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.