mVenus signal, excited by 488nm laser and detected in 530±40nm channel. Untransformed wild type *C. reinhardtii* cells were used as negative control to set mVenus positive gate at ~0%. *C. reinhardtii* cells in the transformed samples that fell into the mVenus positive gate were selected and sorted one cell per well at single cell sort mode into in a 96 well plate containing 200 µL of TAP-paromomycin (10 µg mL^{-1}) medium per well. The selected cells were grown mixotrophically in TAP medium under 50 µmol photons $m^{-2}s^{-1}$ of continuous light in a 25°C shaking incubator (100 rpm) for one week. The surviving cultures were transferred to 4 mL TAP-paromomycin (10 µg mL^{-1}) medium and grown for three days in triplicates. On the third day, *C. reinhardtii* transgenic cell lines and wild types were analysed on Beckman Coulter CytoFLEX S flow cytometer to measure their chlorophyll signals, excited by 488nm laser and detected in 690±50nm channel, and mVenus signals, excited by 488nm laser and detected in 525±40nm channel, to identify the best candidates based on the percentage of positive cells using the same gating strategy used for FACS.

Cultivation and harvesting of *C. reinhardtii* transformants for proteomic analysis

The best five transgenic lines for strain 137c and strain UVM4 were transferred (1:250 inoculum ratio) to 500 mL of TAP medium and grown under mixotrophic conditions at 25°C and 100 rpm with ~50 µmol photons m^{-2} s^{-1} of light. Three replicates of each transgenic cell line were used to analyse growth and fluorescence over time. Fluorescence of the transgenic cell lines was analysed by flow cytometry each day for one week (using the same method as described above). Normalised fluorescence of the samples was measured as described in the equation (Eq3.1) below.

$$Norm\ F = \frac{\%positive\ cells \times MeanF}{MeanS} \times 1000 \qquad \text{(Eq3.1)}$$

With Norm F referring to the normalised fluorescence, MeanF corresponding to the mean fluorescence of positive cells and MeanS to the mean size (determined by the FSC-A) of positive cells.

A forth replicate was harvested for proteomic analysis when cell density reached ~1.5 $OD_{750\ nm}$ (mid-exponential phase, 72 hours of growth). Fluorescence of the samples was measured before harvesting, to confirm the presence of mVenus-positive cells in the culture. Cells and medium were separated by tangential flow microfiltration with a 0.2 µm polyethersulfone (PES) membrane (Vivaflow 200, Sartorius) to limit cell lysis as described in Chapter 2. Samples were stored at -80°C prior proteomics analysis.

Protein extraction and sample preparation

The harvested cells were sonicated to obtain the intracellular proteomes as described in Chapter 2. The intracellular proteomes were precipitated to remove non-protein contaminants, resuspended in 100 mM of triethylammonium bicarbonate buffer (TEAB) with 8M urea, and quantified using the Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific) protein assay as reported in Chapter 2. 40 µg of total protein were purified by single-pot, solid-phase-enhanced sample preparation (SP3) and digested with proteomic grade trypsin (Trypsin Gold, Promega) as shown in Chapter 2. The obtained peptides were quantified using Pierce™ Quantitative Colorimetric Peptide Assay (ThermoFisher Scientific), 2 µg of peptides were concentrated, resuspended in 5 µL of mass spectrometry loading buffer (2% acetonitrile (ACN), 0.2% trifluoroacetic acid (TFA) and analysed by mass spectrometry as described in Chapter 2.

Mass spectrometry proteome analysis

Mass spectrometry proteome analysis was performed using a Q Exactive Plus mass spectrometer (Thermo Scientific) as described in Chapter 2.

Data Analysis

The MS/MS data files were searched using Peaks Studio X against the UniProt *Chlamydomonas reinhardtii* proteome (UP000006906, protein count: 18,829) as described in Chapter 2.

Protein annotation

Intracellular proteins for transgenic cell lines and wild type were quantified using PEAKS label-free quantification as described in Chapter 2. Differentially expressed proteins were selected based on fold change (≥ 2) and identification significance (≥ 20).

Statistical analysis

Statistical analysis was performed as described in Chapter 2.

4. RESULTS AND DISCUSSION

4.1. Low transformation success rates and gene silencing mechanisms still impair genetic engineering in *C. reinhardtii* strain UVM4

Given the limited amount of cells expressing the transgene in *C. reinhardtii* due to the low transformation efficiency repeatedly reported in this species, and the importance of high abundance to report a proteomic profile representative of transgenic cell behaviour, the positive fluorescing single cells were sorted by flow cytometry to obtain single transgenic cell lines.

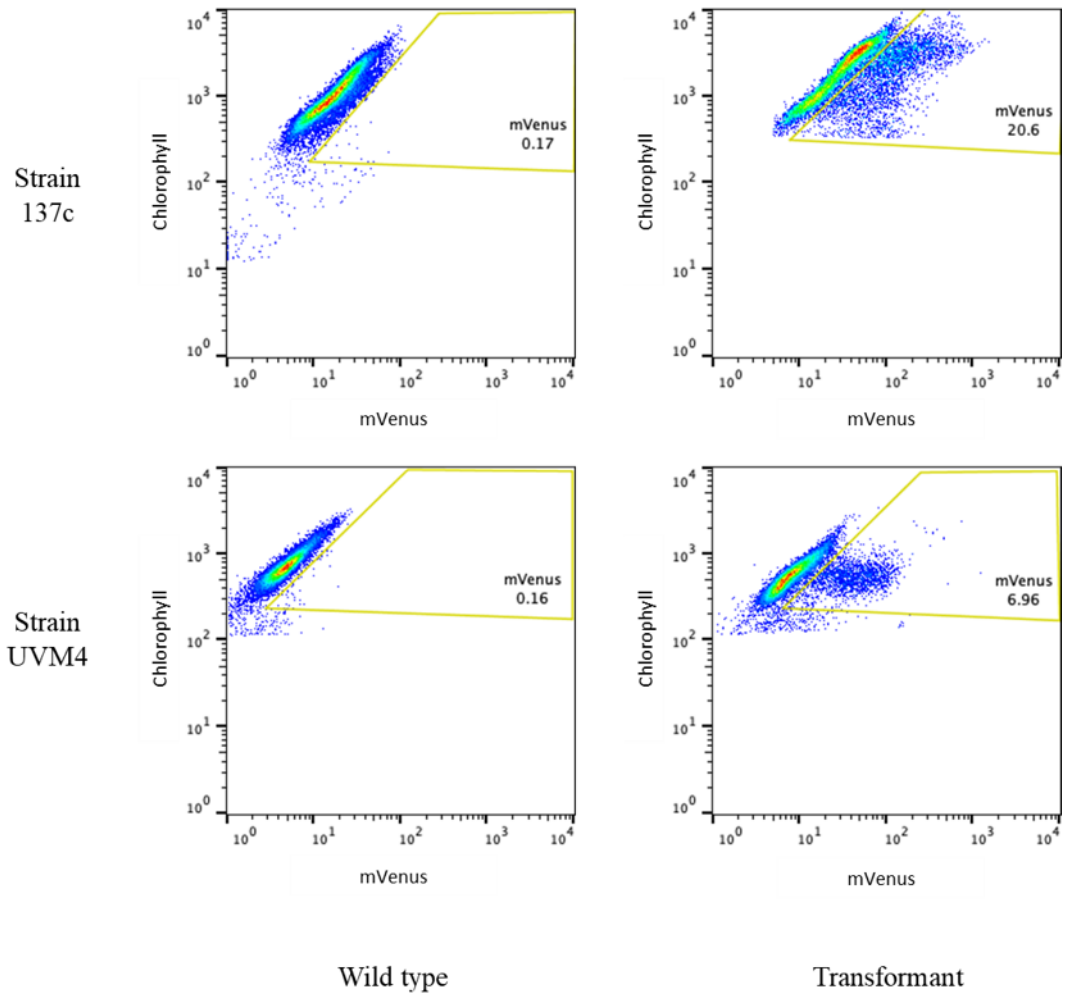


Figure 3.1. Dot plots of fluorescence-activated cell sorting (FACS) for transgenic cells of strains 137c and UVM4. The positive cells showing expression of the transgene were gated in green based on wild type cells fluorescence. Cells included in the "mVenus" yellow gate were selected for single cell sorting.

As expected, the transformation efficiencies for both strains were low; however, surprisingly strain UVM4 showed a lower success rate than strain 137c, with less than 10% of the cells in this strain being considered positive for mVenus fluorescence (based on the selected gate cut-off) (Figure 3.1).

After 7 days, the viability of the cultures derived from individual positive cells was assessed. For both strains, only ~50% of the 96 sorted cells survived. To select the best five candidates for the proteomic analysis, the surviving single cell cultures were split into triplicates and grown under selective pressure for additional screening by flow cytometry. Survival of the scaling-up process from 200 µL to 4 mL was also ~50%; therefore, a total of 11 cultures for strain 137c and 18 cultures for strain UVM4 were subjected to flow cytometry analysis (Figure 3.2).

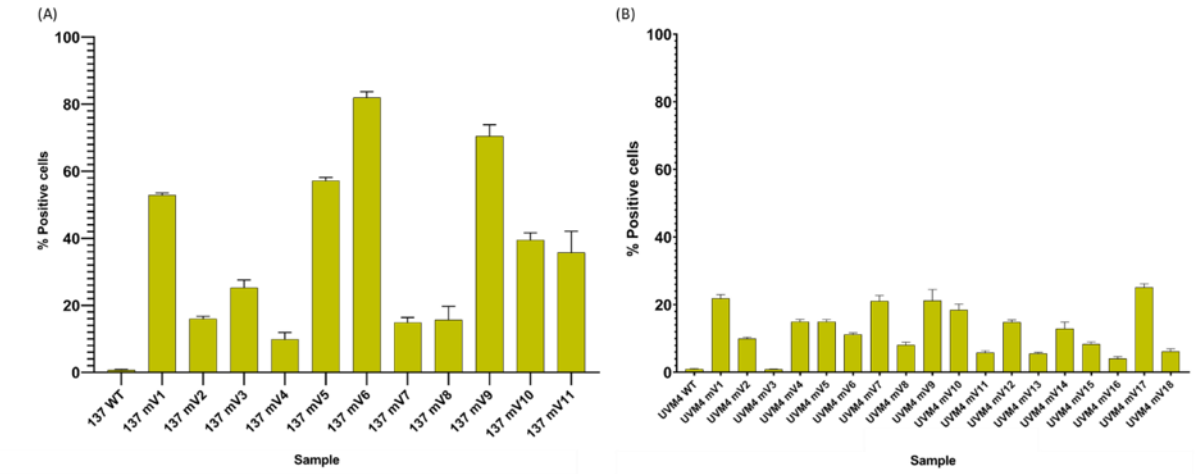


Figure 3.2. Percentage of mVenus positive cells in 11 transgenic lines of strain 137c (A) and 18 transgenic lines of strain UVM4 (B), gated based on their respective wild type behaviour. The data was obtained by flow cytometry analysis after 72 hours of growth under antibiotic selective pressure. n = 3.

Overall, both strains showed high fluctuation of percentage of positive cells among transgenic cell lines, spanning from 82% to 10% for strain 137c and from 25% to 1% for strain UVM4 (Figure 3.2A and 3.2B), possibly due to gene silencing mechanisms. However, strain 137c showed higher percentages of positive cells. In fact, the five best candidates for strain 137c all exhibited percentages between 40 – 80% of fluorescing cells (Figure 3.2A), whilst the top five cultures of strain UVM4 showed values

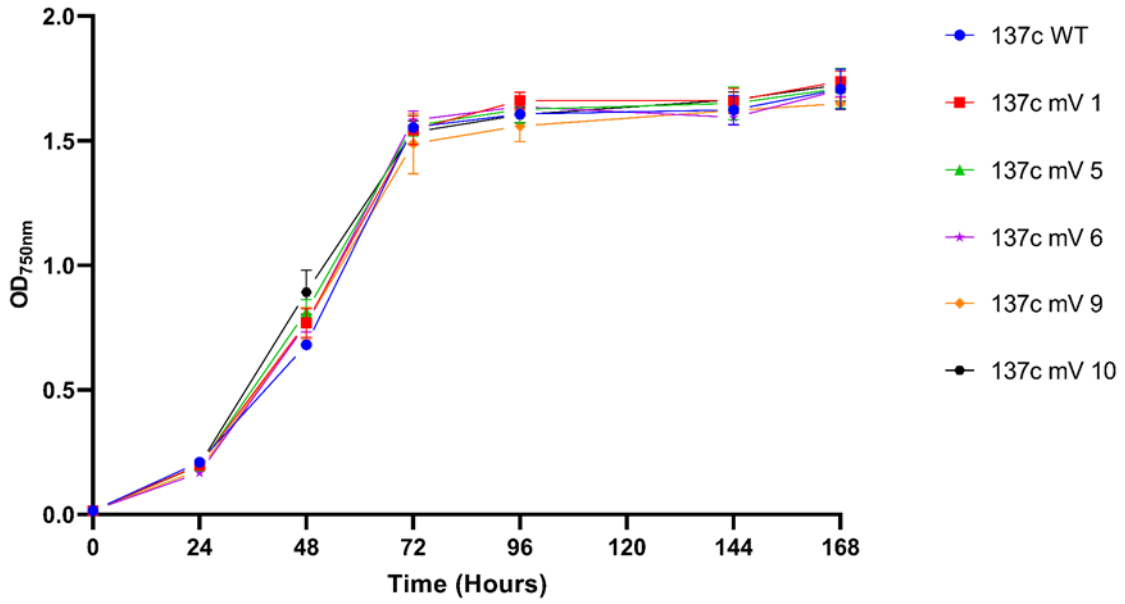
ranging between 20 – 25% (Figure 3.2B). Based on percentage of positive cells, strain 137c transgenic lines mV1, mV5, mV6, mV9, and mV10 were selected for proteomic analysis. For strain UVM4, transgenic lines mV1, mV7, mV9, mV10, and mV17 were selected. The selection criteria was based on percentage of positive cells to maximise the portion of the cell population expressing the transgene, thus obtaining a proteomic profile representative of the positive cells.

An interesting result for both strains is the complete absence of any culture showing >90% of positive cells, suggesting heterogeneity of transgene expression within genetically identical clonal populations. Since the cells were selected and sorted based on fluorescence, and kept under selective antibiotic pressure during the whole growth process, this result suggests gene silencing mechanisms [18,22,26] are extensively active and heavily neutralizing the expression of transgenes in *C. reinhardtii*, especially in strain UVM4.

4.2. Strain UMV4 is more affected by position effect than strain 137c

To elucidate the effects of transgene insertion and expression on *C. reinhardtii* strains 137c and UVM4 fitness, growth of the transgenic cell lines was analysed and compared to the wild types.

(A)



(B)

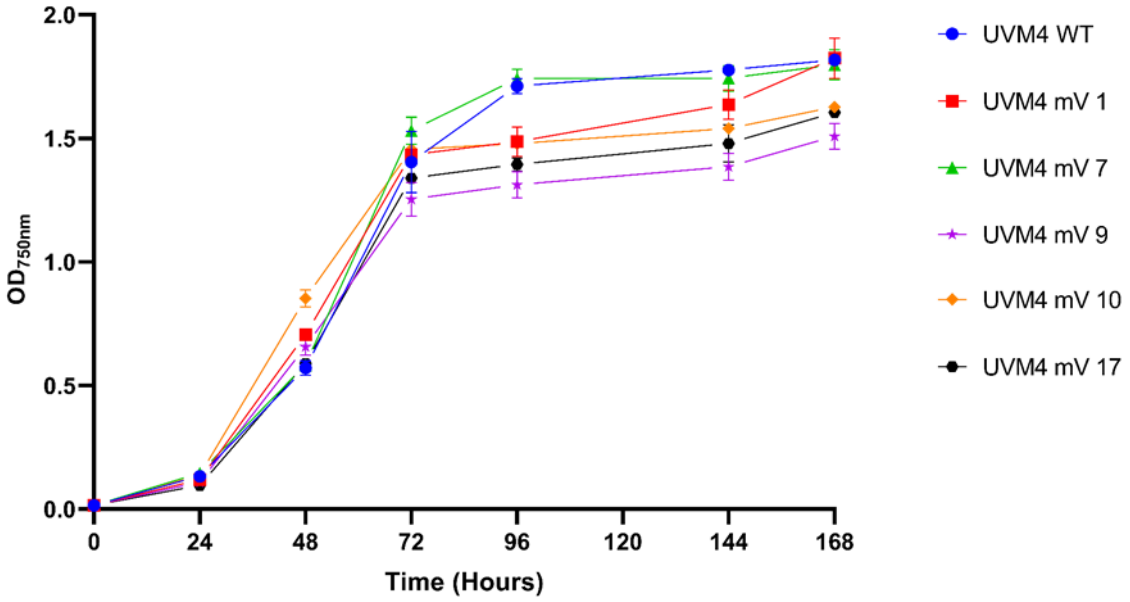
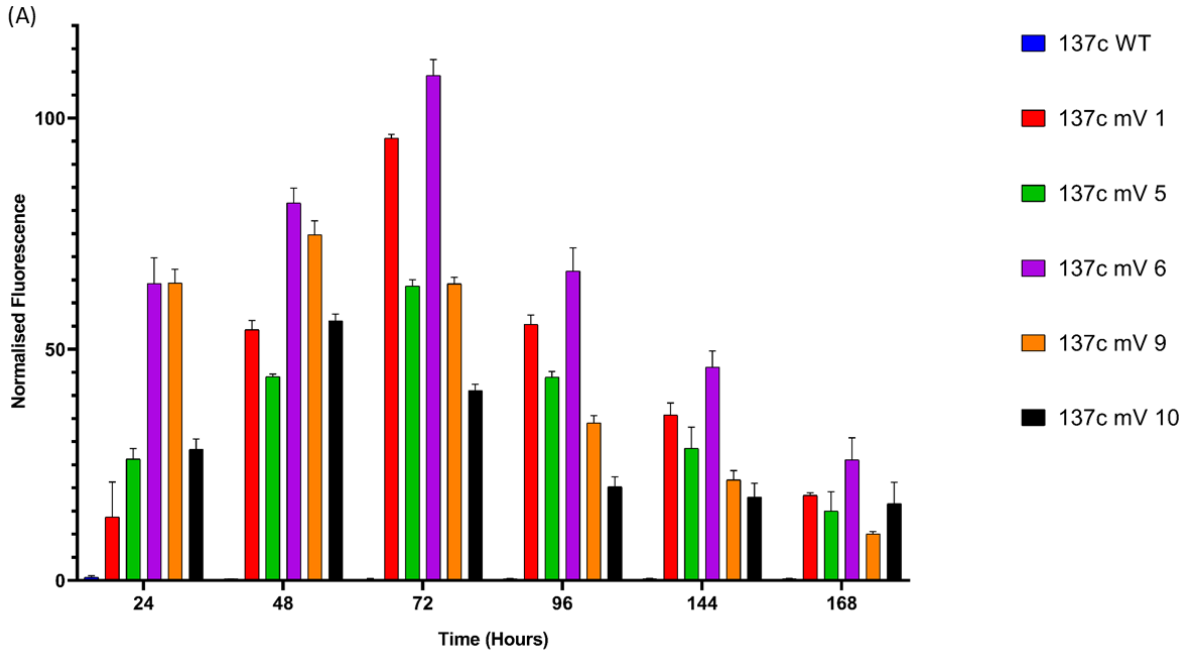


Figure 3.3. Comparison of growth profiles for wild type and transformant lines expressing mVenus for strains A) 137c and B) UVM4. Cells were grown in 500 mL of antibiotic-free media.

The growth curves of strain 137c wild type and transformants showed comparable trends and similar growth rates and all cultures reached stationary phase after 72 hours (Figure 3.3A). Strain UVM4 wild type and transformants, on the other hand, showed two different trends (Figure 3.3B). The wild type reached stationary phase after 92 hours as previously observed (Chapter 2, Figure 2.2); however, four out of five transgenic cell lines reached stationary phase after 72 hours, 24 hours earlier than the wild type. Transgene insertion and expression might have caused early entrance into stationary phase for these transgenic cell lines. It is possible that i) random integration of recombinant DNA disrupted endogenous genes fundamental for cell growth and survival, or ii) the over-production of a recombinant protein consumed an extensive amount of energy and resources, causing early stationary phase.

To analyse transgene expression over time, and to evaluate expression differences among transgenic cell lines, the normalised fluorescence of both strain transgenic cell lines was measured by flow cytometry and graphed (Figure 3.4). The mean fluorescence of gated positive cells (obtained by flow cytometry analysis) was normalised to cell size to be able to compare the expression of transgenes

between strains, considering the differences in cell size between strain 137c and strain UVM4 reported in Chapter 2. The wild type of both strains reported a normalised fluorescence always lower than 1.



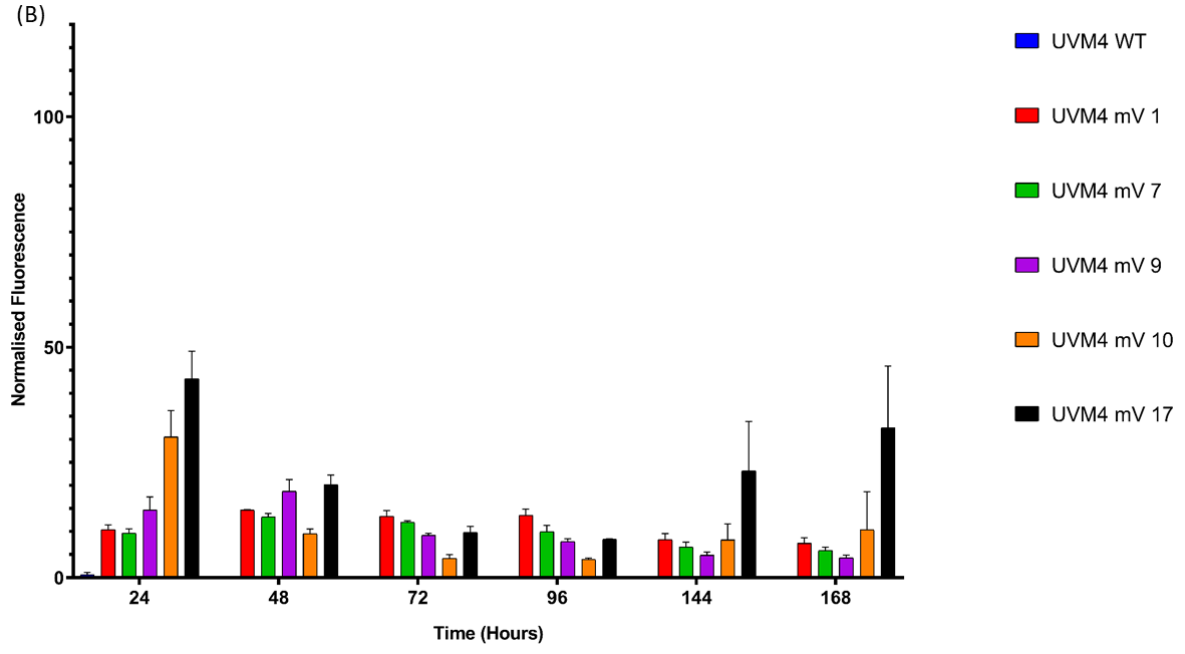


Figure 3.4. Normalised fluorescence in transgenic cell lines of strain 137c (A) and strain UVM4 (B) over time. Cells were grown in 500 mL of antibiotic-free media. Positive cells were detected based on wild type background fluorescence. Normalised fluorescence was calculated using equation (Eq3.1).

Strains 137c and UVM4 showed different patterns of normalised fluorescence over time. Interestingly, expression of mVenus was maintained in strains 137c and UVM4 transformants even when grown in media without selective pressure (Figure 3.4). The mVenus expression cassette was not fused with the antibiotic resistance cassette, therefore, selective pressure and fluorescence are not necessarily linked. However, if mVenus expression causes a disadvantage, the removal of selection may allow a higher number of negative cells to proliferate. The fact that both strains showed fluorescence higher than the wild type in antibiotic-free media, suggests that recombinant protein production can maintain consistency of transgene expression without antibiotics for at least 7 days.

For strain 137c, all transformants showed high values of normalised fluorescence (40 – 110) around 48 – 72 hours. After 168 hours, in stationary phase, these values dropped (< 26) (Figure 3.4A). Interestingly, all strain 137c transgenic cell lines behaved similarly, reaching peak of fluorescence in mid-late exponential phase and showing lower fluorescence in stationary phase. Transgenes were randomly integrated in the endogenous DNA, therefore major variations in transgene expression and protein production were expected among different transgenic cell lines. Considering the similar transgene expression observed among different transgenic cell lines of strain 137c, position effects seem to be weakly effective in this strain. Moreover, peak of expression in mid exponential phase have been previously reported for 137c transgenic lines, suggesting that it is independent of the recombinant protein being produced [Commault et al. 2020 submitted]. In contrast, strain UVM4 transformants showed low values of normalised fluorescence overall, with the exception of sample UVM4 mV10 and mV17, which reached fluorescence values comparable or even higher than strain 137c at stationary phase (Figure 3.4B). For strain UVM4, the transformants behaved differently, and consistency among different cell lines was not observed. Samples UVM4 mV10 and mV17 showed peaks of normalised

fluorescence at 24 hours, to drastically drop in mid-late exponential phase, to eventually rise again (especially sample UVM4 mV 17) around 144 – 168 hours. None of the other transgenic cell lines showed a similar behaviour. Sample UVM4 mV1, mV7, and mV9 showed consistent low values of normalised fluorescence throughout the whole growth. Considering the differences among transgenic cell lines in strain UVM4, position effect seems to have a higher impact in this strain.

Overall, strain UVM4 showed lower normalised fluorescence values than strain 137c. This seems counterintuitive, considering that strain UVM4 is renowned for its higher production and yields of secreted recombinant proteins [38,39]. However, there are few relevant factors to consider, which are possibly responsible for the contrasting results reported in this study. Firstly, random integration and position effects play a major role in *C. reinhardtii* recombinant protein production, especially in strain UVM4 (Figure 3.4B). Possibly a higher number of transformants is necessary to find a transgenic cell line showing higher normalised fluorescence values. Secondly, previous studies analysed yields of secreted recombinant protein, whilst this study focused on an intracellular protein. It is possible that the secretion route and the final location of the protein play a major role in accumulation of protein over time. The secretory pathway might be more active in strain UVM4, resulting in higher abundance of recombinant protein secreted and accumulated in the extracellular space. Moreover, protein degradation mechanisms are less active in the secretome of strain UVM4. In fact, in Chapter 2, I reported a lower abundance of proteins involved in lysosome activity in the secretome of strain UVM4, therefore it is likely that a recombinant protein accumulated in the extracellular space is degraded at slower rates. And finally, it is important to note that the proxy measured in this study (normalised fluorescence) might not reflect accurately recombinant protein yield in the different transformants. Further studies comparing the effect of transgene insertion and expression on strains 137c and UVM4 should therefore include direct measurement of final recombinant protein yield.

4.3. Proteomic analysis

The cells were harvested at 72 hours (late exponential/early stationary phase, Figure 3.3) for proteomic analysis. The intracellular proteomes of strains 137c and UVM4 (wild types and transgenic cell lines)

were compared using label-free quantification to identify the common proteins present in both strains, as reported in Chapter 2. The analysis was performed for two different proteomic datasets:

1. Strain 137c transgenic cell lines compared to strain 137c wild type (used as control);
2. Strain UVM4 transgenic cell lines compared to strain UVM4 wild type (used as control).

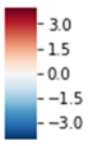
The analysis of transgenic cell lines compared to their respective wild types revealed several pathways altered by transgene insertion and expression. The two strains showed multiple similarities in protein relative abundance and pathways affected by the transformation. Overall, similar pathways were affected by the transformation in the two strains. However, specific proteins showed significantly different abundance.

4.3.1. Nuclear transformation negatively affects cytoskeleton and morphogenesis in both strains

The two strains were transformed using electroporation. Electroporation consists in creating pores in cell membranes using an electric field increasing membrane permeability. This technique has been widely used to genetically engineer prokaryotic and eukaryotic organisms, including microalgae [1,35]. However, applying electrical pulses to an organism can cause many undesired side effects. Depending on electroporation parameters and buffer used, the transformation process can disrupt the cytoskeleton and impair motility [42]. Therefore, the abundance of proteins involved in cell structure development was analysed and compared.

Table 3.1. Relative abundance of cytoskeleton proteins and proteins involved in morphogenesis after comparison of the transgenic lines with their respective wild type strain (n = 5). ND = Not Detected.

Accession	Protein Name	Strain 137c transgenic cell line relative abundance	Strain UVM4 transgenic cell line relative abundance
A8HYU2	V-type proton ATPase subunit C	-1.556	-1.515
A8J1M5	Flagellar outer dynein arm heavy chain beta	-2.321	ND
A8J8K0	Myosin heavy chain class XI	ND	-1.152
A8JAV1	Actin	-3.321	-1.943
P04690	Tubulin beta-1/beta-2 chain	-1.737	ND
Q540H1	Tubulin alpha chain	-2	-1



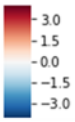
All the cytoskeleton proteins and the proteins involved in morphogenesis detected showed lower abundance in both strain transgenic lines compared to their respective wild types (Table 3.1). The V-type proton ATPase (UniProt accession number: A8HYU2) is potentially involved in a biological process called unidimensional cell growth (protein inferred from homology). Unidimensional growth is the process that leads to morphogenesis. When a cell increases in size and one spatial dimension (namely unidimensional cell growth), it starts to develop its morphological characteristics. A lower abundance of this protein might result in differentially shaped or not fully formed cells. Lower abundance of proteins involved in cell shape and growth might be related to the transformation process. In fact, it has been reported that electroporation severely impairs the tubulin cytoskeleton and motility of cells [42]. Flagellar outer dynein arm heavy chain beta (UniProt accession number: A8J1M5), actin (UniProt accession numbers: A8JAV1), and tubulin alpha chain (UniProt accession number: Q540H1), all reported a lower abundance in both strains after genetic engineering. These proteins are all essential components of the cytoskeleton involved in cell shape, growth, and motility. In addition, strain 137c transgenic samples showed lower abundance of the protein tubulin beta-1/beta-2 chain (UniProt accession number: P04690), whilst strain UVM4 transgenic cell lines showed lower abundance of the protein myosin heavy chain, class XI (UniProt accession number: A8J8K0), both components of the cytoskeleton. A lower abundance of cytoskeleton and morphogenesis proteins might result in not completely formed cells, negatively affecting shape and growth of transformed cells.

4.3.2. Nuclear transgene insertion and expression induce chromatin remodelling

Insertion of recombinant DNA in the endogenous genome is an invasive procedure, therefore after genetic engineering, multiple differences in proteins involved in DNA processing were expected.

Table 3.2. Relative abundance of proteins involved in DNA binding and chromatin remodelling after comparison of the transgenic lines with their respective wild type strain (n = 5). ND = Not Detected.

Accession	Protein Name	Strain 137c transgenic cell line relative abundance	Strain UVM4 transgenic cell line relative abundance
A8HSA7	Histone H3	ND	1.820
A8IA86	Nucleolar protein component of C/D snoRNPs	-2.396	ND
A8J3F0	High mobility group protein	-1.152	-2.836
A8J724	Transcriptional coactivator-like protein	-1.556	-1.556
A8JGJ7	Regulator of chromosome condensation	ND	-3.059
A8JGX5	Protein arginine N-methyltransferase	-1.218	ND



The transcriptional coactivator-like protein (*TSN1*; UniProt accession number: A8J724) showed lower abundance in both strains after genetic engineering. In *A. thaliana*, the TSN1 protein is involved in gene silencing at the RNA level in response to stress [43]. The low abundance of TSN1 in transgenic lines suggests a reduced influence of mechanisms that cause transcripts silencing.

Other proteins found in both proteomic profiles were: nucleolar protein, component of C/D snoRNPs (UniProt accession number: A8IA86), high mobility group protein (UniProt accession number: A8J3F0), and protein arginine N-methyltransferase (UniProt accession number: A8JGX5). In other eukaryotic organisms, these proteins are all involved in histone modification and chromatin structure regulation [44,45,46,47]. Histones and their interaction with DNA play a major role in transcription regulation by condensing or extending the chromatin structure to affect DNA accessibility [29,30]. Distension or condensation of the chromatin structure results in enhancement or diminishment of transcription and is regulated by histone PTMs [28]. After transformation and insertion of transgene in the nucleus, specific enzymes can target the insertion loci and modify histones to induce a condensation of the chromatin structure in those sites, consequently blocking access of the RNA polymerase to the transgene and the DNA loci around it [18,22,27,28].

The nucleolar protein, component of C/D snoRNPs (gene *NOP1*) and the protein arginine N-methyltransferase (gene *PRMT2*), are both involved in histone modification. More specifically, in yeasts, the activity of the NOP1 protein is directly related to enhanced transcription [44,45]. The gene

PRMT2 in *Homo sapiens* encodes for a protein related to both positive and negative regulation of transcription [47]. The high mobility group protein does not directly modify histones, however it is involved in bending DNA to ease the insertion of histones during formation of the nucleosome (the octamer constituted of histones H2A, H2B, H3, and H4). A lower abundance of this protein enhance transcription by making DNA more accessible.

Taken together, the differential production of these three proteins in the transgenic lines of both strains implies possible DNA restructuring as response to the transformation (Table 3.2).

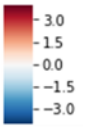
Strain UVM4 transgenic lines showed unique differentially produced proteins (i.e. not detected in strain 137c transgenic lines) involved in chromatin regulation mechanisms. A lower abundance of the protein called Regulator of chromosome condensation (UniProt accession number: A8JGJ7) might be related to a distended DNA structure and possibly to enhanced transcription. The histone H3 (UniProt accession number: A8HSA7) was found in higher abundance, which was surprising considering the lower abundance of multiple proteins involved in histone attachment and chromatin condensation. A higher abundance of core histones was already reported for strain UVM4 wild type when compared to 137c wild type (Chapter 2). In this study, the other three core histones were not detected. These proteins involved in chromatin remodelling showed altered abundance only in strain UVM4 transgenic cell lines, suggesting that they might be responsible for the position effects affecting this strain.

4.3.3. Genetic engineering enhances production of ribosomal proteins but decreases translation initiation and elongation

Considering that the transgene is driven by a constitutive promoter and thus constantly expressed, a higher abundance of proteins involved in translation was expected.

Table 3.3. Relative abundance of ribosomal proteins and proteins involved in translation initiation and elongation. Protein relative abundance was determined by comparing the transgenic lines with their respective wild type strain (n = 5). ND = Not Detected.

Accession	Protein Name	Strain 137c transgenic cell line relative abundance	Strain UVM4 transgenic cell line relative abundance
A0A218N8I0	30S ribosomal protein S3 chloroplastic	1.189	ND
A8HMG7	Ribosomal protein L26	ND	4.062
A8HN50	60S ribosomal protein L18a	2.884	1.899
A8HP90	Ribosomal protein L6	4.148	4.655
A8HS59	Ribosomal protein L17	3.376	3.396
A8HVK4	Ribosomal protein S27a	3.016	3.414
A8HVQ1	40S ribosomal protein S8	3.619	1.339
A8I2T0	Ribosomal protein L27a	3.332	2.568
A8I8Z4	Ribosomal protein	ND	-1.644
A8IMP6	40S ribosomal protein S4	ND	1.974
A8IQE3	Ribosomal protein L14	1.546	2.498
A8IUV7	60S ribosomal protein L13	2.433	2.933
A8IVE2	Ribosomal protein L7	1.604	2.195
A8IVK1	Ribosomal protein L8	1.971	2.208
A8J0I0	Ribosomal protein L4	1.536	1.982
A8J1G8	40S ribosomal protein S6	3.293	2.807
A8J239	Ribosomal protein L23a	ND	1.930
A8J493	Ribosomal protein S15	3.645	1.907
A8J567	Ribosomal protein L7a	ND	1.876
A8J8M9	Ribosomal protein S20	-1.322	-1.836
A8J951	Ribosomal protein L21	1.632	ND
A8JE07	Ribosomal protein S15a	1.546	ND
A8JF05	Ribosomal protein L28	ND	1.599
A8JHC3	Ribosomal protein S11	3.123	1.063
A8JHU2	60S ribosomal protein L36	3.417	4.229
E3SC57	60S ribosomal protein L3 (Fragment)	2.098	2.361
P46295	40S ribosomal protein S14	ND	1.063
P48267	30S ribosomal protein S7 chloroplastic	2.815	ND
P49202	40S ribosomal protein S18	1.561	ND
A8HW92	Eukaryotic translation initiation factor (Fragment)	ND	-1.251
A8HX38	Eukaryotic translation elongation factor 1 alpha 1	-1.152	-1.184
A8IA39	Elongation factor G chloroplastic	ND	-1.120
A8IP17	Eukaryotic initiation factor 4A-like protein	ND	-2.059
A8ISZ1	Elongation factor EF-3	-2.322	-1.836
A8J280	Elongation factor-like protein (Fragment)	-1.670	ND
A8J9A9	Eukaryotic translation initiation factor 3 subunit F	-1.599	ND
A8JBN7	Eukaryotic initiation factor	-1.889	ND
A8JHM2	Eukaryotic initiation factor (Fragment)	-1.786	-1.644
A8JHX9	Elongation factor 2	-1.514	-1.786
P17746	Elongation factor Tu chloroplastic	-3.322	-1.396



The transgenic cell lines for both strains showed a similar trend. Almost all the ribosomal proteins were more abundant in transgenic lines than wild types, whilst proteins involved in initiation or elongation of translation were less abundant. Ribosomal proteins are bound to the rRNA, maintain the structural integrity of ribosomes, and are fundamental for translation processes, thus their higher abundance is directly correlated to higher translation rates [48], probably generated by the continuous expression of transgene. Strain UVM4 wild type (compared to strain 137c wild type, Chapter 2) also showed higher abundance of ribosomal proteins. As described in Chapter 2, strain UVM4 was subjected to transgene

insertion prior to UV mutagenesis, therefore the higher abundance of ribosomal proteins reported in strain UVM4 wild type (Chapter 2) is coherent with the results shown here in Table 3.3.

Interestingly, the proteomic profiles in Table 3.3 showed one common less abundant ribosomal protein, Ribosomal protein S20 (UniProt accession number: A8J8M9). This protein seems to be an ordinary ribosomal protein, similar to all the others shown in Table 3.3, therefore the reasons for its lower abundance remain uncertain. Strain UVM4 transgenic cell lines showed another less abundant ribosomal protein (UniProt accession number: A8I8Z4), not detected in strain 137c. It is possible that the gene *RPS20* (which encodes for the ribosomal protein S20) and the gene *PRPL1* (which encodes for the ribosomal protein A8I8Z4) are both co-transcribed with another down-regulated gene, leading to a lower abundance of these two ribosomal proteins.

Table 3.3 also shows proteins involved in initiation or elongation of translation. These ubiquitous proteins can be found in the nucleus, the cytosol, or the chloroplast, and their role is to initiate the translation of mRNA and to elongate the polypeptide chain during its synthesis in the ribosome. Interestingly, these proteins all showed lower abundance in both strain transgenic lines. This behaviour was not reported in strain UVM4 wild type (compared to strain 137c wild type, Chapter 2), and is possibly related to the expression of nuclear transgenes. The two proteomic profiles showed many common proteins (five out of eight for both strains): Eukaryotic translation elongation factor 1 alpha (UniProt accession number: A8HX38), Elongation factor EF-3 (UniProt accession number: A8ISZ1), Eukaryotic initiation factor (Fragment) (UniProt accession number: A8JHM2), Elongation factor 2 (UniProt accession number: A8JHX9), and Elongation factor Tu, chloroplastic (UniProt accession number: P17746). This reported general lower abundance might be a stress response of the cell to the forced recombinant protein production. The initiation and elongation machinery might slow down to avoid consuming cell resources to produce a protein not useful to the organism. It has been reported in numerous eukaryotes that endoplasmic reticulum (ER) stress can be caused by misfolded proteins accumulating in the ER during protein assembly [49,50], including *C. reinhardtii* [51]. The organism response to ER stress is called unfolded protein response (UPR), a signalling cascade capable of

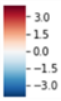
restoring ER homeostasis. This stress response mechanism has multiple pathways, including phosphorylation of translation initiation factor 2 (eIF2α) [49,50,51]. This particular protein was not detected in this study, however the lower abundance of other multiple initiation and elongation translation factors might be related to transgene expression stress and needs to be investigated further.

4.3.4. Chaperones and transporters show general lower abundance, particularly in strain UVM4

After translation, the newly formed proteins are subjected to a quality check and subsequently transported to their final destination. As a consequence of transgene expression, this machinery might be subjected to stress, due to the constant production of an unexpected and contextually function-less protein. Therefore, relative abundance of proteins involved in quality control and transport after genetic engineering was analysed, to detect possible altered pathways. The two strains reported a similar lower abundance of these proteins, and interestingly, strain UVM4 transgenic cell lines showed a higher number of less abundant proteins, including specific classes of proteins absent in strain 137c (Table 3.4).

Table 3.4. Relative abundance of proteins involved in protein folding and transport. Protein relative abundance was determined by comparing the transgenic lines with their respective wild type strain (n = 5). ND = Not Detected.

Accession	Protein Name	Strain 137c transgenic cell line relative abundance	Strain UVM4 transgenic cell line relative abundance
A0A220IUF1	Heat shock protein 90B	-1.322	-1.786
A8HQ74	T-complex protein eta subunit	ND	-1.286
A8HRR9	Coatomer subunit alpha	ND	-1.474
A8I4S9	Clathrin heavy chain	ND	-1.689
A8IF08	T-complex protein theta subunit	ND	-1.599
A8IFK0	Plasma membrane ATPase	ND	-1.184
A8IL29	Small ARF-related GTPase	-1.358	ND
A8IM71	Coatomer subunit gamma	-2	ND
A8IRV0	HSP70-HSP90 organizing protein	-1.251	-1.889
A8IRX5	GTP-binding nuclear protein	-1.358	-1.251
A8IUL5	Importin beta	ND	-1.152
A8IZU0	Heat shock protein 70C	ND	-1.434
A8J146	Guanosine nucleotide diphosphate dissociation inhibitor	ND	-1.434
A8J524	T-complex protein 1 beta subunit	ND	-1.514
A8JE91	Chaperonin 60B1	-1.515	-1.434
A8JGS8	Coatomer subunit beta	-1.286	-1.889
A8JIB7	Chaperonin 60A	ND	-1.089
I2FKQ9	Mitochondrial chaperonin 60	-2	-1.836
O48949	Protein disulfide-isomerase	-1.358	-1.251
P42380	ATP-dependent Clp protease proteolytic subunit	ND	-1.120
Q39571	GTP-binding protein YPTC1	ND	-1.689
Q39603	Heat shock protein 70B	ND	-1.217
Q66T67	Heat shock protein 90C	ND	-1.251
Q9STD3	Calreticulin	-1.029	ND



Heat shock proteins and chaperonins (present in both proteomic profiles), together with calreticulin (present in strain 137c transgenic proteomic profile), are two classes of proteins that bind unfolded proteins (*de novo* or misfolded) and promote their correct folding. Their lower abundance might be related to the stress the transformant cell lines are subjected to, considering the constant production of a recombinant exogenous protein (Table 3.4). In fact, an inserted transgene including a constitutive promoter is constantly transcribed by the cell machinery. Subsequently, the transcript is modified/matured before passing through the nuclear envelope to the cytosol or rough ER. It is then recognized by the ribosomes and translated to protein. At this stage, quality control and transport mechanisms are activated. An overproduction of a recombinant protein might affect negatively the machinery, overwhelmed by the extensive and abnormal amount of work. A possible response of the cell could be to shut down quality control and transport machinery to save energy, subsequently used to activate proteolysis pathways such as the ubiquitin–proteasome pathway (ERAD) and the lysosomal pathway (autophagy) [49].

Another common trait between the two strains is the presence of multiple proteins involved in vesicle-mediated intracellular protein transport (Table 3.4). Small ARF-related GTPase (UniProt accession

number: A8IL29) and Coatomer subunit gamma (UniProt accession number: A8IM71) (both present in strain 137c transgenic cell lines), Coatomer subunit beta (UniProt accession number: A8JGS8) (present in both strains), GTP-binding protein YPTC1 (UniProt accession number: Q39571), Coatomer subunit alpha (UniProt accession number: A8HRR9), Clathrin heavy chain (UniProt accession number: A8I4S9), and Guanosine nucleotide diphosphate dissociation inhibitor (UniProt accession number: A8J146) (all present in strain UVM4 transgenic lines), are all proteins involved in vesicular transport of proteins between organelles. Vesicle-mediated intracellular transport of proteins can be anterograde (ER-to-Golgi transport) or retrograde (Golgi-to-ER). The proteins listed here are involved in both. Moreover, the proteomic profiles of both strains show another class of transport proteins, specifically proteins involved in nucleocytoplasmic transport. GTP-binding nuclear protein (UniProt accession number: A8IRX5) (present in both strains) and importin beta (UniProt accession number: A8IUL5) (present only in strain UVM4) are both involved in transport of proteins into the nucleus and of RNA outside the nucleus into the cytosol. As well as the vesicle-mediated transporter, these proteins are less abundant in both strains. A possible explanation might be a diminished interaction between organelles, to save precious energy and to limit the expansion of unknown and undesired proteins throughout the cell. The recombinant protein accumulated in the cytosol might subsequently be directed to the proteolytic route [52].

Strain UVM4 transgenic cell lines also show some unique proteins (i.e. absent in 137c transgenic lines). Three proteins belonging to the T-complex were found in lower abundance: subunit beta (UniProt accession number: A8J524), subunit eta (UniProt accession number: A8HQ74), and subunit theta (UniProt accession number: A8IF08). The T-complex is a cytosolic chaperon involved in positive regulation of transport and protein folding. None of the T-complex subunits was found in the proteomic analysis of strain UVM4 wild type (compared to strain 137c wild type, Chapter 2), and none was detected in differential abundance in strain 137c transgenic lines (Table 3.4). The lower abundance of the T-complex chaperon in strain UVM4 transgenic cell lines might be a unique response to insertion and expression of transgenes. However, its role is still unclear and needs further analysis.

5. CONCLUSIONS

This study proposes an innovative screening method of *Chlamydomonas reinhardtii* transformants using fluorescence-activated cell sorting (FACS), to overcome the low transformation rates that affect the microalga.

Low transformation rates have been hindering *C. reinhardtii* performances as a cell biofactory since its first biotechnological use [16,17]. Moreover, the generation of strain UVM4 did not fully overcome the issue. In fact, here, before screening transformants using FACS, I reported a low percentage of positive cells after transformation for both strains 137c and UVM4 (with strain UVM4 showing success rate lower than 10%). Separation of single transgenic cells by FACS, and subsequent growth of the cultures as single transgenic cell lines, enhanced the percentage of positive cells showing fluorescence in the analysed cultures. However, even after FACS, none of the transgenic cell lines of both strains showed a total percentage of positive cells higher than 90%. These results suggest that transgene silencing mechanisms are extensively active in *C. reinhardtii* strain 137c and possibly even more in strain UVM4.

In addition, flow cytometry analysis of normalised fluorescence over time for all transgenic cell lines of both strains showed that position effects play a major role in transgene insertion and expression in strain UVM4. Considering the high variation between transgenic cell lines of strain UVM4, a screening of a high number of transformants seems necessary to find a highly expressing candidate. Strain 137c seems somehow to be less affected by position effects. In fact, the normalised fluorescence over time showed less variation between transgenic cell lines of strain 137c. Moreover, while fluorescence might not always reflect accurately recombinant protein final yield, it was interesting to note in this study that strain 137c showed higher values of normalised fluorescence than strain UVM4, which could be related to position effects.

This study also provides the first proteomic analysis and comparison of *Chlamydomonas reinhardtii* strains 137c and UVM4 after genetic engineering, to investigate the effect of transgene insertion and

expression on the organisms, to possibly reveal pathways activated or deactivated by the microalga to counteract the transformation process.

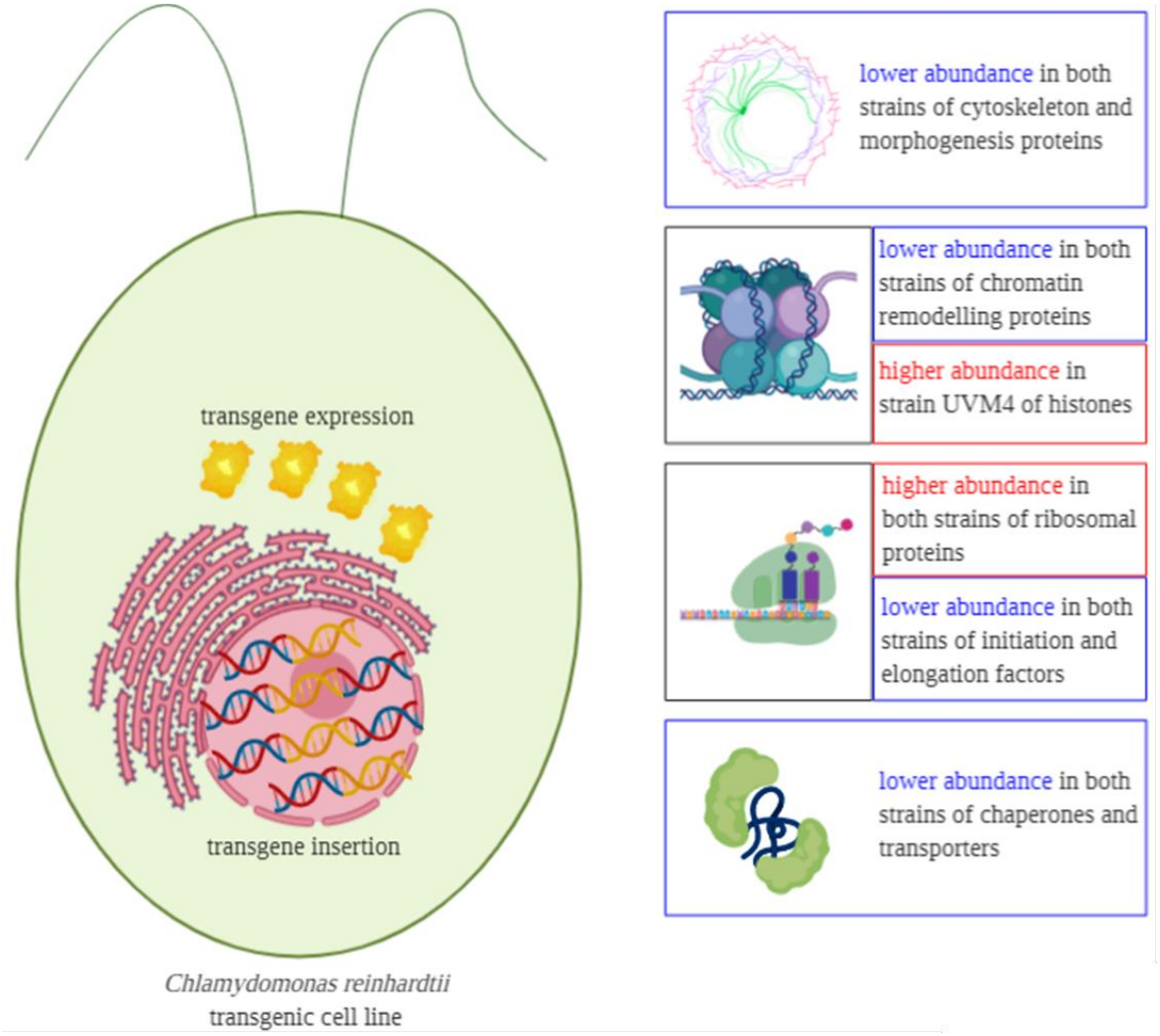


Figure 3.5. Schematic of altered pathways in *C. reinhardtii* strains 137c and UVM4 after transgene insertion and expression.

Transgene insertion and expression had a strong impact on total protein production and altered multiple pathways. Surprisingly, both strains reported lower abundance of proteins directly related to recombinant protein production, such as proteins involved in nucleic acid processing, translation initiation and elongation, protein quality control, and protein transport (Figure 3.5). The only proteins that reported a higher abundance in both strains were ribosomal proteins. I hypothesise that the constant

expression of a transgene causes a strong stress to the organism, which activates a stress response and slows down multiple pathways involved in protein production.

Few differences between the two strains were also reported (Figure 3.5). Strain UVM4 showed an increase in histone abundance and lower abundance of one specific protein involved in chromatin condensation. It is possible that these differences are responsible for increased position effects. Moreover, strain UVM4 shows a unique class of transporters, the T-complex chaperon, differentially abundant after genetic engineering. The role of this complex is still relatively unknown, and so is its involvement in transgene expression.

Future studies should focus their attention on chromatin structure and remodelling in *C. reinhardtii*, to investigate their specific role in transgene insertion and expression and to improve recombinant protein production in the microalga. Moreover, the role of the T-complex chaperon should be investigated further, to fully comprehend its involvement in protein production and its different abundance in strain UVM4. Furthermore, lower abundance of translation initiation and elongation factors are likely related to low recombinant protein yields in *C. reinhardtii* (and possibly in other microalgal species) (compared to more established cell biofactories). Increasing the abundance of proteins involved in these pathways, using knock-in approaches, might enhance recombinant protein production in microalgae. Finally, given that strain UVM4 has been reported to produce higher yields of secreted recombinant proteins [38,39], future studies comparing transgenic cell lines between *C. reinhardtii* strain UMV4 and strain 137c should also investigate secreted recombinant protein.

6. SUPPLEMENTARY FIGURES

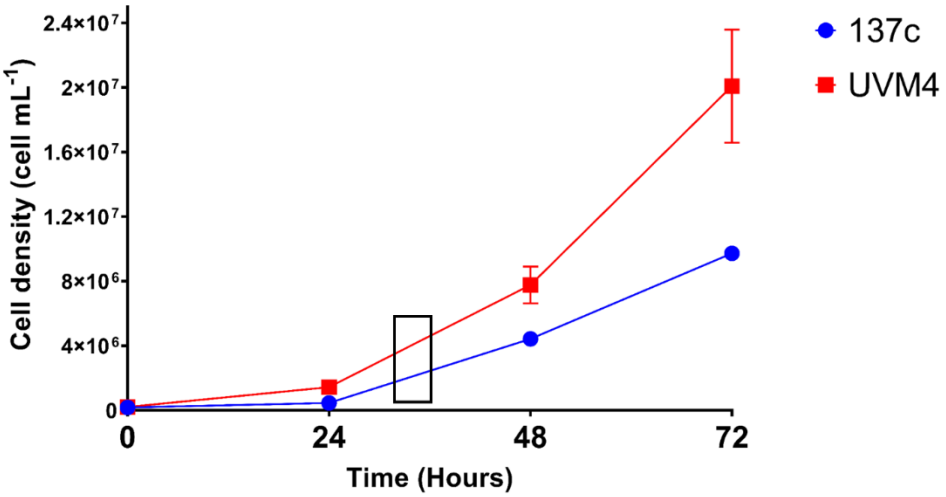


Figure S3.1. Timing of cell sampling for transformation. Cells were harvested at 2 – 4 x10^6 cells/mL (early exponential phase, 24 – 48 hours).

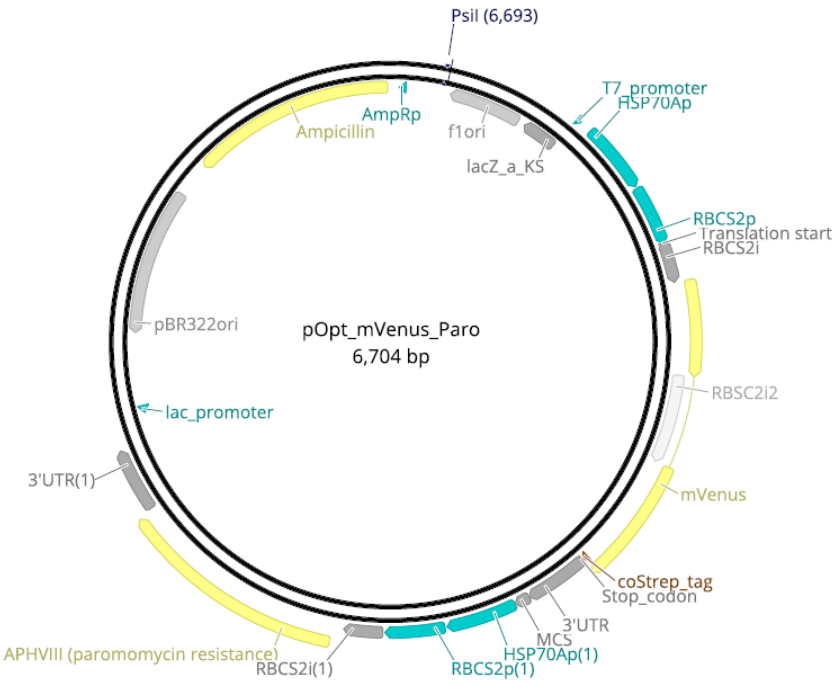


Figure S3.2 Plasmid used for nuclear transformation in *C. reinhardtii* strain UMV4 and strain 137c. The plasmid pOptimized (pOpt) pOpt_mVenus_Paro (NCBI: KM061060) containing the mVenus and the

paromomycin selection marker expression cassettes [32] was purchased from the Chlamydomonas Resource Center (https://www.chlamycollection.org/). Expression cassettes and restriction enzyme (PsiI) site are shown.

7. REFERENCES

1. Shimogawara, K.; Fujiwara, S.; Grossman, A.; Usuda, H. High-efficiency transformation of *Chlamydomonas reinhardtii* by electroporation. *Genetics* **1998**, *148*, 1821-1828.

2. Sasso, S.; Pohnert, G.; Lohr, M.; Mittag, M.; Hertweck, C. Microalgae in the postgenomic era: a blooming reservoir for new natural products. *FEMS Microbiol Rev* **2012**, *36*, 761-785, doi:10.1111/j.1574-6976.2011.00304.x.

3. Lauersen, K.J.; Wichmann, J.; Baier, T.; Kampranis, S.C.; Pateraki, I.; Moller, B.L.; Kruse, O. Phototrophic production of heterologous diterpenoids and a hydroxy-functionalized derivative from *Chlamydomonas reinhardtii*. *Metab Eng* **2018**, *49*, 116-127, doi:10.1016/j.ymben.2018.07.005.

4. Commault, A.S.; Fabris, M.; Kuzhiumparambil, U.; Adriaans, J.; Pernice, M.; Ralph, P.J. Methyl jasmonate treatment affects the regulation of the 2-C-methyl-D-erythritol 4-phosphate pathway and early steps of the triterpenoid biosynthesis in *Chlamydomonas reinhardtii*. *Algal Research* **2019**, *39*, doi:10.1016/j.algal.2019.101462.

5. Chavez, M.N.; Schenck, T.L.; Hopfner, U.; Centeno-Cerdas, C.; Somlai-Schweiger, I.; Schwarz, C.; Machens, H.G.; Heikenwalder, M.; Bono, M.R.; Allende, M.L., et al. Towards autotrophic tissue engineering: Photosynthetic gene therapy for regeneration. *Biomaterials* **2016**, *75*, 25-36, doi:10.1016/j.biomaterials.2015.10.014.

6. Baier, T.; Kros, D.; Feiner, R.C.; Lauersen, K.J.; Muller, K.M.; Kruse, O. Engineered Fusion Proteins for Efficient Protein Secretion and Purification of a Human Growth Factor from the Green Microalga *Chlamydomonas reinhardtii*. *ACS Synth Biol* **2018**, 10.1021/acssynbio.8b00226, doi:10.1021/acssynbio.8b00226.

7. Merlin, M.; Gecchele, E.; Capaldi, S.; Pezzotti, M.; Avesani, L. Comparative evaluation of recombinant protein production in different biofactories: the green perspective. *Biomed Res Int* **2014**, *2014*, 136419, doi:10.1155/2014/136419.

8. Specht, E.; Miyake-Stoner, S.; Mayfield, S. Micro-algae come of age as a platform for recombinant protein production. *Biotechnol Lett* **2010**, *32*, 1373-1383, doi:10.1007/s10529-010-0326-5.

9. Rasala, B.A.; Mayfield, S.P. Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. *Photosynth Res* **2015**, *123*, 227-239, doi:10.1007/s11120-014-9994-7.

10. Merchant, S.S.; Prochnik, S.E.; Vallon, O.; Harris, E.H.; Karpowicz, S.J.; Witman, G.B.; Terry, A.; Salamov, A.; Fritz-Laylin, L.K.; Marechal-Drouard, L., et al. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **2007**, *318*, 245-250, doi:10.1126/science.1143609.

11. Boynton, J.; Gillham, N.; Harris, E.; Hosler, J.; Johnson, A.; Jones, A.; Randolph-Anderson, B.; Robertson, D.; Klein, T.; Shark, K., et al. Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science* **1988**, *240*, 1534-1538, doi:10.1126/science.2897716.

12. Kindle, K.L. High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci U S A* **1990**, *87*, 1228-1232, doi:10.1073/pnas.87.3.1228.

13. Mathieu-Rivet, E.; Kiefer-Meyer, M.C.; Vanier, G.; Ovide, C.; Burel, C.; Lerouge, P.; Bardor, M. Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front Plant Sci* **2014**, *5*, 359, doi:10.3389/fpls.2014.00359.

14. Fahad, S.; Khan, F.A.; Pandupuspitasari, N.S.; Ahmed, M.M.; Liao, Y.C.; Waheed, M.T.; Sameeullah, M.; Darkhshan; Hussain, S.; Saud, S., et al. Recent developments in therapeutic protein expression technologies in plants. *Biotechnol Lett* **2015**, *37*, 265-279, doi:10.1007/s10529-014-1699-7.

15. Shin, S.E.; Lim, J.M.; Koh, H.G.; Kim, E.K.; Kang, N.K.; Jeon, S.; Kwon, S.; Shin, W.S.; Lee, B.; Hwangbo, K., et al. CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Sci Rep* **2016**, *6*, 27810, doi:10.1038/srep27810.

16. Lumbreras, V.; Stevens, D.R.; Purton, S. Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. *The Plant Journal* **1998**, *14*, 441-447, doi:10.1046/j.1365-313X.1998.00145.x.

17. Schroda, M.; Blöcker, D.; Beck, C.F. The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *The Plant Journal* **2000**, *21*, 121-131, doi:10.1046/j.1365-313x.2000.00652.x.

18. Neupert, J.; Karcher, D.; Bock, R. Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J* **2009**, *57*, 1140-1150, doi:10.1111/j.1365-313X.2008.03746.x.

19. Scaife, M.A.; Nguyen, G.T.; Rico, J.; Lambert, D.; Helliwell, K.E.; Smith, A.G. Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host. *Plant J* **2015**, *82*, 532-546, doi:10.1111/tpj.12781.

20. Schroda, M. Good News for Nuclear Transgene Expression in *Chlamydomonas*. *Cells* **2019**, *8*, doi:10.3390/cells8121534.

21. Lodha, M.; Schulz-Raffelt, M.; Schroda, M. A new assay for promoter analysis in *Chlamydomonas* reveals roles for heat shock elements and the TATA box in HSP70A promoter-mediated activation of transgene expression. *Eukaryot Cell* **2008**, *7*, 172-176, doi:10.1128/EC.00055-07.

22. Strenkert, D.; Schmollinger, S.; Schroda, M. Heat shock factor 1 counteracts epigenetic silencing of nuclear transgenes in *Chlamydomonas reinhardtii*. *Nucleic Acids Res* **2013**, *41*, 5273-5289, doi:10.1093/nar/gkt224.

23. Li, B.; Carey, M.; Workman, J.L. The role of chromatin during transcription. *Cell* **2007**, *128*, 707-719, doi:10.1016/j.cell.2007.01.015.

24. Chávez, S.; Beato, M. Nucleosome-mediated synergism between transcription factors on the mouse mammary tumor virus promoter. *Proceedings of the National Academy of Sciences* **1997**, *94*, 2885-2890, doi:10.1073/pnas.94.7.2885.

25. Prado, F.; Koop, R.; Beato, M. Accurate chromatin organization of the mouse mammary tumor virus promoter determines the nature of the synergism between transcription factors. *J Biol Chem* **2002**, *277*, 4911-4917, doi:10.1074/jbc.M110094200.

26. Cerutti, H.; Johnson, A.M.; Gillham, N.W.; Boynton, J.E. Epigenetic silencing of a foreign gene in nuclear transformants of *Chlamydomonas*. *The Plant Cell* **1997**, *9*, 925-945.

27. Casas-Mollano, J.A.; Jeong, B.-r.; Xu, J.; Moriyama, H.; Cerutti, H. The MUT9p kinase phosphorylates histone H3 threonine 3 and is necessary for heritable epigenetic silencing in *Chlamydomonas*. *Proceedings of the National Academy of Sciences* **2008**, *105*, 6486-6491.

28. Moraes, I.; Yuan, Z.F.; Liu, S.; Souza, G.M.; Garcia, B.A.; Casas-Mollano, J.A. Analysis of Histones H3 and H4 Reveals Novel and Conserved Post-Translational Modifications in Sugarcane. *PLoS One* **2015**, *10*, e0134586, doi:10.1371/journal.pone.0134586.

29. Turner, B.M. Cellular Memory and the Histone Code. *Cell* **2002**, *111*, 285-291, doi.

30. Berger, S.L. Histone modifications in transcriptional regulation. *Current Opinion in Genetics & Development* **2002**, *12*, 142-148, doi.

31. Mutskov, V.; Felsenfeld, G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J* **2004**, *23*, 138-149, doi:10.1038/sj.emboj.7600013.

32. Lauersen, K.J.; Kruse, O.; Mussgnug, J.H. Targeted expression of nuclear transgenes in *Chlamydomonas reinhardtii* with a versatile, modular vector toolkit. *Appl Microbiol Biotechnol* **2015**, *99*, 3491-3503, doi:10.1007/s00253-014-6354-7.

33. Lauersen, K.J.; Baier, T.; Wichmann, J.; Wordenweber, R.; Mussgnug, J.H.; Hubner, W.; Huser, T.; Kruse, O. Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga *Chlamydomonas reinhardtii*. *Metab Eng* **2016**, *38*, 331-343, doi:10.1016/j.ymben.2016.07.013.

34. Baier, T.; Wichmann, J.; Kruse, O.; Lauersen, K.J. Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga *Chlamydomonas reinhardtii*. *Nucleic Acids Res* **2018**, *46*, 6909-6919, doi:10.1093/nar/gky532.

35. Rasala, B.A.; Lee, P.A.; Shen, Z.; Briggs, S.P.; Mendez, M.; Mayfield, S.P. Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS One* **2012**, *7*, e43349, doi:10.1371/journal.pone.0043349.

36. Molino, J.V.D.; de Carvalho, J.C.M.; Mayfield, S.P. Comparison of secretory signal peptides for heterologous protein expression in microalgae: Expanding the secretion portfolio for *Chlamydomonas reinhardtii*. *PLoS One* **2018**, *13*, e0192433, doi:10.1371/journal.pone.0192433.

37. Kurniasih, S.D.; Yamasaki, T.; Kong, F.; Okada, S.; Widyaningrum, D.; Ohama, T. UV-mediated *Chlamydomonas* mutants with enhanced nuclear transgene expression by disruption of DNA methylation-dependent and independent silencing systems. *Plant Mol Biol* **2016**, *92*, 629-641, doi:10.1007/s11103-016-0529-9.

38. Lauersen, K.J.; Huber, I.; Wichmann, J.; Baier, T.; Leiter, A.; Gaukel, V.; Kartushin, V.; Rattenholl, A.; Steinweg, C.; von Riesen, L., et al. Investigating the dynamics of recombinant protein secretion from a microalgal host. *J Biotechnol* **2015**, *215*, 62-71, doi:10.1016/j.jbiotec.2015.05.001.

39. Ramos-Martinez, E.M.; Fimognari, L.; Sakuragi, Y. High-yield secretion of recombinant proteins from the microalga *Chlamydomonas reinhardtii*. *Plant Biotechnol J* **2017**, *15*, 1214-1224, doi:10.1111/pbi.12710.

40. Gorman, D.S.; Levine, R.P. Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardi*. *Proc Natl Acad Sci U S A* **1965**, *54*, 1665-1669, doi:10.1073/pnas.54.6.1665.

41. Kong, F.; Yamasaki, T.; Kurniasih, S.D.; Hou, L.; Li, X.; Ivanova, N.; Okada, S.; Ohama, T. Robust expression of heterologous genes by selection marker fusion system in improved *Chlamydomonas* strains. *J Biosci Bioeng* **2015**, *120*, 239-245, doi:10.1016/j.jbiosc.2015.01.005.

42. Harkin, D.G.; Hay, E.D. Effects of electroporation on the tubulin cytoskeleton and directed migration of corneal fibroblasts cultured within collagen matrices. *Cell Motility* **1996**, *35*, 345-357, doi:10.1002/(sici)1097-0169(1996)35:4<345::Aid-cm6>3.0.Co;2-5.

43. Frei dit Frey, N.; Muller, P.; Jammes, F.; Kizis, D.; Leung, J.; Perrot-Rechenmann, C.; Bianchi, M.W. The RNA binding protein Tudor-SN is essential for stress tolerance and stabilizes levels of stress-responsive mRNAs encoding secreted proteins in *Arabidopsis*. *Plant Cell* **2010**, *22*, 1575-1591, doi:10.1105/tpc.109.070680.

44. Tessarz, P.; Santos-Rosa, H.; Robson, S.C.; Sylvestersen, K.B.; Nelson, C.J.; Nielsen, M.L.; Kouzarides, T. Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* **2014**, *505*, 564-568, doi:10.1038/nature12819.

45. Leonhardt, H.; Hake, S.B. Histone glutamine methylation afFACTing rDNA transcription. *Cell Res* **2014**, *24*, 261-262, doi:10.1038/cr.2014.22.

46. Karnavas, T.; Pintonello, L.; Agresti, A.; Bianchi, M.E. Histone content increases in differentiating embryonic stem cells. *Front Physiol* **2014**, *5*, 330, doi:10.3389/fphys.2014.00330.

47. Lakowski, T.M.; Frankel, A. Kinetic analysis of human protein arginine N-methyltransferase 2: formation of monomethyl- and asymmetric dimethyl-arginine residues on histone H4. *Biochem J* **2009**, *421*, 253-261, doi:10.1042/BJ20090268.

48. Zhou, X.; Liao, W.J.; Liao, J.M.; Liao, P.; Lu, H. Ribosomal proteins: functions beyond the ribosome. *J Mol Cell Biol* **2015**, *7*, 92-104, doi:10.1093/jmcb/mjv014.

49. Bravo, R.; Parra, V.; Gatica, D.; Rodriguez, A.E.; Torrealba, N.; Paredes, F.; Wang, Z.V.; Zorzano, A.; Hill, J.A.; Jaimovich, E., et al. Endoplasmic reticulum and the unfolded protein response: dynamics and metabolic integration. *Int Rev Cell Mol Biol* **2013**, *301*, 215-290, doi:10.1016/B978-0-12-407704-1.00005-1.

50. Galluzzi, L.; Yamazaki, T.; Kroemer, G. Linking cellular stress responses to systemic homeostasis. *Nat Rev Mol Cell Biol* **2018**, *19*, 731-745, doi:10.1038/s41580-018-0068-0.

51. Yamaoka, Y.; Choi, B.Y.; Kim, H.; Shin, S.; Kim, Y.; Jang, S.; Song, W.Y.; Cho, C.H.; Yoon, H.S.; Kohno, K., et al. Identification and functional study of the endoplasmic reticulum stress sensor IRE1 in *Chlamydomonas reinhardtii*. *Plant J* **2018**, *94*, 91-104, doi:10.1111/tpj.13844.

52. Robert, F.; Bierau, H.; Rossi, M.; Agugiaro, D.; Soranzo, T.; Broly, H.; Mitchell-Logean, C. Degradation of an Fc-fusion recombinant protein by host cell proteases: Identification of a CHO cathepsin D protease. *Biotechnology and Bioengineering* **2009**, *104*, 1132-1141, doi:10.1002/bit.22494.

CHAPTER 4: Comparative glycomic analysis reveals differences in protein glycosylation between *Chlamydomonas reinhardtii* strains 137c and UVM4

Lorenzo Barolo[1], Matthew P. Padula[2], Raffaela M. Abbriano[1], Audrey S. Commault[1], Unnikrishnan Kuzhiumparambil[1], and Mathieu Pernice[1]

AFFILIATIONS

[1] University of Technology Sydney, Climate Change Cluster, Broadway Campus, Ultimo NSW 2007, Sydney, Australia

[2]School of Life Sciences and Proteomics Core Facility, Faculty of Science, University of Technology Sydney, Ultimo NSW 2007, Sydney, Australia

CONTRIBUTIONS: LB, UK, RMA, ASC, and MP designed the experiment; LB collected and processed the samples; LB performed the protein gels and analysed the results; LB and MPP performed the protein mass spectrometry analysis; UK performed the glycomic analysis; LB analysed the mass spectrometry data for the protein analysis; UK performed data analysis for the glycosylation patterns; LB wrote the manuscript with contributions from all authors.

This chapter has been formatted for submission to the journal: Glycobiology.

1. ABSTRACT

Validation of a cell biofactory for recombinant biopharmaceutical production relies on two major aspects: transgene expression levels (product quantity) and protein post-translational modifications (product quality). The green microalga *Chlamydomonas reinhardtii* has recently arisen as a promising cell biofactory, especially after the generation of the UV mutated strain UVM4 capable of high transgene expression. However, the post-translational modification machinery of this strain has not been characterised yet. Unveiling the configuration of UVM4 glycosylation patterns is critical to assess the quality of recombinant proteins produced in this host. In particular, the presence of non-human and possibly-immunogenic glycans need to be investigated if strain UVM4 is to be considered as a future cell biofactory. In this study, the first glycosylation analysis of the extracellular proteome of the broadly-used strain UVM4 was performed, to unveil its glycosylation status and detect presence/absence of immunogenic glycoforms. These results show that strain UVM4 and strain 137c exhibit very different extracellular glycoprotein profiles. More importantly, I report the presence of immunogenic glycosylation patterns in strain UVM4. I also detected the first "complex" glycoform ever reported in *C. reinhardtii*, a glycan that has been considered completely absent in this species so far. These results force a revaluation of the general glycosylation status and pathways of *C. reinhardtii*, and open the way for further analyses on the immunogenicity of recombinant biopharmaceuticals produced in *C. reinhardtii* strain UVM4.

2. INTRODUCTION

Glycosylation involves the attachment of a polysaccharide chain to an amino acid and is one of the most relevant post-translational modifications (PTMs) found on eukaryotic proteins. The two most frequent types of glycosylation are *N*-linked glycosylation, characterized by the covalent linkage between a glycan and the amidic group of an asparagine (Asn) residue, and *O*-linked glycosylation, presenting a covalent bond between the glycan and the hydroxyl component of a serine (Ser) or a threonine (Thr) residue. Glycosylation is found on more than 50% of native human proteins [1] and it strongly regulates fundamental biological processes within the cell, such as cell adhesion, biological activity, folding, solubility, molecular trafficking and clearance, and proteases resistance [2,3]. Glycosylation also plays a major role in yield, stability, efficacy, and pharmacokinetics of secreted recombinant biopharmaceuticals [4,5]. In eukaryotic cell-based production systems, glycosylation is introduced by directing the recombinant protein to the secretion route with specific signal peptides [6]. The journey through the secretory pathway (endoplasmic reticulum and Golgi apparatus) is necessary to initiate protein glycosylation. Glycosylation is found on more than 40% of approved recombinant biopharmaceuticals [7,8] and is generally responsible for enhanced protein quality and activity.

Glycosylated biopharmaceuticals produced in non-human cell biofactories have the potential to trigger an immune response in patients. The cellular mechanisms that determine protein glycosylation patterns are species-specific and vary among different eukaryotic organisms [9]. Therefore, during recombinant protein production, non-human cell biofactories can attach glycan residues (monosaccharides) that would be absent on the human endogenous protein, potentially resulting in an immunogenic biopharmaceutical (as described in Chapter 1). An immunogenic reaction due to incorrect glycosylation can result in accelerated clearance during therapy or, in some rare cases, life-threatening complications [10]. Glycans trigger immunogenic reactions either by being directly recognised as an exogenous glycan sequence by the patient immune system [11] or by affecting folding, solubility (e.g., formation of aggregates) and structural stability of the recombinant protein, which triggers the immune system of the patient [12].

Considering the importance of glycosylation for the production of recombinant biopharmaceuticals, it becomes fundamental to gather extensive information regarding the glycosylation status of a biofactory to validate their suitability for the biopharmaceutical industry. Over the years, in fact, multiple studies analysed the glycosylation machinery of different well-established biofactories. At least four non-human glycans have been recognized as immunogenic in humans. These residues are galactose-α(1,3)-galactose (α-Gal), *N*-glycolylneuraminic acid (Neu5Gc), β(1,2)-xylose, and α(1,3)-fucose [12]. The α-Gal and Neu5Gc residues are present in recombinant biopharmaceuticals produced in mammalian cells such as CHO cells, while β(1,2)-xylose and α(1,3)-fucose are found on glycoproteins produced in plants and microalgae [12].

Microalgae are alternative biofactories that have gained momentum over the past twenty years for recombinant biopharmaceutical production. These organisms present multiple advantages over other broadly-used cell hosts, such as a eukaryotic PTM machinery (unlike *E. coli*), adequate secretion of recombinant biopharmaceuticals (unlike yeast species), higher growth rates at potential lower costs, and immunity to human pathogens (unlike CHO cells). One of the most utilised microalgal biofactory is the unicellular Chlorophyta *Chlamydomonas reinhardtii* [13,14]. Compared to other well-established biofactories, little is known about the glycosylation status of *C. reinhardtii*. Four studies have profiled glycosylation in multiple *C. reinhardtii* strains (cc-503 cw92, cc-1036 pf18, cc-400 cw15, cc-4375, and cc-4533) [11,15,16,17], and they reported possible absence of the *N*-acetylglucosaminyltransferase I (GnT I) enzyme (responsible for the formation of "complex" and "hybrid" glycans), and presence of both immunogenic β(1,2)-xylose and α(1,3)-fucose residues.

Unfortunately, immunogenic glycosylation is not the only issue that afflicts *C. reinhardtii*; another major drawback in this species is the low recombinant protein yield. Low nuclear transgene expression in *C. reinhardtii* is mainly due to random integration during transformation and gene silencing mechanisms [18,19,20]. To improve recombinant protein yields, Neupert and colleagues [18] generated a highly expressive UV mutated strain called UVM4 that is capable of high yields of secreted

recombinant proteins (>12 mg/L) [21,22]. Today, strain UVM4 is widely used for recombinant protein production in microalgae [21,22]. However, as described in Chapter 2, the altered molecular processes in strain UVM4 are still unknown, including glycosylation. In fact, no study has analysed the glycosylation pattern of strain UVM4 so far.

Here, I present the first analysis of the glycoprotein profile and glycosylation patterns of the cell wall-deficient UV mutated *C. reinhardtii* strain UVM4, compared to the common wild type strain cc-125 *mt+* (also named strain 137c). This work aims to identify the differences in glycoprotein abundance in the extracellular space and in glycosylation patterns between strains. This information is critical to address the issue of immunogenicity in *C. reinhardtii* and to inform future efforts towards recombinant glycoprotein production in this emerging alternative cell biofactory.

3. MATERIALS AND METHODS

C. reinhardtii cultivation conditions and secretomes harvesting

C. reinhardtii wild type strain cc-125 *mt+* (also named strain 137c) was purchased from Invitrogen (GeneArt® Chlamydomonas Protein Expression Kit). *C. reinhardtii* strain UVM4 was graciously provided by Prof. Ralph Bock. *C. reinhardtii* strains 137c and UVM4 were grown under mixotrophic conditions as described in Chapter 2. The secretomes were collected after 72 hours of growth using ultrafiltration technics as described in Chapter 2.

Protein extraction and gel electrophoresis

The concentrated secretomes were precipitated to remove non-protein contaminants, resuspended in 100 mM of triethylammonium bicarbonate buffer (TEAB) with 1M urea, and quantified using the Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific) protein assay as reported in Chapter 2. A portion of the proteins was visualized by gel electrophoresis, while the rest was treated for glycans analysis.

The protein and glycoprotein profile of the samples was investigated on a polyacrylamide gel. The same amount of total protein (~100 µg) was first treated with 1x Laemmli sample buffer (Bio-Rad) containing β-mercaptoethanol (9:1 buffer to β-mercaptoethanol and 3:1 sample to buffer ratio) and boiled for 10 minutes at 95°C. The boiled samples were loaded in triplicates on two 4 – 15% Criterion™ TGX™ Precast Gel (Bio-Rad) and separated according to size by gel electrophoresis in Tris/Glycine/SDS buffer (Bio-Rad) for 40 minutes at 300 V and 290 mA. The ladder used was the Precision Plus Protein™ Dual Colour Standards (Bio-Rad) with protein size range of 10 – 250 kDa. Two protein staining protocols were performed: Coomassie staining and glyco-staining. After gel electrophoresis, one protein gel was incubated with a fixing solution (50% methanol and 10% acetic acid) for 1 hour. Subsequently, the fixed gel was incubated overnight with the Coomassie solution (50% methanol, 10% acetic acid, and 0.25% Coomassie Brilliant Blue G-250 (ThermoFisher)). The gel was then destained by soaking in the

fixing solution for 15 minutes before being washed with distilled water until the background was cleared. For the second protein gel, the Pierce™ Glycoprotein Staining Kit (Thermo Fisher) was used to visualise only glycosylated proteins. Glycoproteins were stained using the periodic acid-Schiff (PAS) method with the fuchsine sulphite dye. Positive and negative control proteins (HRP and STI, respectively) were tested previously (data not shown). The Coomassie-stained and fuchsine-stained gels were visualised using the Typhoon FLA 9000 Gel Imager (GE).

Glycan release and porous graphitized carbon liquid chromatography electrospray liquid tandem mass spectrometry (PGC-LC-ESI-MS/MS)

To release intact *N*-linked glycans attached to the glycoproteins, the extracellular proteomes were treated with the enzyme Protein *N*-glycosidase A (PNGase A). This enzyme was chosen over the *N*-glycosidase F (PNGase F) because the activity of PNGase F is inhibited in presence of α(1,3)-fucose, a residue characterised in *C. reinhardtii*. For the *O*-glycans, a β-elimination detachment was performed [23,24,25]. Sample preparation was performed using the desalting/reduction protocol adapted from Jensen et al. [26]. Different monosaccharides and isobaric isomers were separated by porous graphitized carbon liquid chromatography (PGC-LC) and subsequently analysed by LC-ESI-MS/MS. Sample separation was performed on a PGC-LC column (3µm, 100 mm × 0.18 mm, Hypercarb, Thermo Scientific) column with a binary gradient elution of ultrapure water (with 10 mM ammonium bicarbonate) and 45% acetonitrile (with 10 mM ammonium bicarbonate) at a flow rate of 4 µL/min. Mass spectrometry was performed using a Linear Trap Quadrupole (LTQ) Velos Pro ion trap (Thermo Scientific) mass spectrometer connected to an UltiMate3000 high-performance liquid chromatography (HPLC) system (Dionex). Data acquisition was done in negative ion mode with a source voltage of -3.2 kV at a mass range of m/z 580-2000 (adapted from [26]).

Data analysis

Lists of ions from relevant spectra of each sample were extracted using the Excalibur v4.1 (Thermo Scientific). Masses corresponding to common contaminants were removed and monoisotopic precursor masses potentially corresponding to glycans were searched against GlycoMod

(http://www.expasy.ch/tools/glycomod) to identify putative monosaccharide compositions. Further interpretation and molecular identification were carried out based on the product ions information from MS/MS scans. GlycoworkBench v2.1 (https://code.google.com/archive/p/glycoworkbench/) was used for most glycan product ion annotation (adapted from [26]).

4. RESULTS & DISCUSSION

4.1. *C. reinhardtii* growth

The *C. reinhardtii* strains 137c and UVM4 were grown for 72 hours (growth curves shown in Chapter 2, Figure 2.2B) and harvested using crossflow ultrafiltration with a 3 kDa cut-off as reported in Chapter 2. An aliquot of the secretomes collected was used to perform the glycoprotein profiling (gel electrophoresis) and the glycosylation analysis (PGC-LC-ESI-MS/MS).

4.2. Strains 137c and UVM4 have different glycoprotein profiles

As reported in Chapter 2, strains 137c and UVM4 harvested at 72 hours exhibit major differences in the abundance of proteins found in their extracellular proteomes. To investigate presence/absence, molecular weight, and relative abundance (intensity) of proteins and glycoproteins in the extracellular spaces of strains 137c and UVM4, I performed two protein gels: i) a Coomassie stained gel to detect total proteins profile, (Figure 4.1A) and ii) a glycoprotein stained gel to detect only the glycosylated proteins present in the extracellular spaces of the two strains (Figure 4.1B).

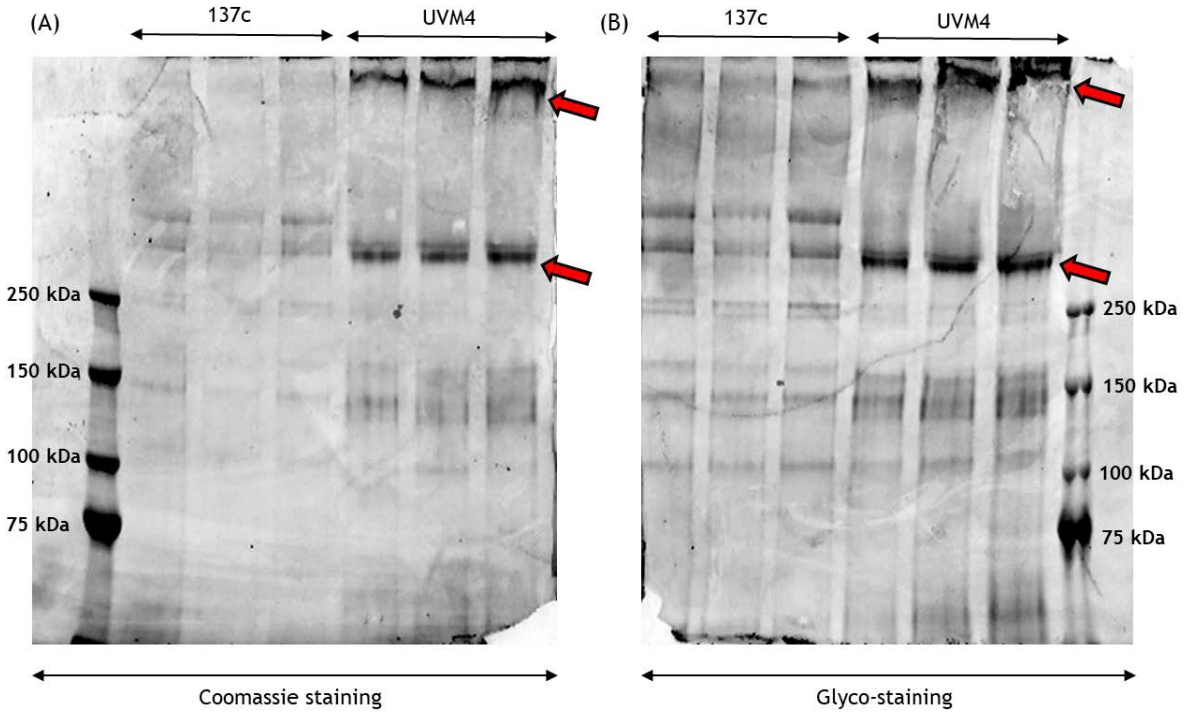


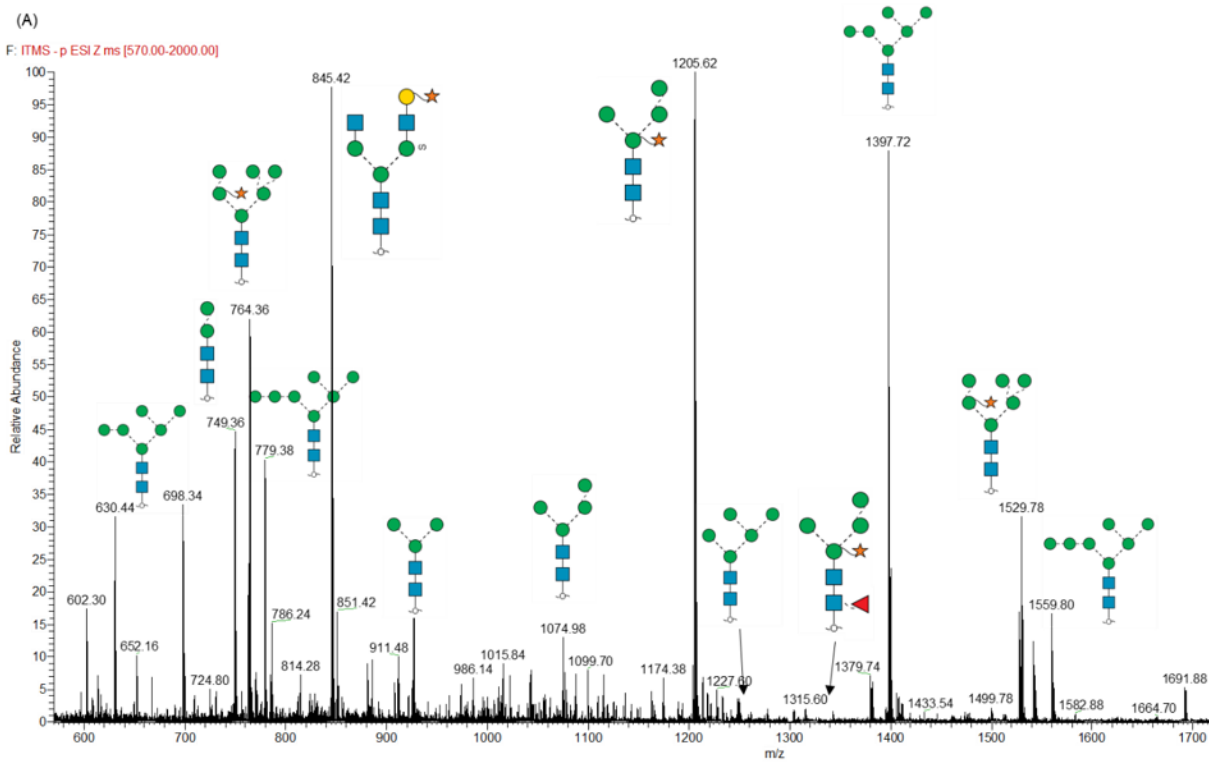
Figure 4.1. Extracellular proteome gel, stained with (A) Coomassie and (B) glycoprotein staining. Sample loaded: ~100 µg total protein, n=3. Intense staining of high molecular weight bands in both Coomassie and glyco-stained gel suggests overabundance of glycosylated extracellular proteins in strain UVM4 relative to strain 137c. Proteins with molecular weight lower than 50 kDa did not show detectable differences. Red arrows indicate the most intense bands detected.

The two strains displayed different protein and glycoprotein profiles in the extracellular space. This was expected considering the differences reported in relative abundance of proteins in the different secretomes (Chapter 2). The differences in pattern and intensity of the bands between Coomassie staining and glycoprotein staining was expected, considering that not all proteins are glycosylated and that the glycostaining is more sensitive and will show bands not visualised on the Coomassie stained gel.

Interestingly, the glycoprotein profiles between the two strains showed extensive differences. The major alterations in pattern and especially in intensity of proteins and glycoproteins between the two strains were noticeable for proteins with molecular weight higher than 250 kDa. In particular, two intense bands > 250 kDa (in triplicates) were found in strain UVM4 (red arrows shown in Figure 4.1A and 4.1B). While glycoprotein stained gel allowed me to indicate the difference in abundance of glycosylated proteins between the two strains (Figure 4.1B), it did not indicate the presence or absence of specific residues, such as immunogenic algal β(1,2)-xylose and α(1,3)-fucose residues.

4.3. Glycosylation analysis shows immunogenic residues in both strain 137c and UVM4

To analyse all the glycosylation patterns of all solubilised proteins present in the extracellular space of strains 137c and UVM4, a mass spectrometry analysis was performed. Given the differences between the two strains in terms of relative abundance of extracellular proteins reported in Chapter 2, and the evident differences shown in the protein gel (Figure 4.1B), major differences in abundance of glycoforms were expected.


(A)
F: ITMS - p ESI Z ms [570.00-2000.00]
Relative Abundance
m/z
845.42
1205.62
1397.72
764.36
749.36
779.38
698.34
630.44
602.30
652.16
786.24
851.42
724.80
814.28
911.48
1074.98
1015.84
986.14
1099.70
1174.38
1227.60
1379.74
1315.60
1529.78
1559.80
1433.54
1499.78
1582.88
1664.70
1691.88

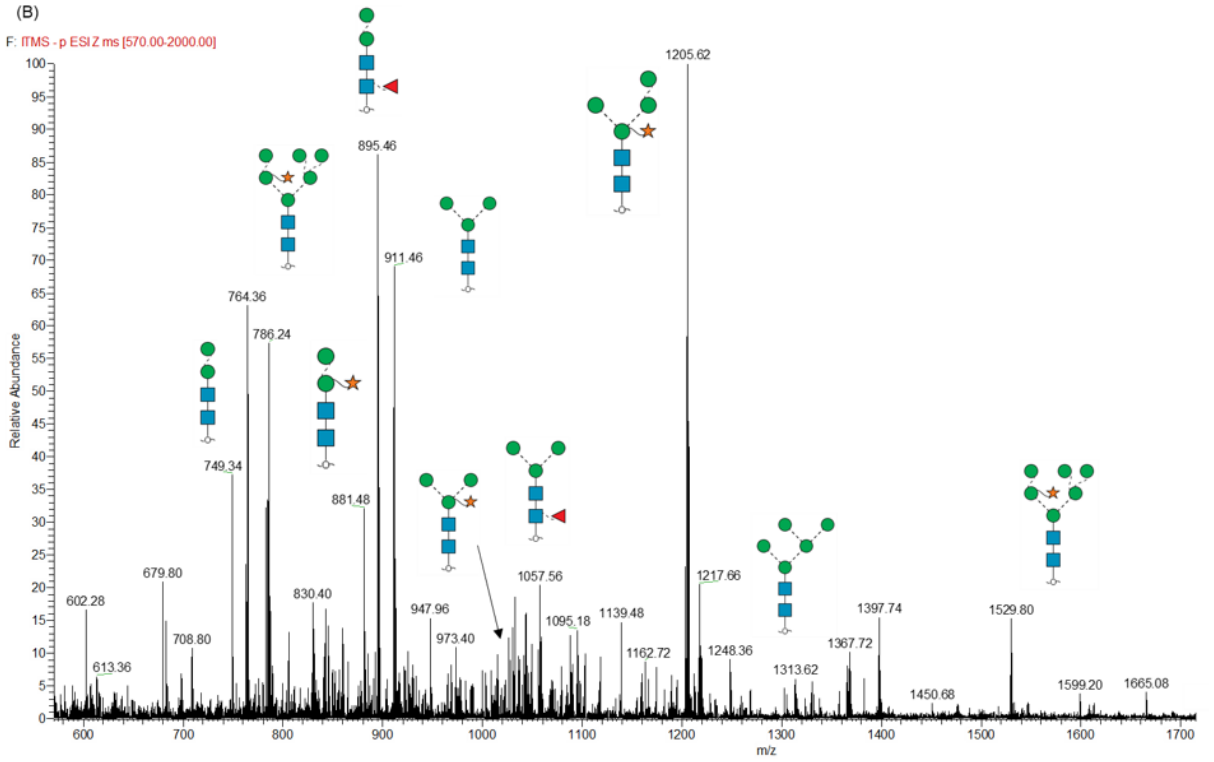

(B)
F: ITMS - p ESI Z ms [570.00-2000.00]
Relative Abundance
m/z
1205.62
895.46
911.46
764.36
786.24
749.34
881.48
679.80
602.28
830.40
708.80
613.36
947.96
973.40
1057.56
1095.18
1139.48
1162.72
1217.66
1248.36
1313.62
1367.72
1397.74
1450.68
1529.80
1599.20
1665.08

Figure 4.2. Mass spectra of glycan population of extracellular proteins of (A) strain 137c and (B) strain UVM4. Green circle = Mannose. Yellow circle = Galactose. Blue square = GlcNAc. Yellow star = Xylose. Red triangle = Fucose. The relative peak abundance was calculated by normalizing the intensity of each peak to the intensity of the highest peak, which occurs at 1205.62 m/z ratio.

The obtained spectra showed some similarities and major differences in peaks reported between strain 137c and strain UVM4 (Figure 4.2). Further, the putative glycans attributed to each peak were identified. Unfortunately, *O*-glycans were not detected in the mass spectra of both strains. Considering the extensive presence of *O*-glycosylated cell wall proteins in the extracellular space of strain UVM4 (Chapter 2, section 4.2.4), it was surprising to be unable to detect any *O*-glycans in this species. It is possible that the glycan detachment procedure or the sample preparation was inadequate to analyse *O*-glycoproteins. Therefore, only the *N*-glycoprofiles of both strains were compared and investigated further.

4.3.1. Similarities and differences in glycosylation status between strains

The most important similarity between strains 137c and UVM4 was the presence of the two microalgal immunogenic residues: β(1,2)-xylose and α(1,3)-fucose.

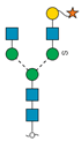
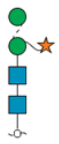
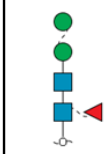
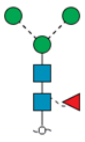
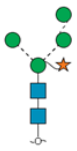
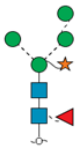
	764.36 m/z	845.42 m/z	881.48 m/z	895.46 m/z	1043.46 m/z	1057.56 m/z	1205.62 m/z	1351.66 m/z
								
Strain 137c	✓	✓					✓	✓
Strain UVM4	✓		✓	✓	✓	✓	✓	

Figure 4.3. Immunogenic glycoforms identified for strains 137c and UVM4. Green circle = Mannose. Yellow circle = Galactose. Blue square = GlcNAc. Yellow star = Xylose. Red triangle = Fucose. A tick symbol (✓) indicates presence of the glycoform in the strain mass spectra.

The most intense peak in both spectra was identified as a core-xylose glycoform, often detected in *C. reinhardtii* (peak 1205.62, Figure 4.2A and 4.2B) [11,15]. Other glycoforms in both strains showed the attachment of a xylose residue on the core mannose or one of the antennae, another well-known *C. reinhardtii* glycosylation pattern (peaks 764.36 m/z, 845.42 m/z, 881.48 m/z, 1043.46 m/z, 1351.66 m/z; Figure 4.3). The reported immunogenic residue α(1,3)-fucose was also detected in both strains (peaks 895.46 m/z, 1057.56, 1351.66 m/z, Figure 4.3); moreover, in strain UVM4 the peak assigned to this glycoform was the second most abundant (peak 895.46 m/z, Figure 4.2B).

The major difference between the two strains was the relative abundance of the different glycoforms. The two mass spectra showed i) presence or absence of very different peaks (attributed to different glycans), and ii) presence of similar peaks but with significantly different relative abundances. As said before, the abundance of glycoproteins has a fundamental impact on the glycosylation profile reported. Not detecting one glycoform does not necessarily relate to the absence of the glycosylation enzyme(s) responsible for that glycan. In fact, the reasons for the absence of a residue or glycoform might depend on detection limit of the instrument used and relative abundance of the glycan. Therefore, while this analysis cannot unambiguously prove that the glycosylation machinery of strain UVM4 is different from 137c, it does show that the population of glycoproteins (glycoprofile) in the extracellular space of the two strains is remarkably different.

Five important differences were detected between the glycoprofiles of strain UVM4 and strain 137c. Firstly, strain 137c presents non-xylosylated “high mannose” glycans (peaks 698.34, 779.38, 1397.72, and 1559.80 m/z, Figure 4.2A) not identified in strain UVM4. “High mannose” glycans are one of the most common glycoforms detected in *C. reinhardtii* [11,15], therefore it was not surprising to find multiple peaks and high relative abundance of these glycans in strain 137c. However, it is interesting to report that non-xylosylated “high mannose” glycans were not detected in strain UVM4 in this study. It would be interesting to determine the role of “high mannose” glycoproteins in *C. reinhardtii* to possibly understand their lower abundance in the secretome of strain UVM4.

Secondly, strain UVM4 presents a peak at 895.46 m/z attributed to a core-fucosylated glycoform (Figure 4.2B) and another peak attributed to a core-fucose glycoform (peak 1057.56 m/z, Figure 4.2B), that were not found in strain 137c. A core-fucose glycan was detected in strain 137c, however its relative abundance was low.

The third important difference between the two strains was the putative detection of a galactose residue (peak 845.42 m/z, Figure 4.2A). This monosaccharide has never been detected before in this species, although galactose has been identified in *Botryococcus braunii*, which belongs to the same phylum as *C. reinhardtii* (Chlorophyta), and therefore the presence of galactose in *C. reinhardtii* is conceivable. In Chapter 1, I reported *in silico* presence or absence of glycosylation enzymes homologs in multiple species, including *C. reinhardtii* and *B. braunii* (Figure 1.3). Neither of them reported presence of galactosyltransferase homologs. Considering the putative presence of this residue in both species, it is possible that an alternative galactosyltransferase might be operating in *C. reinhardtii* and *B. braunii*.

The fourth important difference between the two strains was the different relative abundance of the peak at 911.46 – 911.48 m/z. This peak was identified as the "pentasaccharide core" in both strains, however the relative abundance was significantly lower in strain 137c. The "pentasaccharide core" is the most important structure for *N*-glycosylation. In fact, this structure is well-preserved across all eukaryotes, and it is also present on every *N*-glycan. It is possible that the high relative abundance of this structure in strain UVM4 is related to a general higher abundance of *N*-glycoproteins in this strain compared to strain 137c. The protein gel stained with glyco-staining (Figure 4.1B) showed higher intensity of bands in strain UVM4 (especially high molecular weight proteins, indicated by red arrows in Figure 4.1), reporting a higher abundance of glycoproteins in this strain. However, a glycomics approach is unable to correlate a glycan to its carrier glycoprotein; therefore, to be able to target and characterise specific glycoproteins, and to identify their glycans, a glycoproteomics approach is needed. The selected proteins (such as the high molecular weight proteins more intense in strain UVM4, Figure 4.1) should be excised from the gel and analysed singularly. Glycoproteomics might be useful to identify targeted glycoprotein patterns and possibly to unveil the reasons for the higher abundance in strain UVM4 of the peak associated with the "pentasaccharide core". And lastly, the fifth difference

between strains 137c and UVM4 was the presence of peak 845.42 m/z in strain 137c (peak absent in strain UVM4).

4.3.2. Strain 137c shows a putative "complex" glycoform

Peak 845.42 m/z, present in strain 137c but absent in strain UVM4 (Figure 4.2A and 4.2B), was attributed to a "complex" *N*-glycan. As extensively described in Chapter 1, the formation of complex glycans is possible only in the presence of GnT I. This enzyme in mammalian cells is located in the Golgi apparatus and its role is to attach an *N*-acetylglucosamine residue on the newly formed glycan. Microalgal species are described as either GnT I-dependent or GnT I-independent species. GnT I-dependent species, such as *Phaeodactylum tricornutum* and *B. braunii*, are able to generate "hybrid" and "complex" glycans, whilst GnT I-independent species, such as *Chlorella vulgaris* and *Porphyridium purpureum*, only produce "high mannose" glycoforms. "Complex" and "hybrid" glycans have never been experimentally detected in *C. reinhardtii* previously, therefore it was considered a GnT I-independent species. However, the putative identification of "complex" glycoforms in *C. reinhardtii* results shown in Figure 4.2A suggest that a reinvestigation of the *N*-glycosylation pathway in this species might be needed.

Only four studies have analysed glycosylation in *C. reinhardtii* before this study [11,15,16,17], none of which reported "complex" or "hybrid" glycans. However, considering the direct correlation between abundance of a glycoprotein and the amount of the released glycan in the sample, and the importance of glycan abundance for a mass spectrometry analysis, a different experimental design performed in the previous studies might have affected the obtained glycoprofile. In fact, several factors such as strain, culture conditions, harvesting time, and fractions analysed can all have an impact on the abundance of glycoproteins populating the extracellular space of the cultures, resulting in different glycoprofiles obtained and analysed. The four cited studies treated and analysed the samples differently. First, none of the previous studies have analysed the glycoprofile of strain 137c nor strain UVM4. The strains analysed were cc-503 cw92, cc-1036 pf18, cc-400 cw15, cc-4375, and cc-4533. Mathieu-Rivet and colleagues [15] hypothesised that different strains will show the same glycosylation status. However, in

this study, major differences between strain 137c and strain UVM4 were reported; therefore it seems possible to find strain-specific differences within the same species. Moreover, the other experiments were conducted under different environmental conditions and with a different experimental aim and design [11,15,16,17]. In fact, this study is the first that focuses solely on total native glycosylation status of extracellular proteins. Mathieu-Rivet and colleagues [15] performed targeted native glycosylation analysis of multiple fractions. They collected different fractions, such as chloroplast, plasma membrane, intracellular, and extracellular proteins, loaded onto a SDS-PAGE, prepared in-gel, and analysed by mass spectrometry [15]. This approach is called glycoproteomics, which focuses on subcellular location of the glycoproteins. In contrast, this analysis was focused on the whole extracellular glycoprotein profile, an approach called glycomics. In addition, the other three studies [11,16,17] performed genetic engineering of *C. reinhardtii* to produce specific exogenous glycosyltransferases, and subsequently analysed glycoproteins using a glycoproteomic approach. Schulze and colleagues [16] and Oltmanns and colleagues [17] focused on analysis of methyltransferase, xylosyltransferase, and fucosyltransferase activity, while Vanier and colleagues [11] analysed the activity of heterologous GnT I enzyme. Interestingly, Vanier and colleagues [11] successfully expressed an exogenous GnT I (cDNAs originated from *Arabidopsis sp.* and *P. tricornutum*); however, no modification of the glycans was reported. They hypothesised three possible causes for inactivity of heterologous GnT I: i) the recombinant enzyme was not located in the Golgi apparatus, ii) absence or inadequate activity of the GlcNAc-transporter enzyme responsible for the accumulation of GlcNAc in the Golgi apparatus, and iii) absence of the substrate glycan targeted by the GnT I enzyme [11]. The putative presence of "complex" glycoforms reported in this study in strain 137c, and the inactivity of the conventional GnT I enzyme, suggests the presence of an alternative GnT I enzyme in *C. reinhardtii*. This hypothesis is also supported by computational data. In fact, an *in silico* analysis of glycosylation pathway genes in multiple microalgal species did not identify a possible GnT I homolog in *C. reinhardtii* (Mathieu-Rivet et al. 2014[6]). I performed a similar analysis (Chapter 1, Figure 1.3) and obtained the same result. Therefore, considering that the GnT I enzyme was not detected in *C. reinhardtii* but its product was, there might be an alternative pathway operating in this species. While these results are interesting, they require confirmation with increased replicates and analysis of patterns throughout a growth phase to

clearly confirm the presence of more “complex” glycans, and the existence of an alternative *N*-acetylglucosaminyltransferase in *C. reinhardtii*.

5. CONCLUSION

Here I provide the first glycosylation analysis and comparison of the extracellular proteome of *Chlamydomonas reinhardtii* strain UVM4 with the non-mutant strain 137c. This glycomics analysis offers new insight on the glycosylation status of a well-established *C. reinhardtii* strain, currently broadly used for biotechnological applications.

I reported major differences in the glycoprofiles of the two strains, indicating that different strains grown under the same condition can show different abundance of glycosylated proteins in the extracellular space. I also reported the presence of immunogenic residues in multiple glycoforms of strain UVM4, suggesting possible attachment of immunogenic glycans on recombinant glycoproteins produced in this host. Moreover, I also identified a putative "complex" glycoform in the extracellular space of strain 137c, a glycan never reported in *C. reinhardtii* before. The possible presence of this glycoform in *C. reinhardtii* imposes a reconsideration of this species as a GnT I-independent microalga, and requires further investigations.

Future studies should focus on the analysis and identification of other "complex" or "hybrid" glycoforms, to validate the results shown in this study and identify the enzyme responsible for these glycoforms. In fact, the formation of these glycans is directly related to the presence of the GnT I enzyme. Previous *in silico* studies (including Chapter 1) reported the absence of GnT I homologs [15]. Moreover, another study reported the inactivity of a heterologous GnT I produced in *C. reinhardtii* [11]. Therefore, it is possible that an alternate GnT I enzyme is present in this species and responsible for the formation of "complex" and "hybrid" glycans, but future studies will be necessary to prove it.

In conclusion, the knowledge generated in this study brings new insights into the glycosylation status of proteins in *C. reinhardtii*. Obtaining the full picture is fundamental to validate this species as a cell biofactory for recombinant biopharmaceutical production.

6. REFERENCES

1. An, H.J.; Froehlich, J.W.; Lebrilla, C.B. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr Opin Chem Biol* **2009**, *13*, 421-426, doi:10.1016/j.cbpa.2009.07.022.

2. Marshall, R.D. The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp* **1974**, 17-26.

3. Ohtsubo, K.; Marth, J.D. Glycosylation in cellular mechanisms of health and disease. *Cell* **2006**, *126*, 855-867, doi:10.1016/j.cell.2006.08.019.

4. Bürkle, A. Posttranslational Modification. In *Encyclopedia of Genetics*, Brenner, S., Miller, J.H., Eds. Academic Press: New York, 2001. 1533.

5. Mizukami, A.; Caron, A.L.; Picanço-Castro, V.; Swiech, K. Platforms for Recombinant Therapeutic Glycoprotein Production. In *Recombinant Glycoprotein Production: Methods and Protocols*, Picanço-Castro, V., Swiech, K., Eds. Springer New York: New York, NY, 2018; 10.1007/978-1-4939-7312-5_1pp. 1-14.

6. Mathieu-Rivet, E.; Kiefer-Meyer, M.C.; Vanier, G.; Ovide, C.; Burel, C.; Lerouge, P.; Bardor, M. Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front Plant Sci* **2014**, *5*, 359, doi:10.3389/fpls.2014.00359.

7. Walsh, G.; Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* **2006**, *24*, 1241-1252, doi:10.1038/nbt1252.

8. Higgins, E. Carbohydrate analysis throughout the development of a protein therapeutic. *Glycoconjugate Journal* **2010**, *27*, 211-225, doi:10.1007/s10719-009-9261-x.

9. Butler, M.; Spearman, M. The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr Opin Biotechnol* **2014**, *30*, 107-112, doi:10.1016/j.copbio.2014.06.010.

10. Kessler, M.; Goldsmith, D.; Schellekens, H. Immunogenicity of biopharmaceuticals. *Nephrol Dial Transplant* **2006**, *21 Suppl 5*, v9-12, doi:10.1093/ndt/gfl476.

11. Vanier, G.; Lucas, P.L.; Loutelier-Bourhis, C.; Vanier, J.; Plasson, C.; Walet-Balieu, M.L.; Tchi-Song, P.C.; Remy-Jouet, I.; Richard, V.; Bernard, S., et al. Heterologous expression of the N-

acetylglucosaminyltransferase I dictates a reinvestigation of the N-glycosylation pathway in *Chlamydomonas reinhardtii*. *Sci Rep* **2017**, *7*, 10156, doi:10.1038/s41598-017-10698-z.

12. van Beers, M.M.C.; Bardor, M. Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins. *Biotechnology Journal* **2012**, *7*, 1473-1484, doi:10.1002/biot.201200065.

13. Chavez, M.N.; Schenck, T.L.; Hopfner, U.; Centeno-Cerdas, C.; Somlai-Schweiger, I.; Schwarz, C.; Machens, H.G.; Heikenwalder, M.; Bono, M.R.; Allende, M.L., et al. Towards autotrophic tissue engineering: Photosynthetic gene therapy for regeneration. *Biomaterials* **2016**, *75*, 25-36, doi:10.1016/j.biomaterials.2015.10.014.

14. Baier, T.; Kros, D.; Feiner, R.C.; Lauersen, K.J.; Muller, K.M.; Kruse, O. Engineered Fusion Proteins for Efficient Protein Secretion and Purification of a Human Growth Factor from the Green Microalga *Chlamydomonas reinhardtii*. *ACS Synth Biol* **2018**, 10.1021/acssynbio.8b00226, doi:10.1021/acssynbio.8b00226.

15. Mathieu-Rivet, E.; Scholz, M.; Arias, C.; Dardelle, F.; Schulze, S.; Le Mauff, F.; Teo, G.; Hochmal, A.K.; Blanco-Rivero, A.; Loutelier-Bourhis, C., et al. Exploring the N-glycosylation pathway in *Chlamydomonas reinhardtii* unravels novel complex structures. *Mol Cell Proteomics* **2013**, *12*, 3160-3183, doi:10.1074/mcp.M113.028191.

16. Schulze, S.; Oltmanns, A.; Machnik, N.; Liu, G.; Xu, N.; Jarmatz, N.; Scholz, M.; Sugimoto, K.; Fufezan, C.; Huang, K., et al. N-Glycoproteomic Characterization of Mannosidase and Xylosyltransferase Mutant Strains of *Chlamydomonas reinhardtii*. *Plant Physiol* **2018**, *176*, 1952-1964, doi:10.1104/pp.17.01450.

17. Oltmanns, A.; Hoepfner, L.; Scholz, M.; Zinzius, K.; Schulze, S.; Hippler, M. Novel Insights Into N-Glycan Fucosylation and Core Xylosylation in C. reinhardtii. *Front Plant Sci* **2019**, *10*, 1686, doi:10.3389/fpls.2019.01686.

18. Neupert, J.; Karcher, D.; Bock, R. Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J* **2009**, *57*, 1140-1150, doi:10.1111/j.1365-313X.2008.03746.x.

19. Scaife, M.A.; Nguyen, G.T.; Rico, J.; Lambert, D.; Helliwell, K.E.; Smith, A.G. Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host. *Plant J* **2015**, *82*, 532-546, doi:10.1111/tpj.12781.

20. Schroda, M. Good News for Nuclear Transgene Expression in *Chlamydomonas*. *Cells* **2019**, *8*, doi:10.3390/cells8121534.

21. Lauersen, K.J.; Huber, I.; Wichmann, J.; Baier, T.; Leiter, A.; Gaukel, V.; Kartushin, V.; Rattenholl, A.; Steinweg, C.; von Riesen, L., et al. Investigating the dynamics of recombinant protein secretion from a microalgal host. *J Biotechnol* **2015**, *215*, 62-71, doi:10.1016/j.jbiotec.2015.05.001.

22. Ramos-Martinez, E.M.; Fimognari, L.; Sakuragi, Y. High-yield secretion of recombinant proteins from the microalga *Chlamydomonas reinhardtii*. *Plant Biotechnol J* **2017**, *15*, 1214-1224, doi:10.1111/pbi.12710.

23. Carlson, D.M. Structures and immunochemical properties of oligosaccharides isolated from pig submaxillary mucins. *J Biol Chem* **1968**, *243*, 616-626.

24. Lloyd, K.O.; Burchell, J.; Kudryashov, V.; Yin, B.W.; Taylor-Papadimitriou, J. Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines. Demonstration of simpler and fewer glycan chains in tumor cells. *J Biol Chem* **1996**, *271*, 33325-33334, doi:10.1074/jbc.271.52.33325.

25. Morelle, W.; Canis, K.; Chirat, F.; Faid, V.; Michalski, J.-C. The use of mass spectrometry for the proteomic analysis of glycosylation. *PROTEOMICS* **2006**, *6*, 3993-4015, doi:10.1002/pmic.200600129.

26. Jensen, P.H.; Karlsson, N.G.; Kolarich, D.; Packer, N.H. Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc* **2012**, *7*, 1299-1310, doi:10.1038/nprot.2012.063.

CHAPTER 5: Synthesis, limitations, future perspectives, and concluding remarks

Lorenzo Barolo[1], Raffaela M. Abbriano[1], Audrey S. Commault[1], Matthew P. Padula[2], and Mathieu Pernice[1]

AFFILIATIONS

[1] University of Technology Sydney, Climate Change Cluster, Broadway Campus, Ultimo NSW 2007, Sydney, Australia

[2]School of Life Sciences and Proteomics Core Facility, Faculty of Science, University of Technology Sydney, Ultimo NSW 2007, Sydney, Australia

1. SUMMARY OF THESIS

Production of recombinant proteins from microalgae is an emerging field with great potential despite some significant drawbacks. These obstacles need to be overcome before microalgae, including *Chlamydomonas reinhardtii,* can be successfully employed as an industrial biofactory. As highlighted in Chapters 1 & 4, glycosylation is a particularly important issue that can affect the function and immunogenicity of recombinant biopharmaceuticals produced from *C. reinhardtii* (and other microalgal species), and clearly needs to be further characterized. However, the major stumbling block to commercial relevance of *C. reinhardtii* as a platform for recombinant protein production is the inadequate yield, which is significantly lower than what can be achieved in more established systems. The ultimate goal of this thesis was to report, analyse, and possibly overcome the two drawbacks that afflict *C. reinhardtii*: low yield and potential immunogenicity of recombinant proteins. To do so, I performed the first comparative protein analysis of the highly expressive and broadly-used *C. reinhardtii* UV mutated strain UVM4. This strain was generated by UV mutagenesis [4] and is capable of higher secreted recombinant protein yields than *C. reinhardtii* non-mutated strains (up to 3-fold higher) [1,2]. However, the molecular mechanisms driving these higher yields remain unknown. The total comparative proteome analysis of wild type and genetically engineered cell lines of strains 137c and UVM4 (Chapters 2 & 3) enabled the identification of molecular processes altered in strain UVM4 wild type (Chapter 2) and in *C. reinhardtii* after genetic engineering (Chapter 3). The detailed list of altered proteins and modified pathways reported in this study provide potential protein targets for genetic engineering to optimise recombinant protein production in *C. reinhardtii*. In addition, the comparative glycosylation analysis performed in Chapter 4 provided the first insight on the glycosylation pattern of strain UVM4, while contributing novel information on the glycosylation machinery of the non-mutant strain 137c. The analysis of glycosylation status in *C. reinhardtii* reported the presence of immunogenic glycans in both strains 137c and UVM4. However, considering the extensive information presented in Chapter 1 regarding glyco-engineering, it might be possible to target the enzymes responsible to improve the quality of the recombinant proteins produced in *C. reinhardtii*.

In conclusion, this thesis provides a detailed analysis of the two major obstacles in recombinant protein production in *C. reinhardtii*, unravelling possible causes and potential solutions, with the ultimate goal of establishing *C. reinhardtii* as a commercial and sustainable platform for recombinant protein production.

2. SYNTHESIS

2.1. Comparative proteomic analysis of *C. reinhardtii* strains 137c and UVM4 (Chapter 2)

C. reinhardtii UV mutated strain UVM4 is a well-established model used worldwide for microalgae research and more particularly for recombinant protein production. Its biotechnological value lies in its high secreted recombinant protein yields. However, as frequently observed for organisms generated by mutagenesis, the strain has been chosen based on the phenotype and the mutated DNA loci have not been characterised. Therefore, the molecular mechanisms driving enhanced recombinant protein production in strain UVM4 are still unknown. Considering that yields obtained in strain UVM4 can be up to 3-fold higher that non-mutated strains of *C. reinhardtii* [1,2], but still far from yields obtained in CHO cells, the analysis of the altered pathways responsible for higher yields in UVM4 might help to further optimise *C. reinhardtii* strains for recombinant protein production with the long-term intention to match mammalian cell yields. In Chapter 2, I conducted comparative proteomics on the wild type strain 137c and strain UVM4 to elucidate some of the possible molecular processes driving enhanced recombinant protein production in strain UVM4. I analysed and compared the intracellular proteomic profiles, to reveal potential altered pathways involved in transcription and translation. In addition, I also analysed and compared the proteins populating the extracellular space, a relevant issue for production of secreted recombinant proteins.

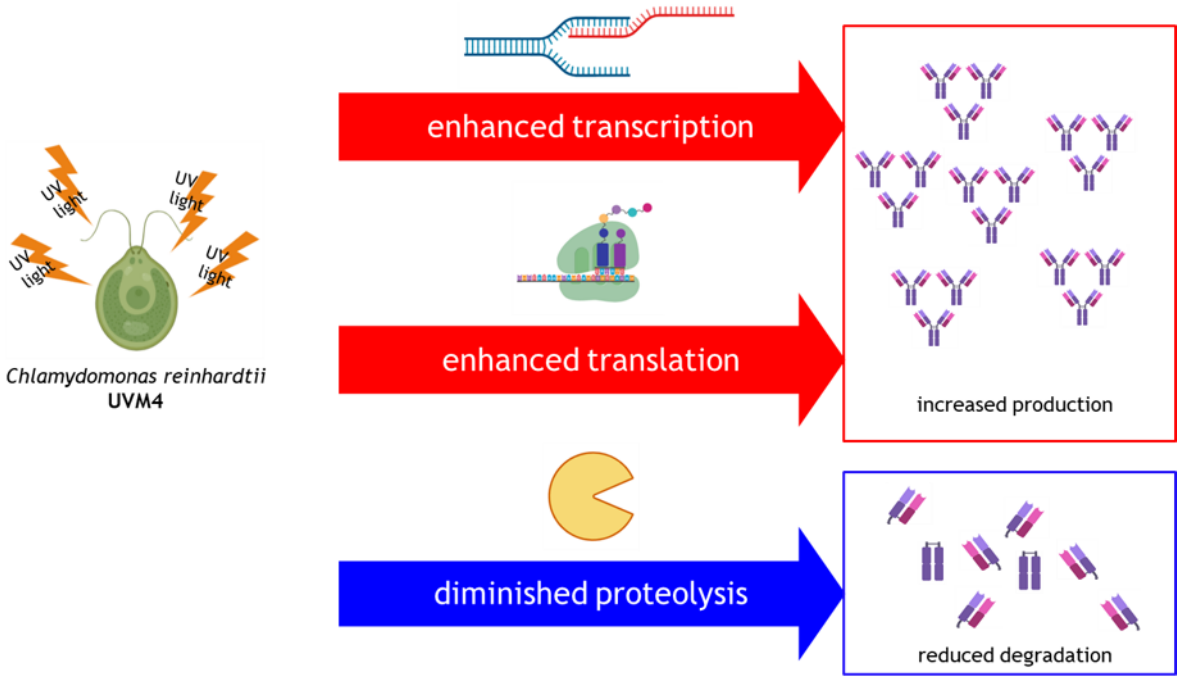


Figure 5.1. Schematic diagram showing the altered molecular processes detected in strain UVM4.

These results revealed several proteins showing significantly higher or lower abundance in strain UVM4 than strain 137c and consequently multiple altered cellular and molecular processes in the UV mutated strain. Many of these altered processes are involved in protein production and therefore they might be responsible for recombinant protein final yield. Strain UVM4 showed high abundance of proteins involved in DNA transcription and RNA translation, and low abundance of proteins involved in proteolysis (Figure 5.1). These results provided new insights on potential molecular mechanisms involved in recombinant protein production in *C. reinhardtii* and on enhanced performances of strain UVM4 as a biofactory. However, this Chapter did not investigate the effect of genetic engineering on *C. reinhardtii* and the differences in behaviour between transgenic cell lines of strains 137c and UVM4, which were covered in Chapter 3.

2.2. Comparative proteomic analysis of *C. reinhardtii* strains 137c and UVM4 after genetic engineering (Chapter 3)

Among the altered molecular processes detected by the proteomic analysis performed in Chapter 2, significant differences were observed in relative abundance of proteins involved in chromatin remodelling between strain 137c and strain UVM4; however, to detect and/or analyse pathway

alterations due to genetic engineering, it was necessary to insert and express nuclear transgenes. Therefore, in Chapter 3, I generated transgenic cell lines of both strains 137c and UVM4 and subsequently analysed and compared their intracellular proteomes. The analysis focused on similarities and differences between the proteomes of the two strains after genetic engineering, to unveil the molecular and cellular processes modified after transgene insertion and expression.

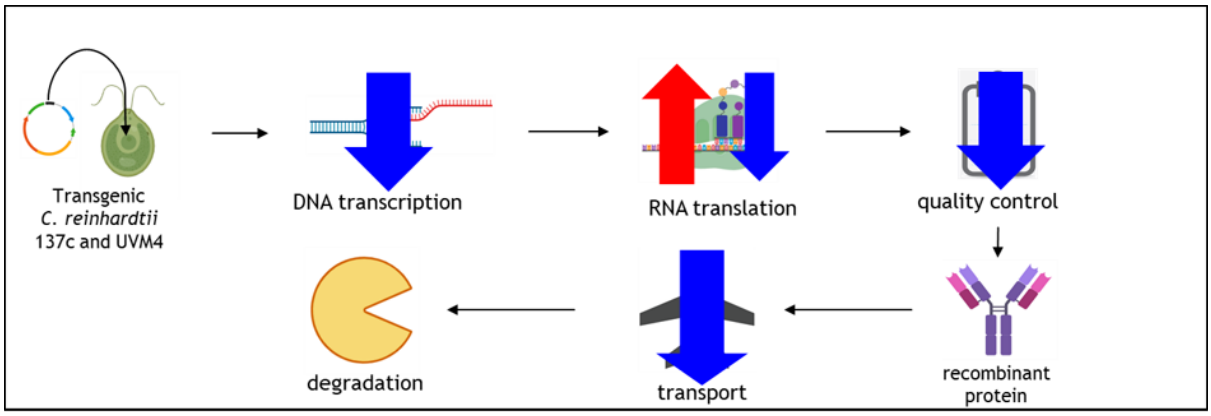


Figure 5.2. Simplified schematic of protein production pathways in *C. reinhardtii*, showing the common pathways altered due to transgene expression in strains 137c and UVM4

These results showed major differences before and after the genetic engineering in both strains. Figure 5.2 shows the similar trends of the altered pathways in both strains after genetic engineering. This analysis reported a general relative higher abundance of ribosomal proteins involved in RNA translation in both strains, probably related to transgene insertion and continuous expression. However, both strain 137c and UVM4 showed a general lower abundance of almost all the other proteins involved in protein production. All proteins involved in DNA transcription, initiation and elongation of translation, quality control, and transport pathways showed lower abundance. This is likely related to the stress of transgenic insertion and expression, the cells of both strains shutting-down cell machinery in order to mitigate energetic costs related to the production an unwanted and unnecessary protein.

Apart from the low yields, another important issue for recombinant protein production in microalgae is possible incorrect glycosylation, which is directly responsible for immunogenicity and consequent low quality of recombinant biopharmaceutical products, addressed in Chapter 4.

2.3. Glycosylation comparison of strains 137c and UVM4 (Chapter 4)

The analysis performed in Chapter 2 & 3 provided useful information on the mechanisms that govern the production of recombinant protein in *C. reinhardtii* and on the altered pathways responsible for enhanced yields in strain UVM4. However, besides quantity, quality of the product is also important for industrial biopharmaceutical production. Therefore, it was necessary to analyse the glycosylation status of strain UVM4 to assess the quality of recombinant proteins produced in this strain. As extensively described in Chapter 1, glycosylation is species-specific and plays a fundamental role in immunogenicity of recombinant biopharmaceuticals; glycans absent in humans but present in recombinant biopharmaceuticals produced in other organisms can be recognised as exogenous and trigger an immunogenic reaction in the patient, leading to reduce efficacy and potential detrimental side effects. Therefore, analysis of glycosylation pattern of a biofactory is necessary to validate its suitability in the industrial biopharmaceutical market. I hypothesised that the engineering and exposure to UV might have affected the glycosylation mechanism in UVM4 resulting in different glycosylation pattern compared to the wild-type. This analysis focused on similarities and differences in glycan abundance between the two strains, with a special focus on immunogenicity.

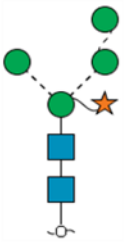
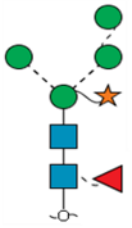
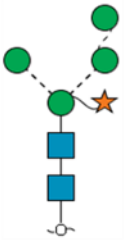
	Glycans showing β(1,2)-xylose	Glycans showing α(1,3)-fucose
Strain 137c		
Strain UVM4		

Figure 5.3. Immunogenic glycans detected in strains 137c and UMV4. Green circle = Mannose; Blue square = *N*-Acetylglucosamine; Red triangle = Fucose; Yellow star = Xylose.

The glycosylation profiles of the two strains showed few similarities and multiple differences. Unfortunately, both strains reported the immunogenic monosaccharides β(1,2)-xylose and α(1,3)-fucose (Figure 3). These immunogenic residues are specific to plant and microalgae. However, the presence of immunogenic native glycoproteins is not necessarily related to production of immunogenic recombinant biopharmaceuticals, as they are not necessarily attached to a recombinant protein during its formation. While this chapter provided the very first comparative glycomics analysis of strain 137c and UVM4, it also highlighted the need for future studies to characterise immunogenicity of recombinant protein produced in *Chlamydomonas reinhardtii*.

3. LIMITATIONS

Chlamydomonas reinhardtii genome was sequenced in 2007 by Merchant and colleagues [3]. Since then, numerous proteins were identified and annotated via experimental analysis, at either the transcriptional or translational level. However, the majority of *C. reinhardtii* proteins are still unpredicted, inferred from homology, or uncharacterised. Moreover, various pathways are incomplete, characterised based on homology with higher plants, or still unknown. Therefore, in Chapters 2 & 3, reconstructing the biological processes with limited information was challenging, and it is likely that important pieces of the puzzle were missing.

The comparative analysis performed in Chapter 2 revealed the potential molecular mechanisms driving the phenotypic differences between two of the most widely used strains of *C. reinhardtii*, strain 137c (wild type strain cc-125 *mt*+) and strain UVM4 (cell wall deficient strain cw15-302 cwd *mt*+ arg7 subjected to genetic engineering and UV mutation), more particularly in the context of recombinant protein production. However, strain 137c is not the direct parental strain of strain UVM4. The parental strain is strain cw15-302 cwd *mt*+ arg7 (also called cc-4350) after nuclear co-transformation of the *ARG7* gene and the *CRY1-1* gene. This genetically modified strain (called Elow47) was generated by Neupert and colleagues in 2009 prior to UV mutagenesis of the microalga [4]. Comparative analysis of strain UVM4 and strain Elow47 might have helped to entirely discern between proteomic profile alterations uniquely generated by UV mutagenesis and alterations due to genetic engineering of strain cc-4350. However, strain cc-4350 is arginine-auxotrophic, therefore it requires specific growing conditions that made impossible a correct proteomic comparison for this PhD thesis. Moreover, both cc-4350 and Elow47 are cell wall deficient strains, therefore a comparative proteomic analysis of strain UVM4 using one of those two strains as control would have not unveil the fundamental molecular processes related to cell wall deficiency, which is supposed to play a fundamental role in the improved potential for recombinant protein production in strain UVM4.

In Chapter 3, both strain 137c and UVM4 were genetically engineered. To overcome the low transformation success rates (common in *C. reinhardtii*), the recombinant protein of interest was fused to mVenus to discriminate positive transformants from others by fluorescence-based cell sorting flow cytometry. This allowed me to select microalgal cultures showing high percentage of cells carrying the recombinant DNA sequence and producing the desired recombinant protein. However, strain UVM4 selected transgenic cell lines showed percentages of positive cells between 18 – 25%. This result confirmed that the high transgene suppression mechanisms reported for *C. reinhardtii* are still active in strain UVM4. Moreover, albeit a fusion fluorescent reporter is very helpful to screen positive cultures and separate them from the rest of the culture, partially enhancing the total amount of positive cells, it cannot be used for industrial application. In addition, a high percentage of positive cells does not directly translate into high yields of recombinant protein. In fact, random nuclear integration transformation mechanism generates very different transgenic cell lines starting from the same wild type culture. Random integration adds another level of complexity to the generation of highly expressive transgenic cell lines in *C. reinhardtii*. Until targeted genetic engineering is perfected in *C. reinhardtii* and microalgae in general, selection of the best transgenic candidates will remain challenging.

In Chapter 2 & 3, the label-free quantification method allowed me to compare the intracellular and extracellular proteomes of strains 137c and UVM4. However, this method is valid only for proteins detected in both strains. The proteins only found in one of the two strains (called unique proteins) were analysed and used as validation of the label-free quantification; however, being absent in the other strain, these proteins could not be quantified. Unique proteins of strain UVM4, as well as absent proteins in this strain, could define the modification of strain UVM4 genome, and also the response of this strain to genetic engineering and its increased transgene expression. Therefore, a different quantification method could have been used to obtain the proteomic datasets. Quantification methods involving an internal or external standard could result in an absolute quantification of these unique proteins, leading to fundamental information regarding strain UVM4 behaviour.

In Chapter 4, I was able to analyse only *N*-glycosylated proteins. Considering the extensive amount of cell wall proteins detected in the extracellular space of UVM4, and the fact that these cell wall proteins (such as hydroxyproline-rich proteins and pherophorins) are heavily *O*-glycosylated, the absence of *O*-glycans in the glyco-profile of the secretome of strain 137c and especially strain UVM4 was surprising. It is plausible that the sample preparation and the detachment of *O*-glycans was inadequate to successfully characterise *O*-glycosylation in *C. reinhardtii*.

As described in Chapter 1 & 4, presence of native immunogenic residues in an organism does not directly related to attachment of immunogenic residues on recombinant glycoproteins. To analyse the glycosylation patterns and potential immunogenicity of recombinant biopharmaceuticals produced in *C. reinhardtii*, I tried to produce two secreted recombinant proteins in strain UVM4: one native (Fe-assimilating protein I, UniProt accession number: Q9LD42) and one of human origin (interferon α2A, *O*-glycosylated human protein with anti-cancer and anti-viral properties, UniProt accession number: P01563). I constructed two plasmids carrying the recombinant DNA encoding for the Fe-assimilating protein 1 and the interferon α2A. Both proteins were carrying a fusion fluorescent tag and a secretion tag, therefore the recombinant proteins would pass through the secretory pathway (enabling glycosylation) and they would be located and detectable in the extracellular space. I performed nuclear transformation and measured fluorescence in the media by high-performance microplate reader to select positive transgenic cell lines for the two constructs. Lastly, I needed to purify the recombinant proteins from the extracellular space to perform glycosylation analysis. Unfortunately, I was not able to purify the proteins from the secretome or even visualize them using western blot (using anti-His antibody). This could be attributed to different reasons: i) the amount of recombinant proteins was too low to be detected; ii) the His-tag was cleaved during secretion, impeding the detection via anti-His antibody (behaviour reported in plants [5]); or iii) the recombinant protein was "trapped" in an extracellular matrix, impeding proper migration on electrophoresis gels, proper binding of the antibody, and successful purification using liquid chromatography methods. Baier and colleagues [6] associated impaired purification of recombinant proteins from the secretome of strain UVM4 with an extensive amount of cell wall protein aggregates populating the extracellular space of this strain. In cell wall-

deficient strains (such as strain UVM4), cell wall proteins are still produced, however they are not assembled in the cell wall and they are secreted in the extracellular space, where they aggregate [6]. In Chapter 2, I analysed the population of cell wall proteins in the secretome of *C. reinhardtii* and compared relative abundances of these proteins in strain UVM4 (against the control strain 137c). I reported an extensive amount of these proteins in the extracellular space and a general higher abundance of cell wall proteins in strain UVM4 (up to 64-fold higher than in 137c). Moreover, in Chapter 4, I analysed and compared the proteomic profile of the extracellular space of both strains using a TGX precast gel and two different protein staining methods, Coomassie staining (to detect the total protein profile) and glyco-staining (to visualise uniquely *N*- and *O*-glycoproteins).

Considering the difficulties I experienced in purification of secreted recombinant proteins from the extracellular space of strain UVM4, I analysed further the protein gels obtained in Chapter 4. In this case, I was looking for intense bands stained with both Coomassie and glyco-staining (as stated before, cell wall proteins are heavily glycosylated) that were only present in strain UVM4.

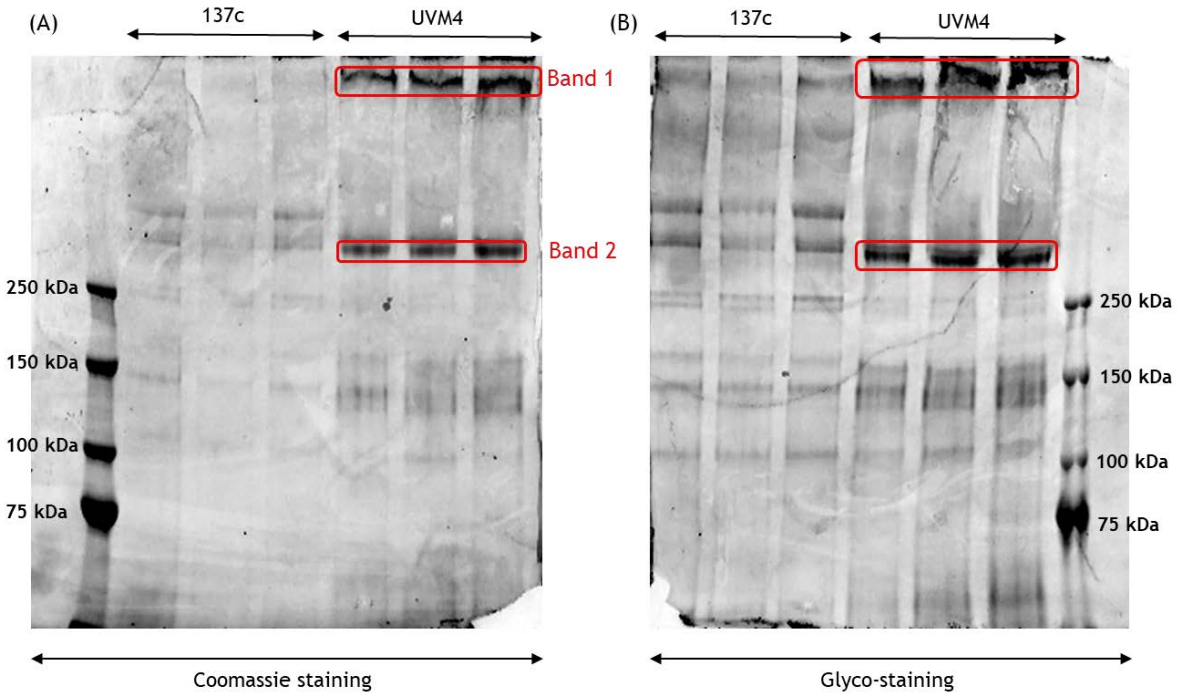


Figure 5.4. SDS-PAGE gel showing the extracellular proteomes of 137c and UVM4 strains stained with (A) Coomassie and (B) glycoprotein staining. Selected bands are highlighted (in red).

Using both staining techniques, I was able to detect two intense glycoprotein bands only present in strain UVM4. The selected bands were extracted (from the gel stained with Coomassie) and the proteins were released from the gel matrix using trypsin digestion. The peptide mixture was subsequently analysed by mass spectrometry to identify the proteins. The identification was performed using the software PEAKS (manually validating the results) (Tables 5.1 and 5.2).

Table 5.1. Identified proteins from strain UVM4 gel band 1

Accession number	Protein	#Peptides	#UniqueP	Avg. Mass (Da)
A8HP37	Cell wall protein pherophorin-C13	66	18	52902
A8HP36	Cell wall protein pherophorin-C2	55	7	52642
A8JJQ2	Cell wall protein pherophorin-C1	26	15	50057
Q6PLP6	Cell wall protein GP2 (Fragment)	20	20	120312
Q3HTK4	Cell wall protein pherophorin-C3	14	14	47118

Table 5.2. Identified proteins from strain UVM4 gel band 2

Accession number	Protein	#Peptides	#UniqueP	Avg. Mass (Da)
Q6PLP6	Cell wall protein GP2 (Fragment)	86	86	120312
A8HP36	Cell wall protein pherophorin-C2	13	3	52642

This analysis showed multiple interesting results: i) both band 1 and band 2 showed multiple protein identifications, ii) the proteins found in band 2 are also present in band 1, and iii) all the proteins identified are cell wall proteins. Moreover, all proteins reported a high number of identified peptides, repeatedly showing multiple hits for the same peptide (often unique peptides were identified multiple times). Furthermore, both bands on the gel were showing molecular weights higher than 250 kDa, whilst all the identified proteins have molecular weights equal or lower than 120 kDa. I hypothesise that the aggregates described by Baier and colleagues [6] are formed by multiple different cell wall proteins, and that these aggregates are not separated in mild denaturing conditions. I also identified the proteins possibly responsible for those aggregates that impair recombinant protein purification in strain UVM4 and consequently responsible for lower yields.

4. FUTURE PERSPECTIVES AND CONCLUDING REMARKS

As stated before, there are still some challenges hindering *C. reinhardtii* to become a commercially established biofactory. Here I provide multiple molecular targets possibly responsible for low recombinant protein yields in this species. Future research should apply the information gathered in this study to improve yields in *C. reinhardtii*.

In this thesis, I mainly used correlative approaches to identify possible protein targets responsible for low recombinant protein yields in *C. reinhardtii*. Combining existing protein annotation, homology with higher eukaryotes, and literature, I was able to speculate on the role of multiple proteins and their involvement in specific molecular processes. However, to fully demonstrate my hypotheses, functional genomics approaches are necessary. Targeted knock-in (KI), knock-down (KD), and/or knock-out (KO) of gene(s) encoding for proteins identified in this study, and subsequent analysis of the transgenic cell line generated, could possibly validate the role(s) hypothesised. The application of functional genomics based on this study could i) improve the currently inadequate molecular knowledge of *C. reinhardtii*, and ii) enhance recombinant protein production and yields in this species, to validate *C. reinhardtii* as a cell biofactory for industrial applications. For example, in Chapter 3, I reported lower abundance of proteins involved in translation initiation and elongation after genetic engineering in both *C. reinhardtii* strains; I hypothesised that lower abundance of these proteins might be directly related to lower transgene expression in *C. reinhardtii*. Overexpressing genes involved in these pathways by genetic engineering might result in higher translation of recombinant protein and consequently higher yields, validating my speculation. Other possible targets for KI, KD, or KO approaches are gene(s) encoding for proteins involved in chromatin remodelling. In Chapter 2 & 3, I reported a higher abundance of histones in strain UVM4 compared to strain 137c, before and after transgenic insertion and expression. Moreover, in Chapter 3, I detected lower abundance of proteins involved in histone modification and chromatin remodelling in both strains after genetic engineering. Differential abundance of these proteins might be linked to the position effects and gene silencing mechanisms that afflict strain UVM4. Unfortunately, the mechanisms regulating core histone modifications and gene silencing in *C.*

reinhardtii are poorly described and will gain to be studied further. Unveiling the mechanisms regulating histone-DNA binding and histone modifications in *C. reinhardtii* might help understanding their role in gene and transgene expression and consequently in recombinant protein yields. Another important altered molecular process reported in strain UVM4 was lysosome activity. Lysosome is a membrane-bound organelle involved in protein degradation. The mechanisms regulating the activity of this organelle, and the interaction between intracellular and extracellular lysosome enzymes are still unknown. It is crucial to extend our general understanding of lysosome activity in *C. reinhardtii*, to elucidate protein degradation mechanisms in *C. reinhardtii*, and possibly to enhance recombinant protein final yields. Another functional genomics application might target extracellular cell wall proteins. Here I provide a list of secreted cell wall proteins possibly involved in impaired recombinant protein purification. Future research should target these cell wall proteins, perform gene KD or KO to impede their production, and subsequently produce secreted recombinant proteins in these cell wall-protein-free cell lines. Recombinant protein purification in these KD or KO cell lines might improve, leading to enhanced secreted protein yields.

However, there are two important aspects to consider before undertaking a genetic engineering approach: i) modifying existing pathways might have detrimental effects on survival and growth of the organism, and ii) multiple target proteins involved in similar pathways were reported, and overexpressing them all might be impractical and/or counterproductive. A possible solution is a trial-and-error approach, selecting and overexpressing a specific gene encoding for one target protein and report the transgenic cell line behaviour.

Future research should focus on producing recombinant secreted proteins and analyse the intracellular and especially the extracellular proteomes. The journey of a recombinant protein through the secretory pathway will definitely alter the proteomic profile of *C. reinhardtii* differently than for a recombinant protein produced intracellularly and accumulating in the cytosol (in reference to the recombinant mVenus protein used in this thesis, Chapter 3). The analysis of the extracellular proteome of a strain

UVM4 transgenic cell line producing a secreted protein could also bring significant information on lysosome activity after genetic engineering.

In Chapter 4, I tried to analyse both *N*- and *O*-glycosylation, however only *N*-glycan patterns were detected and identified. I hypothesised (in Chapter 5, section *3*) that sample preparation played a major role for the failure of *O*-glycosylation analysis in my project. Nevertheless, I think that a comprehensive analysis of native *O*-glycans in *C. reinhardtii* and in microalgae in general is urgently needed. To achieve successful *O*-glycosylation analysis, I suggest to separate the *O*-glycoproteins from the rest of the sample mixture prior to the detachment of the glycans. A possible procedure might involve the use of lectins, carbohydrate-binding proteins that can target uniquely *O*-glycosylation [7].

In Chapter 4, I analysed native *N*-glycosylation patterns in *C. reinhardtii*, reporting multiple immunogenic glycans. However, because of some technical and time limitations, I was not able to purify secreted recombinant proteins and analyse glycosylation of recombinant proteins. The capacity of a cell biofactory to attach immunogenic residues on native proteins does not directly correlate to presence of immunogenic glycans on recombinant proteins. Therefore, targeted analysis of recombinant glycosylation patterns from *C. reinhardtii* is urgently needed to validate *C. reinhardtii* as a cell biofactory for industrial applications.

Finally, while my research unveiled the limitations of *Chlamydomonas reinhardtii* as a cell biofactory, multiple solutions are clearly available to overcome these obstacles and there is no doubt that *C. reinhardtii* will increasingly play significant role in the biopharmaceutical industry in the near future.

5. REFERENCES

1. Lauersen, K.J.; Huber, I.; Wichmann, J.; Baier, T.; Leiter, A.; Gaukel, V.; Kartushin, V.; Rattenholl, A.; Steinweg, C.; von Riesen, L., et al. Investigating the dynamics of recombinant protein secretion from a microalgal host. *J Biotechnol* **2015**, *215*, 62-71, doi:10.1016/j.jbiotec.2015.05.001.

2. Ramos-Martinez, E.M.; Fimognari, L.; Sakuragi, Y. High-yield secretion of recombinant proteins from the microalga *Chlamydomonas reinhardtii*. *Plant Biotechnol J* **2017**, *15*, 1214-1224, doi:10.1111/pbi.12710.

3. Merchant, S.S.; Prochnik, S.E.; Vallon, O.; Harris, E.H.; Karpowicz, S.J.; Witman, G.B.; Terry, A.; Salamov, A.; Fritz-Laylin, L.K.; Marechal-Drouard, L., et al. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **2007**, *318*, 245-250, doi:10.1126/science.1143609.

4. Neupert, J.; Karcher, D.; Bock, R. Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J* **2009**, *57*, 1140-1150, doi:10.1111/j.1365-313X.2008.03746.x.

5. Zhang, N.; Wright, T.; Caraway, P.; Xu, J. Enhanced secretion of human alpha1-antitrypsin expressed with a novel glycosylation module in tobacco BY-2 cell culture. *Bioengineered* **2019**, *10*, 87-97, doi:10.1080/21655979.2019.1604037.

6. Baier, T.; Kros, D.; Feiner, R.C.; Lauersen, K.J.; Muller, K.M.; Kruse, O. Engineered Fusion Proteins for Efficient Protein Secretion and Purification of a Human Growth Factor from the Green Microalga *Chlamydomonas reinhardtii*. *ACS Synth Biol* **2018**, 10.1021/acssynbio.8b00226, doi:10.1021/acssynbio.8b00226.

7. Poiroux, G.; Barre, A.; van Damme, E.J.M.; Benoist, H.; Rouge, P. Plant Lectins Targeting O-Glycans at the Cell Surface as Tools for Cancer Diagnosis, Prognosis and Therapy. *Int J Mol Sci* **2017**, *18*, doi:10.3390/ijms18061232.