UNIVERSITAT DE
BARCELONA

# Strategies to improve *Chlamydomonas reinhardtii* as a recombinant protein host: from a small growth factor to a complex monoclonal antibody production

Maria Teresa López Cerro

**ADVERTIMENT**. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA**. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING**. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

UNIVERSITAT DE BARCELONA

# Strategies to improve *Chlamydomonas reinhardtii* as a recombinant protein host: from a small growth factor to a complex monoclonal antibody production

This dissertation is submitted for the degree of Doctor of Philosophy at the University of Barcelona

Maria Teresa López Cerro

May 2018

| Director: | Director: | Supervisor: |
|---|---|---|
| Jaume Mercadé Roca | Cristina López Paz | Josefa Badia Palacín |

Greenaltech
Microalgae Technologies

# *Abstract*

*Chlamydomonas reinhardtii* has emerged as a promising alternative host for recombinant protein expression. Despite its advantageous characteristics and low-cost production, its use is hampered by low expression levels of nuclear transgenes. In this thesis we test several strategies designed to reduce or overcome this limitation. As a result, on the base of a secreted fusion protein comprising a small growth factor and a reporter, the use of regulatory and stabilizing regions resulted in expression levels ranging from 1 to 100 µg /L of culture. We report the expression of a fully-assembled monoclonal antibody in Chlamydomonas nucleus, therefore, validating Chlamydomonas as a host for complex protein production. The cassettes and high throughput screenings developed emerge as innovative tools expanding the molecular toolbox available for Chlamydomonas nucleus. In addition, a scalable purification method to recover the target protein from culture medium has been developed and validated indicating a simple downstream processing for secreted recombinant protein production. Finally, we report that Chlamydomonas secreted components induce proliferation of murine fibroblasts and have a synergistic effect with supplied hEGF, unveiling the potential of extracellular components of Chlamydomonas for a variety of applications.

# *Acknowledgements*

I would like to be able to say that writing this thesis was not that tough, but I can't. This thesis would have not been possible without all the help and support I received and I am deeply grateful.

First, I want to thank Olga, Jaume and Jordi for creating Greenaltech, devoted to such a difficult activity as it is research. Greenaltech has been a great adventure from the very beginning and a unique environment for my personal and professional development. Thanks to Olga, Jaume and Jordi for trusting me from the beginning and thanks to Joan for understanding all the implications of this project. I would like to specially express my gratitude to Jaume, who has lead this period of my professional life with his way of thinking, scientific advice and the freedom environment that he always promotes.

This thesis would have not been possible without the expert advice of Cristina and I will be grateful forever and ever to her. With her incredible analytic capacity and knowledge, she made possible a lot of challenging work. Without her, the learning I have gained during this period would not have reached the same level. Thank you, Cristina, for teaching me that we can do incredibly good things if we make our best.

This project started with Derek, who showed me the first steps in Greenaltech and in this project. I am thankful to him and Eugènia for making the long working days fun and inspiring. I am deeply grateful because we have been for a long time in the same boat, being always supportive to each other in all aspects. Without them this adventure would have not been the same. I don't have words to express my joy, not only for having their scientific advice and for all the days we shared, but also for being today in the position to call them friends.

I would like to say thanks to my colleagues from Greenaltech for always being supportive and for helping me in everything. Thank you very much Ariadna, Esther, Rosa and Xavi. I have to say thanks to all the members (past and present) of the Molecular Biology department for providing an unbeatable friendly environment that made everything easier: Griselda, Pau, Borja, Noelia and Tania. Thanks to Dani for being always so supportive and for buying my happiness (in the bad days) with smiles and delicious home-made food. Ariadna deserves a special mention because she, with her never-ending smile, made possible part of this thesis with a superb working capacity the past year.

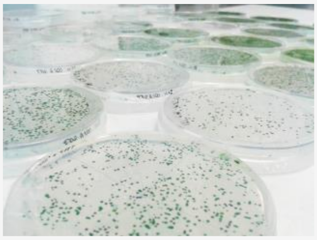
I need to say thanks to Leitat Biomed team for our collaborations, including the RETOS project we shared. Thanks for your scientific advice, response capacity and willingness to develop new things.

I want to thank also Pepita Badia for her help and willingness to make everything easier.

In my personal life I have to thank my friends and family for understanding my absences, even when nobody knows exactly what I am still studying 10 years after starting my degree.

It is impossible to express the gratitude I feel to my parents for their unconditional support, encouragement and love, even when my objectives keep me away from home for months. I owe everything to them for all they have done, and still do, for me. Thanks also to my brother for understanding that “his sister sometimes exists and sometimes does not” and for helping me to imagine all the things we will do when I have free time.

Finally, thanks to Marc, because without him I am not sure if I would have ever completed my thesis. In a tough period for both, he has been my biggest support. For always getting the best out of me and always giving me the best of you, thank you.



# Contents

# Contents

List of abbreviations .......... i

List of figures .......... iii

List of tables .......... ix

Introduction .......... 1

Objectives .......... 19

CHAPTER 1. Improving expression levels from Chlamydomonas nuclear genome .......... 21

- Introduction .......... 23
- Results .......... 33
  - Strategy 1. Design of DNA cassette and optimization of a high-throughput screening method to obtain a high RP expressing transformant .......... 33
    1. Developing a high throughput screening method: fusion of a selectable marker gene and the protein of interest .......... 34
      - 1.1. Use of polycistronic mRNAs: analysis of two different viral 2A sequences .......... 34
        - 1.1.1. Use of minimal FMDV 2A peptide for processing of a dicistronic vector .......... 34
        - 1.1.2. Use of extended FMDV 2A peptide for processing of a dicistronic vector .......... 41
      - 1.2. Use of a fusion protein consisting of a resistance gene and the protein of interest linked by TEV protease recognition sequence .......... 45

            - 1.2.1. Identification of a highly expressed chimeric protein resulting from the partial insertion of exogenous DNA cassette........................................................................ 54
    - 2. Developing a high throughput screening method: fusion of luciferase reporter to the protein of interest................................................. 59
        - 2.1. Fusion of 3XHA tagged Gaussia Luciferase reporter to the protein of interest to enable screening of transformants..... 59
        - 2.2. Comparison of the efficiency of two different cis regulatory regions to drive transgene expression ................................. 67
        - 2.3. Comparison of the effect of several secretion peptides on expression yields................................................................. 69
        - 2.4. Use of glycomodules to enhance expression and secretion of nuclear transgenes............................................................... 76
        - 2.5. Introduction of an intronic sequence within a reporter gene to enhance nuclear expression ................................................ 86
        - 2.6. Characterization and quantification of transgene expression during cell culture growth................................................... 88
- Strategy 2. Study of the effect of transgene copy number on nuclear transgene expression levels..................................................................97
    - 1. Analysis of recombinant protein expression in diploid cells ........... 97
    - 2. Study of the relation between transgene copy number and recombinant protein expression in haploid cells......................... 102
- Strategy 3. Obtaining a Chlamydomonas strain with improved expression of nuclear transgenes ........................................................................105

Discussion ..........................................................................................117

Methodology .........................................................................................129

CHAPTER 2. Purification of human Epidermal Growth Factor produced in Chlamydomonas ........................................................................................147
Introduction........................................................................................149
Results ........................................................................................153
1. Purification of different hEGF isoforms ........................................153
1.1. Development of a purification procedure at lab scale............. 153
1.2 Purification and processing of different hEGF isoforms ........... 155
1.3 IEX purification of processed hEGF .......................................... 158
2. hEGF activity assessment of isoforms...........................................161
2.1 Development of *In vitro* activity assay conditions ................... 161
2.2 *In vitro* activity assessment of partially purified hEGF isoforms 164
Discussion........................................................................................169
Methodology ........................................................................................175
CHAPTER 3. Expression of a monoclonal antibody from Chlamydomonas nuclear genome........................................................................................181
Introduction........................................................................................183
Results ........................................................................................191
1. Cassette design........................................................................191
2. Screening of transformants ........................................................193
2.1 Initial screening of transformants............................................ 193
2.2 Retransformation and screening of transformants ................. 202
2.3 mAb expression characterization in selected transformants.... 206
3. *In vitro* mAb characterization ..........................................................210

3.1 mAb evaluation by immunoblot in reducing and non-reducing conditions ........................................................................ 210

3.2 Evaluation of recognition of Ag by ELISA ................................ 211

Discussion ......................................................................................213

Methodology ..................................................................................219

Discussion and future prospects..................................................................225

Conclusions..........................................................................................233

Appendix 1..........................................................................................239

Recombinant protein expression and purification from *E.coli* to allow quantification based on densitometry of immunoblots ........................239

Bibliography.........................................................................................243

# *List of abbreviations*

- **Ag:** antigen
- **AGPs:** Arabinogalactan protein
- **AR:** *HSP70/RBCS2* chimeric promoter
- **shBle:** protein conferring resistance to bleomycin family of antibiotics
- **DDT:** Dithiothreitol
- **DMEM:** Dulbecco's Modified Eagle's medium
- **DMSO:** dimethylsulfoxide
- **ELISA:** Enzyme-Linked Immunosorbent Assay
- **FBS:** fetal bovine serum
- **FCS:** fetal calf serum
- **FMDV 2A:** food and mouth disease virus 2A peptides
- **FT:** flow through
- **gLuc:** Gaussia princeps Luciferase (protein)
- **GOI:** gene of interest
- **HC:** heavy chain (of an antibody)
- **hEGF:** Epidermal growth factor (protein)
- **HRGP:** Hydroxyproline-rich glycoprotein
- **Hyp:** hydroxyproline
- **IMAC:** immobilized metal affinity chromatography
- **LC:** light chain
- **mAb:** monoclonal antibody
- **MAR:** matrix attachment region
- **MBS:** mean background signal
- **MT:** mating type
- **MW:** molecular weight

- **OD:** optical density
- **PRPs:** proline rich protein
- **PTM:** post-translational modification
- **RP:** recombinant protein
- **RT:** room temperature
- **TFF:** tangential flow filtration
- **TRXH:** thioredoxin H1 (protein)
- **TSP:** total soluble protein

# *List of figures*

*Figure 1*. Schematics of FMDV 2A function ................................................ 35

*Figure 2.* Schematics of pARshBle2AhEGF, pARshBle2ATRXH and pARshBle2AmCherry constructs ................................................................. 36

*Figure 3*. Zeocin sensitivity test of pARshBle2AhEGF and pARshBle2ATRXH transformants ........................................................................................ 37

*Figure 4.* Immunoblot of pARshBle2AhEGF and pARshBle2ATRXH high resistant transformants ........................................................................................ 38

*Figure 5.* Immunoblot of pARshBle2AhEGF and pARshBle2ATRXH selected transformants at different culture stages.................................................... 39

*Figure 6*. Immunoblot of pARshBlemCherry selected transformants ............ 41

*Figure 7*. Schematic comparison of minimal FMDV 2A (2A) and extended FMDV 2A ($2A^E$) sequences and schematics of pARshBle2A$^E$hEGF and pARshBle2A$^E$TRXH constructs ..................................................................... 42

*Figure 8.* Zeocin sensitivity test of pARshBle2A$^E$hEGF and pARshBle2A$^E$TRXH transformants ........................................................................................ 43

*Figure 9.* Immunoblot of pARshBle2A$^E$hEGF and pARshBle2A$^E$TRXH selected transformants ........................................................................................ 44

*Figure 10.* Schematics of pPLshBleTEVhEGF construct ................................. 46

*Figure 11*. Zeocin sensitivity test of pPLshBleTEVhEGF transformants .......... 47

*Figure 12.* Immunoblot of pPLshBleTEVhEGF selected high expressing transformants ........................................................................................ 48

*Figure. 13* Immunoblot of pPLshBleTEVhEGF selected transformants at different growth stages ....................................................................... 49

*Figure 14.* PCR analysis of the integration of the cassette in selected high resistant pPLshBleTEVhEGF transformants .................................................. 50

*Figure 15*. Analysis of pPLshBleTEVhEGF 21gr transformants ....................... 52

*Figure 16.* Immunoblot used to estimate expression levels by densitometry in selected transformants........................................................................ 53

*Figure 17*. Immunoblot of hEGF fusion protein obtained transformant......... 54

*Figure 18.* Immunoprecipitation of shBle-X chimeric protein........................ 55

*Figure 19.* Schematics of fusion protein (designed) and chimeric protein found in pPLshBleTEVhEGF#16. .................................................................... 55

*Figure 20.* Analysis of the distribution of identified peptides along the five Ble:unknown putative candidates ................................................................ 57

*Figure 21.* PCR analysis to determine the sequence transcribing for the chimeric protein expressed in clone pPLshBleTEVhEGF#16........................................ 58

*Figure 22*. Schematics of pAR-ARSss-glucTEVhEGF-HygB$^R$ construct............. 60

*Figure 23.* gLuc activity test results of pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ initial transformants............................................................................ 61

*Figure 24.* gLuc activity assay results for the selected pAR-ARSss-glucTEVhEGF-HygB$^R$ transformants .......................................................................... 62

*Figure 25.* Expression analysis of pAR-ARSss-glucTEVhEGF-HygB$^R$ selected transformants by immunoblot..................................................................... 63

*Figure 26.* Immunoblots of intracellular fraction of transformant pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ #H2.5........................................................ 64

*Figure 27*. Expression analysis of pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ #H2.5 selected transformant by gLuc activity test at different growth stages. .................................................................................................................. 65

*Figure 28.* Immunoblot of pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ #H2.5 transformant .......................................................................................... 66

*Figure 29*. Schematics of pPL-ARSss-gLucTEVhEGF-HygB$^R$ construct............. 67

*Figure 30.* gLuc activity results for pAR-ARSss-gLucTEVhEGF (pAR) and pPL-ARss-gLucTEVhEGF (pPL) transformants...................................................... 68

*Figure 31.* Immunoblot of pAR-ARSss-gLucTEVhEGF-hygB$^R$ (pAR) and pPL-ARSss-gLucTEVhEGF-hygB$^R$ (pPL) highest expressing transformants of CC-124 strain.......................................................................................................... 69

*Figure 32.* Schematics of pPL-GAMss-gLucTEVhEGF-paro$^R$ (A) and pPL-ARSss-gLucTEVhEGF-paro$^R$ (B) vectors.................................................................... 70

*Figure 33.* Distribution of normalized luciferase expression (RLU) values from Chlamydomonas strains UVM4 (A) and CC-124 (B) transformed with vectors pPL-GAMss-gLucTEVhEGF (GAMss) and pPL-ARSss-gLucTEVhEGF (ARSss)..... 72

*Figure 34.* Schematics of pPL-CAHss-gLucTEVhEGF-paro$^R$ vector.................. 73

*Figure 35.* Distribution of normalized luciferase expression (RLU) values from CC-124 Chlamydomonas strain transformed with vectors pPL-GAMss-gLucTEVhEGF (GAMss), pPL-ARSss-gLucTEVhEGF (ARSss) and pPL-CAHss-gLucTEVhEGF (CAHss)......................................................................... 74

*Figure 36.* Secretion efficiency analysis of GAMss, ARSss and CAHss secretion signals by gluciferase activity...................................................................... 75

*Figure 37*. Distribution of normalized luciferase expression (RLU) values from Chlamydomonas strains UVM4, UVM11 and CC-1883 transformed with constructs pPL-ARSss-glucTEVhEGF-hygB$^R$, pPL-ARSss-glucTEVhEGF-(SP)$_{10}$-hygB$^R$ and pPL-ARSss-glucTEVhEGF-(SP)$_{20}$-hygB$^R$.......................................... 78

*Figure 38.* Schematics of cassettes used to test the selected glycomodules.. 80

*Figure 39.* Distribution of normalized luciferase expression (RLU) values from Chlamydomonas strains CC-124 (A) and UVM4 (B) transformed with constructs illustrated in figure 38 ......................................................................... 82

*Figure 40.* Immunoblot of selected high expressing transformants of fusion proteins comprising a glycomodule sequence ............................................. 85

*Figure 41*. Schematics of pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF vector .... 86

*Figure 42.* Distribution of normalized luciferase expression (RLU) values from Chlamydomonas strains CC-124 and UVM4 transformed with pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF vector........................................................... 87

*Figure 43.* Immunoblot of pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF highest expressing transformants ........ 88

*Figure 44.* Expression analysis of selected transformants by gLuc activity test at different growth stages under different growth conditions ........ 91

*Figure 45.* Immunoblot analysis of selected transformants at different growth stages under different growth conditions ........ 93

*Figure 46.* Immunoblots used to estimate quantification of expression from CC-124 (A), UVM4 (B) transformants and (C) UVM4#35 transformant........ 95

*Figure 47*. Schematics of vectors used to study the effect of transgene copy number on Chlamydomonas nuclear transgene expression levels........ 98

*Figure 48.* Schematics of crossings performed to obtain diploid progeny...... 99

*Figure 49.* Confirmation by PCR of the diploid nature of the selected progeny. ........ 100

*Figure 50*. Expression analysis of diploid cells by gLuc activity at different growth stages........ 101

*Figure 51.* Luciferase activity assessment of gene copy number effect on RP expression from Chlamydomonas nucleus ........ 103

*Figure 52*. Schematic diagram of the experimental proposal to screen high expressing strains. ........ 105

*Figure 53.* Schematics of pAR-his-shBle vector ........ 106

*Figure 54.* Zeocin sensitivity test and immunoblot of selected strains for mutagenesis ........ 107

*Figure 55.* Viability of mutagenized cells with increasing irradiation. ........ 107

*Figure 56.* Zeocin sensitivity test and immunoblot analysis of mutants obtained from strain shBlelowA ........ 109

*Figure 57*. Zeocin sensitivity test and immunoblot analysis of mutants obtained from strain shBlelowB........ 110

*Figure 58.* Transformation efficiency (number of colonies obtained/µg transformed DNA) fold increase of the mutants referred to CC-124 ........ 112

*Figure 59*. Comparison of pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF highest expressing transformants from CC-124, mutB12 and mutB13 strains ......... 114

*Figure 60*. Resulting progeny from a cross between a high expressing strain and gLuc wild type transformants ..................................................................... 115

*Figure 61.* Schematic representation of initial purification proposal........... 153

*Figure 62.* Immunoblot assessment of IMAC purification with different resins and sample conditions......................................................................... 154

*Figure 63.* Analysis of RP purification from selected transformants............ 157

*Figure 64*. Schematic representation of improved purification proposal. .... 158

*Figure 65.* Immunoblot assessment of RP IMAC purification from CC-124#3 transformant ........................................................................................ 159

*Figure 66*. IEX purification of RP from CC-124#3 transformant ................... 160

*Figure 67.* hEGF proliferation induction in NIH 3T3 cells in complete medium (CM) and basal medium (BM) by MTT test. ............................................... 162

*Figure 68.* hEGF proliferation induction in NIH 3T3 cells under different medium by MTT test......................................................................................... 163

*Figure 69.* hEGF proliferation induction assessment in NIH3T3 cells in basal medium (BM) + 0.5% FCS ........................................................................ 163

*Figure 70.* MTT proliferation induction assessment of hEGF isoforms expressed in Chlamydomonas ............................................................................... 165

*Figure 71*. Proliferation induction assessment of Chlamydomonas extracellular fractions in presence or absence of hEGF by MTT assay ............................ 167

*Figure 72.* Schematic representation of the vector used for the expression of m5C3 antibody. ..................................................................................... 192

*Figure 73.* High throughput screening designed to select Chlamydomonas transformants expressing a fully-assembled mAb ...................................... 194

*Figure 74.* Sandwich ELISA standard curves in PBS, in culture supernatant from CC-124 and in total cell extract from CC-124. ............................................ 195

*Figure 75*. Sandwich ELISA screening of CC-124 transformants .................. 196

*Figure 76.* Sandwich ELISA analysis of intracellular and extracellular mAb expression in 19A3 transformant ............................................................ 197

*Figure 77.* Kinetics of mAb expression in 19A3 transformant analysed by sandwich ELISA ........................................................................................ 197

*Figure 78*. Analysis of mAb expressed in 19A3 clone after concentration by filtration ........................................................................................................ 199

*Figure 79.* Sandwich ELISA screening of UVM4 transformants. .................... 200

*Figure 80.* Immunoblot evaluation of mAb expressed in UVM4 selected transformants ............................................................................................... 201

*Figure 81.* Immunoblot evaluation of mAb expressed by UVM4-2 transformant. ..................................................................................................................... 201

*Figure 82.* HC-paro$^R$ vector map ................................................................ 202

*Figure 83*. Sandwich ELISA screening of 19A3 transformants...................... 203

*Figure 84.* Sandwich ELISA confirmation of 19A3 selected transformants ... 203

*Figure 85.* Sandwich ELISA quantification of intracellular and extracellular mAb expression in CC-124 selected clones......................................................... 204

*Figure 86*. Immunoblot evaluation of mAb expressed in 19A3-1 and 19A3-6 transformants and comparison to its background strain 19A3 ................... 205

*Figure 87*. Kinetics of growth in transformants expressing m5C3. ............... 206

*Figure 88*. Kinetics of mAb expression analysed by sandwich ELISA in selected transformants. ........................................................................................... 207

*Figure 89.* Immunoblot evaluation of mAb expression in selected transformants at different culture stages ........................................................................ 209

*Figure 90.* Immunoblot evaluation of concentrated mAb by TFF ................ 210

*Figure 91.* Analysis of human S100A4 recognition by ELISA ........................ 211

# *List of tables*

*Table 1*. Current expression systems available ............................................... 6

*Table 2.* Most relevant alternative hosts for recombinant protein production 9

*Table 3*. Microalgae used for recombinant protein expression. .................... 12

*Table 4*. Characteristics of recombinant protein expression in Chlamydomonas nucleus or chloroplast.. ........................................................................ 14

*Table 5.* Estimation of expression in selected transformants ........................ 53

*Table 6*. Candidates to be the unknown part of the chimeric protein expressed by the transformant pPLshBleTEVhEGF#16. ................................................ 56

*Table 7*. pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ H2.5 transformant expression level estimation by densitometry. .............................................. 64

*Table 8.* Chlamydomonas native glycomodules selected to test its effect on RP expression.......................................................................................... 79

*Table 9.* Densitometry based quantification results in selected transformants. .......................................................................................................... 96

*Table 10.* P-values resulting of the statistical analysis to determine candidates to be a high expressing strain ..................................................................... 113

*Table 11.* Sequences of oligonucleotides used for vectors construction in this study. ........................................................................................... 130

*Table 12.* Clonings performed for the present study. ................................. 131

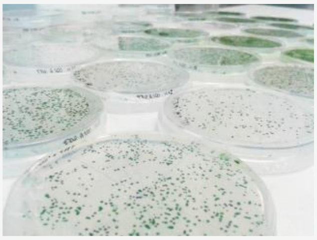
*Table 13.* Sequences of oligonucleotides used for PCR amplification in this study. ........................................................................................... 139

*Table 14.* Sequences of oligonucleotides used for mating type determination by PCR amplification in this study ................................................................. 140

*Table 15.* Sandwich ELISA quantification of extracellular ( E ) and intracellular ( I ) mAb expression in selected transformants. ............................................ 204

*Table 16*. Sandwich ELISA quantification of mAb expression in selected transformants at different culture stages.................................................... 208

*Table 17.* Sequences of oligonucleotides used for vectors construction in this study. ........................................................................................... 220



# Introduction

# *Introduction*

## *The need for recombinant proteins*

Recombinant proteins (RP) are essential in a variety of applications in cosmetic, industrial, veterinary and therapeutic markets. The expression of RPs in heterologous hosts is a core technique in biotechnology, however, the transfer to a suitable expression system is not always easy (Ferrer-Miralles et al. 2009; Corchero et al. 2013). The therapeutic market, although it is not the only one, is the largest market for recombinant technology (market value of biopharmaceuticals reached a total cumulative sales value of $140 billion for 2013 (Walsh G 2014)). Many therapeutic RPs today come from the culture of transgenic mammalian cells in fermentation facilities (Walsh G 2014). Due to media costs and the inherent risks and complexity of these production systems, RPs produced in this manner are very expensive (Heitzer et al. 2007; Corchero et al. 2013). Thus, there is a challenge in improving the expression hosts towards a cheaper, safe, effective and scalable heterologous platform allowing to address the demands for a variety of applications in diverse markets such as cosmetics, industry or veterinary (Fletcher et al. 2007; Ferrer-Miralles et al. 2009).

## *Available expression hosts*

Current RP expression hosts include bacteria, yeasts, insect cells, plants, mammalian cells and transgenic animals, each of these production platforms having its advantages and disadvantages (Demain and Vaishnav 2009; Ferrer-Miralles et al. 2009; Corchero et al. 2013). None of the systems can be considered superior to any other (see Table 1 for a comparison) and for each individual product and application the most suitable expression system has to be selected and optimized (Schmidt 2004). Functionality, production speed, production cost and yield are the key factors to consider when choosing RP expression host (Demain and Vaishnav 2009).

Bacteria are the first option as RP host since they can produce RPs at low cost and quickly (Demain and Vaishnav 2009). Among bacteria, *E.coli* is the preferential host due to a vast molecular toolbox available, rapid growth and simple cultivation techniques that allow high density cultures (Ferrer-Miralles et al. 2009; Jia and Jeon 2016). Thanks to its low-cost production and high yields, RP produced in bacteria are marketed for a wide range of applications in all the markets impacted for recombinant technologies including cosmetics, industry, research and therapeutics. However, *E.coli* is generally best suited for the production of small and structurally simple proteins because it does not perform post-translational modifications (PTMs), such as disulphide bond formation and glycosylation, and is inefficient at protein folding and secretion. Furthermore, complex proteins can often end up in insoluble inclusion bodies requiring a costly downstream processing (Demain and Vaishnav 2009; Ferrer-Miralles et al. 2009; Corchero et al. 2013). In addition, endotoxin level must be controlled in the final product.

In a comparable way to bacteria, yeasts can quickly produce low-cost RP trough an easy genetic manipulation. High-density yeast cultures can be reached in bioreactors offering generally higher yields of RPs and better scalability than insect or mammalian cells (Demain and Vaishnav 2009; Corchero et al. 2013). Yeasts can secrete RP and are able to perform many PTMs. Thus, RP production in yeast is usually approached when the production of the target protein is not feasible in bacteria due to the requirement of PTM (Porro et al. 2011). However, even though yeasts are able to perform many PTMs those are different from mammalian ones. The main limitation of this expression system is related to N-linked glycosylation patterns in which sugar side chains of high mannose content affect the serum half-life and immunogenicity of the final product (Gemmill and Trimble 1999; Demain and Vaishnav 2009; Ferrer-Miralles et al. 2009). However, the methylotrophic yeast *Pichia pastoris* has been genetically engineered to produce a human type of glycosylation (De Pourcq et al. 2010). RP produced in *Saccharomyces cerevisiae* and *Pichia pastoris* among other yeasts are marketed for a range of applications including therapeutics (Nielsen 2013).

Another available expression system are insect cells infected with baculovirus. This system holds an important potential because offers eukaryotic processing of complex proteins and a similar PTM pattern to mammalian cells at high yields and at a lower cost than mammalian cells. However, insect PTM pattern represents a limitation to the use of this expression system because the differences with the mammalian PTM affect its immunogenicity and pharmacokinetics (Sinclair and Elliott 2005). RP produced in insect cells have much simpler side chains than mammalian glycoproteins which often have complex type N-glycans with terminal sialic acids (Shi and Jarvis 2007). In addition, insect-derived N-glycans may contain core α1,3-linked fucose residues, which are known to be allergenic (Staudacher et al. 1999). Furthermore, the cell growth is slow (compared to bacteria and yeast) and each protein batch preparation has to be obtained from fresh cells since viral infection is lethal (Ferrer-Miralles et al. 2009). RP products expressed in insect cells have been approved for veterinary and human therapeutic uses (Cox 2012).

Transgenic plants have been used as hosts for the production of RP for a variety of applications due to the advantageous low cost of cultivation, high mass production, flexible scale-up, lack of human pathogens and addition of eukaryotic PTMs (Ferrer-Miralles et al. 2009). Again, the main disadvantage of this expression system is related to the plant specific PTMs introduced: both plant-complex N-glycans and O-glycans are different from human glycans

(Gomord et al. 2010). Moreover, the possibility of gene flow and the negative public perception of transgenic plants hinders its use as an expression system (Ferrer-Miralles et al. 2009). Main challenges for plant-derived RPs include low yields, the long process required to generate transgenic plant lines, and the associated scale-up costs (Hernández et al. 2014). However, there is the possibility of expression of RP in plant cells in a suspension culture allowing the growth in containment in an analogous way to well-established fermentation-based technologies and thus avoiding public concerns and gene flow risks.

The exploration of plants as expression platforms achieved a historic milestone in 2012 when the first protein pharmaceutical expressed in plants was approved for commercial use in humans (Mor 2015). The product is a recombinant form of the human enzyme β-glucocerebrosidase expressed in carrot cells and is intended for replacement therapy in type 1 Gaucher's disease, which is an orphan disease. The product, marketed as ELELYSO, was developed by Protalix Biotherapeutics (Carmiel, Israel) and its approval means a key step towards market acceptance and commercial viability for plant biopharmaceuticals (Mor 2015). It is important to highlight that this product is produced in closed bioreactors facilitating the accomplishment of good manufacturing practices (GMP) and avoiding public controversies. In this case, plant glycosylation resulted in a biobetter product: the enzyme secreted by mammalian cells does not have exposed mannose residues and thereby it requires a costly post-production as part of its formulation. In contrast, the enzyme expressed in plant cells naturally carries high-mannose glycans and thus the production process is significantly cheaper (Tekoah et al. 2013).

Finally, it is worth it to remark an interesting advantage of plants as an expression system that consists in the possibility of using it as oral vehicle of vaccines, therapeutic RP or food supplement for veterinary or human purposes (Haq et al. 1995; Martínez-González et al. 2011; Fragoso et al. 2017; Miletic et al. 2017).

The production of RPs in mammalian cells is characterized by performing appropriate PTMs and high yields of functional proteins (Almo and Love 2014). Main limitations of mammalian cell lines expression systems are high cost of the culture media and production process and slow growth (Barnes et al. 2003; Wurm 2004; Butler 2005; Fischer et al. 2015). In addition, there is an important inherent risk of contamination with pathogenic virus or prions since hemoderivates and animal products are used in growth media (Fischer et al. 2015) requiring strict purification processes and quality control procedures that result in a costly final product.

Several transgenic animals have been successfully used to produce RPs secreted into blood, urine, milk, seminal plasma and silk worm cocoons (Demain and Vaishnav 2009; Houdebine 2009). However, RP production using transgenic farm animals is a challenge in terms of safety concerns such as transmission of infectious diseases (including viral and prion infections) or adverse allergenic, immunogenic and autoimmune responses (Ferrer-Miralles et al. 2009). Also, the production in transgenic animals is time-consuming, expensive and allows only a limited production.

**Table 1. Current expression systems available.** Modified from (Demain and Vaishnav 2009).

| Advantages | Disadvantages |
|---|---|
| **Bacteria (*E.coli*)** | |
| - Cost effective<br>- High yields<br>- Rapid growth<br>- Rapid expression<br>- Easy scale up. | - Produce simple proteins:<br>    - Difficulty to express proteins with disulphide bonds<br>    - Inability to produce glycosylated proteins<br>- Safety risks: presence of endotoxins<br>- Inefficient at protein folding and secretion. Protein aggregation or produced as inclusion bodies - inactive; require refolding. |
| **Yeasts** | |
| - Cost effective<br>- High yields<br>- Rapid growth<br>- Rapid expression<br>- Can express complex proteins:<br>    - Can handle S–S rich proteins<br>    - Can assist protein folding<br>    - Can glycosylate proteins<br>- Can secrete proteins<br>- Easy scale up. | - PTM patterns differ from mammalian ones. |
| **Insect cells** | |
| - Cost effective<br>- High yields<br>- Can express complex proteins:<br>    - Can handle S–S rich proteins<br>    - Can assist protein folding<br>    - Can glycosylate proteins<br>- Can secrete proteins<br>- Safety: low risk of animal pathogens | - Time-consuming,<br>    - Cell growth is relatively slow<br>    - RP batches have to be obtained from fresh cells (viral infection is lethal)<br>- Each protein batch preparation must be obtained from fresh cells (viral infection is lethal).<br>- PTM patterns differ from mammalian ones. |

| Advantages | Disadvantages |
|---|---|
| **Plants** | |
| - Cost effective<br>- High yields<br>- Rapid growth<br>- Can express complex proteins:<br>  - Can handle S–S rich proteins<br>  - Can assist protein folding<br>  - Can glycosylate proteins<br>- Can secrete proteins<br>- Safety: low risk of animal pathogens<br>- Easy scale up. | - PTM patterns differ from mammalian ones.<br>- Negative public perception of the transgenic plants<br>- Difficult gene containment in open fields.<br>- Difficult protein extraction and purification. |
| **Mammalian cells** | |
| - High yield<br>- Can express complex proteins:<br>  - Can handle S–S rich proteins<br>  - Can assist protein folding<br>  - Similar PTMs to human cells - Can glycosylate proteins<br>- Can secrete proteins | - Safety risks: transmission of infectious diseases (including viral and prion infections)<br>- Time-consuming: slow growth<br>- Expensive |
| **Transgenic animals** | |
| - Can express complex proteins:<br>  - Can handle S–S rich proteins<br>  - Can assist protein folding<br>  - Similar PTMs to human cells - Can glycosylate proteins<br>- Can secrete proteins | - Safety risks:<br>  - transmission of infectious diseases (including viral and prion infections)<br>  - Risk of adverse allergenic, immunogenic and autoimmune responses<br>- Time-consuming: long time needed to assess production level<br>- Expensive<br>- Limited production |

Nowadays, glycosylated proteins are usually produced in mammalian cells which mimic human glycosylation, but the resulting product is expensive. Yeasts, plants and insect cells are generally unable to provide mammalian glycosylation (Demain and Vaishnav 2009) and this is crucial as PTMs play a key role in protein folding, processing, stability, solubility, final biological activity, tissue targeting, serum half-life and immunogenicity of the protein (Sinclair and Elliott 2005; Ferrer-Miralles et al. 2009). Therefore, there is a need for the exploitation of better RP expression hosts able to improve quality of proteins while lowering its costs. The continuous improvement of the established expression systems by metabolic engineering is a proof of this need (De Pourcq et al. 2010; Gomord et al. 2010; Fischer et al. 2015). In addition to improving the current systems, the wide diversity of metabolisms prompts to explore new RP expression hosts. These unconventional expression systems (see table 2) might offer opportunities for RP production for several applications by improving RP characteristics or by lowering its production costs (Ferrer-Miralles et al. 2009; Corchero et al. 2013). It is worth mentioning an interesting concept in the context of therapeutic RP: biobetters. Biobetters are products with improved characteristics when compared to products already marketed: cheaper, more efficacious, less frequently dosed, better targeted, and/or better tolerated (Reski et al. 2015). For example, human α- galactosidase produced in moss for enzyme replacement therapy in Fabry disease is nowadays in clinical trials and is a biobetter because it lacks the terminal mannose phosphorylation showing improved pharmacokinetics in Fabry mice (Reski et al. 2015).

Algae as an alternative RP expression host is introduced in the next section.

**Table 2. Most relevant alternative hosts for recombinant protein production.** Modified from Corchero et al. 2013.

| Host platform | Model species | Main features | References |
|---|---|---|---|
| **Cold adapted bacteria** | *Pseudoalteromonas haloplanktis* | Improved protein folding | (Duilio et al. 2004a, b; Giuliani et al. 2011) |
| **Pseudomonads** | *Pseudomonas fluorescens* | Efficient secretion | (Retallack et al. 2007, 2012; Jin et al. 2011) |
| **Coryneform bacteria** | *Corynebacterium glutamicum* | High-level production and secretion | (Kikuchi et al. 2008, 2009) |
| **Bacilli** | *Bacillus megaterium* | High-level production and secretion | (Stammen et al. 2010; David et al. 2011) |
| **Lactic acid bacteria** | *Lactococcus lactis* | Secretion; GRAS (Regulatory issues) | (Le Loir et al. 2005; Miyoshi et al. 2006; Morello et al. 2007; Innocentin et al. 2009) |
| **Filamentous fungi** | *Tricoderma reseei* | High-level production and secretion; post-translational modifications | (Uusitalo et al. 1991; Nyyssönen et al. 1993; Collén et al. 2005; Guillemette et al. 2007; Jun et al. 2011) |
| **Moss** | *Physcomitrella patens* | Post-translational modifications | (Decker and Reski 2007, 2008; Reski et al. 2015) |
| **Algae** | *Chlamydomonas reinhardtii* | Cheap nutrients; high potential for scalability | (Mayfield and Franklin 2005; Specht et al. 2010; Potvin and Zhang 2010; Gong et al. 2011) |
| **Unconventional plant** | *Lemna minor* | cheap nutrients, Post-translational modifications | (Cox et al. 2006; Bertran et al. 2015) |

### *Microalgae as an expression host*

The term "microalgae" refers to a diverse photosynthetic group of prokaryotic (cyanobacteria) and eukaryotic microrganisms with biotechnological potential. Historically, microalgae have been used in several applications, ranging from enhancing the nutritional value of animal feed to as producers of high-value molecules, like polyunsaturated fatty acid oils or human nutritional supplements and pigments (Spolaore et al. 2006). Furthermore, microalgae have been studied extensively in the last decades due to their potential to generate biofuel (Demirbas 2010; Wijffels and Barbosa 2010; Kilian et al. 2011). Microalgae have also attracted attention as an alternative RP production system because they offer the benefits of plants coupled with high productivities associated with microbial production. Thus, microalgae are promising as expression hosts due to low production cost, high growth rates, high yields, better safety and, in the case of eukaryotic microalgae, the ability to accomplish PTMs (Rosales-Mendoza 2016).

Despite the fact that some species can be grown as heterotrophs in fermenters without light as energy source, thus requiring a supply of sugars for energy and as a carbon source, most microalgae are photoautotrophs requiring only light, water and basic nutrients for their culture (Walker et al. 2005). Besides the low-cost production, there is no gene flow by means of pollen or other vehicles of gene escape. This fact coupled with the possibility of growth in closed bioreactors (Pulz 2001) makes transgenic microalgae an environmentally safe alternative (Walker et al. 2005; Rasala and Mayfield 2015).

The safety due to the absence of contamination by human pathogens is an advantage in front other RP hosts (Walker et al. 2005). Also, some algae species hold a GRAS (generally recognized as safe) status and thus, for many applications including food, agriculture, and cosmetic industries, little purification may be required. Also, they can be used as oral vehicle of vaccines, therapeutic RP or food supplement both for veterinary or human purposes (Specht and Mayfield 2014; Beltrán-López et al. 2016).

Furthermore, an advantage is the quick generation of RPs, requiring only a few weeks between the production of transformants and their scale up to production volumes.

Finally, it has to be taken into account that a plant-made biopharmaceutical approved by the FDA in 2012 is a key reference in favour of algae-made therapeutic RPs regarding accomplishment of approvals (Mor 2015; Rosales-Mendoza 2016).

The diatom *Phaedodactylum tricornutum* has been studied as a feasible alternative host for mAb production. Several fully-assembled mAbs have been expressed in *P. tricornutum* retained in the endoplasmic reticulum (Hempel et al. 2011) and secreted to medium (Hempel and Maier 2012; Hempel et al. 2017) with a maximum expression level reported of 2500 ng/ml (Hempel and Maier 2012). One of the expressed mAbs showed a reduced affinity to its receptors compared to mammalian mAb, and it has been suggested to be due to differences in glycosylation of the Fc region (Vanier et al. 2017).

Despite the increasing examples of successful transformation and RP expression from many microalgal species (see table 3), current work is performed with *Chlamydomonas reinhardtii* that, since has been a model organism for decades, is the best characterized microalgal species. *Chlamydomonas reinhardtii* is introduced in the next section.

**Table 3. Microalgae used for recombinant protein expression.** Adapted from (Rasala and Mayfield 2015; Rosales-Mendoza 2016).

| Microalgae | Use in RP expression | References |
|---|---|---|
| *C.reinhardtii* | +++ | (Rasala et al. 2012; Gimpel et al. 2015) |
| *Phaeodactylum tricornutum* | ++ | (Hempel et al. 2011; Hempel and Maier 2012) |
| *Dunaliella salina* | ++ | (Geng et al. 2003; Feng et al. 2014) |
| *Schizochytrium sp.* | + | (Bayne et al. 2013; Bañuelos-Hernández et al. 2017; Ortega-berlanga et al. 2018) |
| *Chlorella ellipsoidea* | + | (Chen et al. 2001; Kim et al. 2002) |
| *Dunaliella tertiolecta* | + | (Georgianna et al. 2013) |
| *Scenedesmus sp.* | + | (Chen et al. 2017) |

## *Chlamydomonas reinhardtii*

*Chlamydomonas reinhardtii* (Chlamydomonas hereafter) is a freshwater, green microalga that has been a popular model organism in plant biology for physiological, molecular, biochemical and genetic studies. Because of that, it has a well-developed molecular toolkit (Jinkerson and Jonikas 2015). Several key features of this organism make it a useful model organism for a variety of topics (Harris 2001). Thanks to its metabolic flexibility (ability to grow photoautotrophically or heterotrophically) Chlamydomonas enables the study of mutations in genes essential for photosynthesis, which would be lethal in many other systems. Furthermore, the unicellular nature of Chlamydomonas simplifies the study of responses to environmental stresses by eliminating the complexity associated with the different tissues found in higher plants. Liquid cell cultures can be grown easily in the laboratory, and their cell cycles can be easily synchronized by growth in a light/dark diurnal cycle. Vegetative cells are haploid and divide by multiple fission (Cross and Umen 2015), being thus suitable for loss-of function genetic studies because recessive mutations immediately show a phenotype. Chlamydomonas has also a well-characterized sexual cycle induced under nitrogen starvation conditions, and gametes can be mated to allow tetrad analysis. Both the chloroplast and the mitochondrial genomes can be genetically transformed through highly efficient homologous recombination techniques and this feature, along with the availability of numerous nuclear mutants and nuclear genome transformation and together with recent genome editing technologies, facilitates the study of nuclear/organellar interactions (Jinkerson and Jonikas 2015).

Chlamydomonas has also gained attention as an expression host for the production of therapeutic proteins (Mayfield et al. 2007; Rasala, B.A., Mayfield 2011; Rosales-Mendoza et al. 2012). Although genetic transformation techniques are well established for both the nucleus and the chloroplast, proteins of pharmaceutical interest have been expressed to economically viable RP yields only in chloroplast (Mayfield et al. 2003, 2007; Mayfield and Franklin 2005; Tran et al. 2009; Rasala, B.A., Mayfield 2011). In fact, the Chlamydomonas chloroplast, which occupies up to 40 % of the cellular volume, possesses prokaryotic machinery together with proper synthesis of complex proteins requiring disulphide bonds for correct folding (Rasala, B.A., Mayfield 2011; Tran et al. 2013b). However, expression from the nuclear genome is necessary for the production of some RPs because RPs expressed in nucleus can be produced with PTMs that are absent from proteins produced in the chloroplast (Mayfield et al. 2007). Moreover, whereas chloroplast produced proteins remain within the plastid, proteins expressed in the nuclear genome can be targeted to the endoplasmic reticulum for secretion into the external medium (Eichler-Stahlberg et al. 2009; Rasala et al. 2012; Lauersen et al. 2013a). Thus, proteins expressed in the chloroplast involve the need of a complex extraction while secretion of proteins is very appreciated for downstream processing.

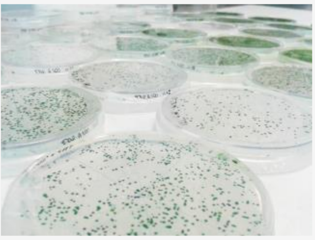
**Table 4. Characteristics of recombinant protein expression in Chlamydomonas nucleus or chloroplast.** Adapted from (Rasala and Mayfield 2015).

| | Nucleus | Chloroplast |
|---|---|---|
| **Gene integration** | Random integration | Homologous recombination |
| **Stability of transgene** | Susceptible to gene silencing | Highly stable once homoplasmic |
| **Levels of recombinant protein accumulation** | Low (0.25% TSP)<br>Levels susceptible to position effect | High (1-21% TSP)<br>Levels consistent across clones |
| **Post-translational modification** | Disulphide bond formation, phosphorylation, glycosylation, sulfation | Disulphide bond formation, phosphorylation |
| **Protein localization** | Cytosol, organelle targeting, secretion | Protein retained in chloroplast |

### *Limitations of Chlamydomonas reinhardtii for nuclear transgene expression*

As exposed, Chlamydomonas has long served as a model organism in plant biology. Even though the availability of a molecular toolbox and the high-efficiency transformation of the nuclear genome (Kindle 1990; Shimogawara et al. 1998), there are a difficulties associated with the integration and expression of foreign DNA. First, transgenes are integrated randomly through non-homologous end joining leading to a high degree of expression variability due to position effects (Kindle 1990). Second, transgenes are often fragmented or rearranged during integration (Zhang et al. 2014) and that makes necessary the screening of large numbers of transformants to identify the ones desired. Third, transgenes are commonly subject to silencing (Cerutti et al. 1997; Wu-Scharf et al. 2000; Jeong Br et al. 2002). Moreover, transgene silencing has been demonstrated at both the transcriptional and post-transcriptional levels (Ceruttti et al. 1997; Wu-Scharf et al. 2000) and this results in unreliable long-term expression of genes of interest (GOI).

Several improvements have been made to enhance transgene expression from the nuclear genome of Chlamydomonas, the strategies followed towards this end will be discussed together with our proposed approaches in Chapter 1.



# Objectives

# *Objectives*

The global objective of this thesis is the enhancement of *Chlamydomonas reinhardtii* (Chlamydomonas hereafter) as a recombinant protein expression system and its validation as a versatile host with the nuclear expression of two protein models: a small growth factor and a complex monoclonal antibody (mAb). Chlamydomonas is an excellent model organism for plant biology and an attractive biotechnological production host. Despite the extensive molecular toolbox available, transgene expression levels are usually low, and this limitation has hampered the use of Chlamydomonas as a host for RP production. Several strategies designed to reduce or overcome these limitations in recombinant protein expression are tested (chapter 1) and the versatility of the host is assessed through the expression of two proteins (chapters 2 and 3). Specific objectives are listed below:

1. **Improving RP expression levels from Chlamydomonas nuclear genome** (Chapter 1).

   - *STRATEGY 1. Design of DNA cassette and optimization of a high-throughput screening method to obtain a high RP expressing transformant.* The high variability in the expression of the same cassette from independent clones of Chlamydomonas makes it necessary to screen a high number of transformants to select those which express the highest levels of RP. Since our high-throughput screening must be suitable for a GOI without an observable phenotype, two approaches are tested:

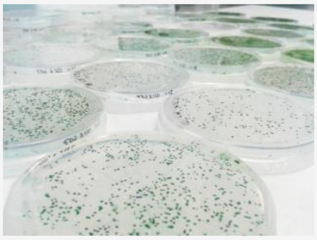
     - Use of a fusion protein consisting of a gene conferring resistance to an antibiotic and the protein of interest. A high-throughput screening is proposed based on fusion proteins comprising cleavage regions or skipping peptides.

  - Use of a fusion protein consisting of a reporter gene and the protein of interest. A high-throughput screening is proposed based on fusion proteins comprising several secretion peptides, regulatory regions and stabilizing regions. Influence of the effect of said different regions to increase RP expression yield is studied.

- *STRATEGY 2. Study of the effect of transgene copy number on nuclear transgene expression levels.* Two approaches are studied: i) the relation between transgene copy number and recombinant protein expression in haploid cells, and ii) recombinant protein expression in diploid cells.
- *STRATEGY 3. Obtaining a Chlamydomonas strain with improved expression of nuclear transgenes.* A screening to obtain strains with impaired silencing mechanisms is designed and performed.

2. **Purification and functional characterization of human Epidermal Growth Factor expressed from Chlamydomonas nucleus** (Chapter 2). A purification protocol and downstream processing from Chlamydomonas culture supernatant is proposed. Functional activity of expressed hEGF is tested.

3. **Expression of a monoclonal antibody from Chlamydomonas nucleus** (Chapter 3). A high-throughput screening to select a transformant expressing a fully-assembled murine mAb is designed. Characterization of produced antibody is performed.



# Chapter 1. Improving expression levels from Chlamydomonas nuclear genome

# Introduction

As introduced previously, each expression system has its own limitations and there is a continuous effort to improve those expression systems and the products obtained. There are several levels at which actions can be taken to enhance gene expression: increasing transcription, increasing translation, favouring stability of RNA and/or stability of protein and favouring folding and PTMs of the recombinant protein. These can be achieved by improving strains, optimizing the heterologous DNA introduced (e.g. regulatory regions, fusion proteins, codon usage) into the cells or co-expressing specific proteins (e.g. foldases).

The molecular mechanisms limiting the RP expression in Chlamydomonas nucleus are not completely understood (Dong et al. 2017). However, several strategies to improve transgene expression in the nuclear genome of Chlamydomonas have been reported, even though they are not always successful. Some of these strategies are commonly employed in other hosts while some are novel. An overview of these strategies is presented below.

### *Codon optimization improves transgene expression*

Modifying the coding region to improve the level of gene expression is an established practice in RP expression from *E.coli* to mammalian cells (Makrides 1999; Ferrer-Miralles et al. 2009).

The Chlamydomonas nuclear genome has an unusually high G+C nucleotide content of 68% in coding sequences (Merchant et al. 2007) and the use of transgenes with low GC content or strongly deviating codon usage are expressed only poorly or not at all (Fuhrmann et al. 2004; Shao and Bock 2008; Barahimipour et al. 2015). Furthermore, the optimization of genes to match the GC content and codon usage found in Chlamydomonas has been successful to improve transgene expression in Chlamydomonas. It has been reported that, while GC content exerts an influence on local chromatin structure, codon usage determines translational efficiency and mRNA stability (Barahimipour et al. 2015). It has been shown that the optimization of genes may introduce codons containing splicing signals (Weiner et al. 2018). Also, the same study showed the use of a frequent subset of preferred codons to increase transcript levels and demonstrated the effect on gene expression of mRNA folding energy in the translation initiation vicinity.

### *Use of optimized regulatory elements to increase transgene expression*

In plants, bacteria and mammalian cells, viral and homologous genes are the two major sources of promoters used (Peremarti et al. 2010; Jia and Jeon 2016; Romanova and Noll 2017). The use of homologous cis regulatory regions to drive expression is also a well-established practice in RP expression in yeasts and molds (Porro et al. 2005; Demain and Vaishnav 2009). Based on the promoters used, chimeric promoters have been designed for several expression systems (Brown et al. 2014). Regarding Chlamydomonas, several native promoters of highly expressed genes and chimeras of native promoters have been reported to date, being heat shock protein 70A and Rubisco2 (*HSP70-RBCS2*) chimeric promoter improved by insertion of the *RBCS2* first intron downstream (AR) the one most commonly used (Schroda et al. 2000; Sizova et al. 2001). The promoter-terminator from the *PSAD* gene (without introns in its sequence) has also been successfully used to express cDNAs (Fischer and Rochaix 2001; Heitzer and Zschoernig 2007; Kumar et al. 2013). Recently it has been shown that designed synthetic promoters can outperform AR promoter (Scranton et al. 2016). Finally, the native cis regulatory regions of *RPL23* gene have been shown to increase expression in comparison to AR promoter (López-Paz et al. 2017). In this latter work it was also shown that the use of 5’ and 3’ cis regulatory regions from a single gene outperforms the combination between cis regulatory regions of different genes. In the same work it was reported that not only promoter but also UTR play a role on efficiency to drive transgene expression.

The effect of introns on expression of genes is an evolutionarily conserved feature found in diverse organisms as yeast, insects, plants and mammals. This positive effect on gene expression has been named intron-mediated enhancement (IME) and, even though the mechanism behind is not entirely clear (Moabbi et al. 2012; Laxa 2017), intronic sequences are used to increase transgene expression (Vain et al. 1996; Wurm 2004). It has been demonstrated that the inclusion of an intron from a native gene can improve transgene expression in Chlamydomonas (Lumbreras et al. 1998; Berthold et al. 2002; Eichler-Stahlberg et al. 2009). The initial and strongest evidence for the improvement of transgene expression by native introns comes from resistance markers in which the addition of introns increases transformation efficiencies and tolerance to higher antibiotic concentrations (Lumbreras et al. 1998). Intronic sequences have been generally used in expression of transgenes, for example in combination with *HSP70A/RBCS2* chimeric promoter (Sizova et al. 2001; Hu et al. 2014).

Matrix attachment regions (MARs) are eukaryotic regulatory DNA elements that mediate attachment to the nuclear scaffold. These elements promote more open chromatin structure (Dietz-Pfeilstetter 2010) thus increasing transformation efficiency and reducing expression variability (Mlynarova et al. 1996; Petersen et al. 2002). Although the mechanism of MARs action *in vivo* is still not completely elucidated (Romanova and Noll 2017) successful results have been obtained in plants and mammalian cells (Allen et al. 1993, 1996; Girod et al. 2007). It has also been shown that the incorporation of MARs into transgenes may protect against transgene silencing and improve the stability of expression over a number of generations (Ülker et al. 1999; Vain et al. 1999; Mlynarova et al. 2003).

_Fusion proteins_

The use of fusion proteins is a usual resource to facilitate or increase recombinant expression of GOI. As an example, in *E.coli* solubility of RP can be increased when genes are fused to the *E.coli* thioredoxin gene (LaVallie et al. 1993). The Fusion to Fc regions or stable proteins such as hSA are used not to increase RP expression but to improve pharmacokinetic properties (increased serum half-life) of the target protein (Capon et al. 1989; Yeh et al. 1992; Czajkowsky et al. 2012; Strohl 2015).

In Chlamydomonas, transgene expression can be increased by fusing the coding region of a GOI to a selectable marker gene, in a manner that a single promoter is used to drive expression of the fusion protein (Fuhrmann et al. 1999). This method increases the chances that primary transformants express the desired GOI, allows more reliable gene expression over time, and reduces the amount of screening required to identify suitable transformants. However, fusion of a GOI to a marker gene may hinder the function of the GOI, the marker or both. It has been reported that the fusion consisting of the intron-containing bleomycin gene ($Ble^R$) and GOIs support higher level expression of these GOIs as compared to their unfused counterparts (Fuhrmann et al. 1999, 2004; Rasala et al. 2012; Kong et al. 2015).

_Secretion_

A common strategy in industrial RP production is targeted secretion of RPs to the periplasmic or extracellular space. It is considered that protein localization into the periplasmic space has reduced exposure to protease activity, less inhibitory feedback to gene expression processes, and allows continuous cultivation in perfusion-style bioreactors (Fischer et al. 2004; Schmidt 2004;

Porro et al. 2005). In addition, proteins that are secreted through the ER and Golgi exocytosis pathway are subjected to post-translational modifications (PTMs) (Eichler-Stahlberg et al. 2009). Furthermore, secretion to medium facilitates the isolation of the protein. Regarding Chlamydomonas, in an study using a secreted fusion protein comprising a reporter gene and a GOI an expression level of 12 mg RP/L has been achieved (in the highly expressing mutant UVM4 strain) (Lauersen et al. 2015b).

In plant cell cultures, a successful strategy to enhance secretion yields and the stability of recombinant proteins has been demonstrated based on $(SP)_n$ glycomodules, which are O-glycosylated with arabinogalactan polysaccharides attached to hydroxylated Pro residues (Hyp) (Shpak et al. 1999; Xu et al. 2007). Glycosylation serves a variety of structural and functional roles and it is well established that glycosylation increases the stability of proteins (Sinclair and Elliott 2005). The presence of O-linked glycans, for instance, strongly enhances the secretion yields and physicochemical properties of a protein by influencing protein folding, solubility, stability and resistance to heat or proteolysis (Walsh and Jefferis 2006; Gomord et al. 2010). From green algae to plants, hydroxyproline (Hyp)- O-glycosylation characterizes an ancient and diverse group of structural glycoproteins associated with the cell wall (Shpak et al. 1999). Peptide sequence periodicity and glycosylation distinguish the three major HRGP families: arabinogalactan proteins (AGPs), extensins, and proline-rich proteins (PRPs). Moreover, the conserved (Ser-Hyp)n motif has been identified both in Chlamydomonas (Woessner and Goodenough 1992) and in higher plants AGPs (Kieliszewski and Lamport 1994). When adding $(SP)_{10}$ or $(SP)_{20}$ glycomodules the expressions achieved 7.5 mg/L and 15 mg/L respectively, 6 and 12-fold higher, than the secreted RP without glycomodules (Ramos-Martinez et al. 2017). To our knowledge this is the highest yield of protein expression in Chlamydomonas nucleus reported until today.

### *Gene dosage*

One strategy that has been investigated to increase transgene expression is gene dosage, involving multiple copies of an heterologous gene introduced into the genome (Aw et al. 2012; Zhou et al. 2013). For some organisms is one of the best established methods for increasing expression yields (Aw et al. 2012). In practice however, for some other organisms there is no direct correlation between copy number and expression yields (De Buck et al. 2004; Schubert 2004; De Paepe et al. 2012; Zhou et al. 2013). Despite having originally been described over 20 years ago, the use of multicopy clones is an aspect that is still being investigated in depth today (Aw et al. 2012).

In plants, the analysis of the correlation between copy number and gene expression is complicated because increasing the copy number can both increase and decrease the level of expression of heterologous genes in transgenic plants, the key criterion being the nature of the copy that is involved (Hobbs et al. 1993): some insertions act in an additive manner but others are silenced and are able to cause silencing on other introduced high expressing copies. This correlation has been controversial during years and the comparison between different plant expression systems has made more difficult the establishment of a relation between gene dosage and transgene expression by yielding apparently conflicting results (Schubert 2004). Although repetitiveness is an important feature in the induction of silencing, single-copy transgenes can also be silenced, either by the influence of trans-acting mechanisms or by the neighbouring heterochromatic DNA (Dorokhov 2007). Besides, the nature of the coding region is an important determinant of the gene-specific threshold for silencing (Schubert 2004).

Concerning Chlamydomonas, reported results indicate that transgene copy number is not positively correlated with expression level (Barahimipour et al. 2015, 2016). However, transformants carrying more than one copy of transgene have been identified for robust expression of a GOI (Lauersen et al. 2016). Therefore, the high variability in the expression of the same cassette in independent clones of Chlamydomonas makes it difficult to establish a correlation between copy number and expression level of the transgene from published studies.

In plants, induction of polyploidy for crop improvement dates to the 1930s, and doubling of the chromosome number is associated with increased biomass and tolerance to various stresses (Lee et al. 2010). These improved traits are essential for the maximal production of crops on a large agricultural scale. Chlamydomonas vegetative diploid colonies can be obtained (Ebersold 1967). These strains are different from haploids by cell and nuclear size, DNA content per nucleus, and chromosome number (Ebersold 1967). Recently, it was reported that under nitrogen starvation conditions diploid cells grow better in terms of biomass because its cellular weight triplicate the cellular weight of haploid cells (Kwak et al. 2017). Interestingly, the diploids showed enhanced tolerance to other stresses, including oxidative and cold stresses (Kwak et al. 2017).

To our knowledge, there hasn’t been a study to determine the relation between copy number and transgene expression in Chlamydomonas nucleus or to determine the expression of RP in diploid Chlamydomonas cells.

### *Strains improved for high transgene expression*

The use of mutant backgrounds is one of the key strategies in molecular biology for both basic research and applied. For example, the creation of protease deficient mutants to avoid product degradation is a reality in bacteria (Joseph et al. 2015) or molds (Demain and Vaishnav 2009). Also, metabolic engineering of yeast strains has been achieved in order to mimic human glycosylation (De Pourcq et al. 2010) and *E.coli* strains had been developed to facilitate the expression of rare codons (Loyevsky et al. 2003).

Specific mutant backgrounds can be used to reduce silencing of transgenes by epigenetic mechanisms. It has been shown that transgene inactivation in Chlamydomonas occurs at transcriptional and post-transcriptional levels (Ceruttti et al. 1997; Wu-Scharf et al. 2000; Jeong Br et al. 2002). A transgene that is inserted multiple times into the nuclear genome may be transcriptionally silenced through DNA methylation (Ceruttti et al. 1997) or histone modification (Casas-Mollano et al. 2007). However, single-gene copies of transgenes can also be silenced. For instance, *E.coli* aadA gene, conferring spectinomycin resistance, stably integrates into the nuclear genome but in approximately half of the transformants its expression is reversibly suppressed (Ceruttti et al. 1997). This gene silencing was shown to be a result of transcriptional inactivation and did not correlate with methylation of the integrated DNA. To identify the mechanism of silencing, an insertional mutagenesis screen was conducted in a background with a silenced aadA resistance marker, and mutants were selected by transgene expression reactivation (Wu-Scharf et al. 2000; Jeong Br et al. 2002). Two proteins were found to be involved in transcriptional gene silencing, MUT9 and MUT11, which likely function in the formation of chromatin structure (Casas-Mollano et al. 2008) and as a global transcriptional repressor (Zhang et al. 2002) respectively. Protein MUT6, that encodes a DEAH-box RNA helicase (Wu-Scharf et al. 2000), was found to be involved in post-transcriptional gene silencing. This protein is necessary to silence transgenes by aiding in degradation of aberrant RNAs, formed when transcripts are incorrectly processed (Wu-Scharf et al. 2000).

A Chlamydomonas genetic screening was developed using a selectable marker gene (CRY1-1) whose expression level is directly proportional to the phenotypic resistance to the selecting agent (Neupert et al. 2009). With this system, the selection pressure could be increased to select mutants with high transgene expression. Thanks to this selection strategy, two UV mutants that showed increased transgene expression relative to its parental strain were isolated (Neupert et al. 2009): UVM4 and UVM11. When these mutants where transformed to express two fluorescent proteins (GFP and YFP) they showed increased fluorescence over non-selected strains and a higher percentage of its transformants incorporated the full-length transgene cassette into the genome than wild-type strains. UVM4 was used to produce a secreted luciferase protein and produced more than threefold increase in luciferase than a wild-type strain using the same expression cassette. Additionally, not only the UVM4 background produced more luciferase luminescence positive transformants but more robust bioluminescence signals than wild-type (Lauersen et al. 2013a). Another study used UVM mutants for the expression of three genes in the biosynthetic pathway of the terpene squalene and while transgene levels were increased in the UVM mutants over wild-type, expression levels were variable between individual UVM transformants. These results suggest that the epigenetic repression mechanism is alleviated but not eliminated in the UVM strains (Kong et al. 2014). Finally, it is important to highlight that UVM strains harbour unknown mutation(s).

A Chlamydomonas mutant with a tag insertion in the maintenance-type DNA methyltransferase gene (*MET1*) which encodes one of the key enzymes for epigenetic gene silencing was identified (Kong et al. 2015). This mutant significantly improved the expression of transgenes but was susceptible to transcriptional gene silencing and expression varied among its transformants (Kong et al. 2015). Later, this Met1-strain was used as a background for random mutagenesis (Kurniasih et al. 2016) and some of the resulting mutant were reported to reach the expression level of UVM strains, showing an enhanced nuclear transformation efficiency and increased transgene expression irrespective of the insertion position (Kurniasih et al. 2016).

While mutant strains have already been useful for transgene expression, new screenings are needed to identify and isolate new improved strains and to help in further describing silencing mechanisms in Chlamydomonas.

With the aim of improving expression levels from Chlamydomonas nucleus three strategies have been performed: i) design of DNA cassette and optimization of a high-throughput screening method to obtain a high RP expressing transformant, ii) study of the effect of transgene copy number on nuclear transgene expression levels and Iii) obtainment of a Chlamydomonas strain with improved expression of nuclear transgenes.

# Results

## *Strategy 1. Design of DNA cassette and optimization of a high-throughput screening method to obtain a high RP expressing transformant*

As previously introduced, the high variability in the expression of the same cassette from independent clones of Chlamydomonas makes it necessary to screen a high number of transformants to select those with high expression. As we intend to develop a high-throughput screening suitable for GOIs without an observable phenotype, we need a marker or a reporter gene. In the present work two approaches are tested: a fusion protein comprising an antibiotic resistance gene and a GOI, and a fusion protein comprising a reporter gene and a GOI. Three GOIs were initially selected for the present study.

The protein of interest was human Epidermal Growth Factor (*EGF*, Gene ID: 1950). The mature form of EGF is a small polypeptide of 6045 Da and six half-cysteines that exist in disulphide linkage. This protein acts as a potent mitogenic factor that plays a key role in the growth, proliferation and differentiation of numerous cell types. Since its discovery (Cohen 1962) all aspects of EGF biology have attracted an intense research interest (Wong et al. 2001). EGF had shown a potential utility in various areas of human health care and has also attracted commercial interest in the cosmetic market.

Another protein of interest was *Chlamydomonas reinhardtii* Thioredoxin h1 (*TRXH*, gene Cre09.g391900). Thioredoxins are small proteins which are involved in the cell redox regulation. These ubiquitous proteins are present in all organisms from prokaryotes to higher eukaryotes. 10 different thioredoxins have been identified in Chlamydomonas (Lemaire and Miginiac-Maslow 2004).

Chlamydomonas cytosolic Thioredoxin h1, with a molecular weight (MW) of 11.8 kDa, was selected to be expressed in the present study as a native protein of Chlamydomonas with commercial interest in the cosmetic market.

The fluorescent protein mCherry was also included as a protein easy to follow. mCherry, a protein of 26.7 kDa, has been previously reported as a good fluorescent candidate because Chlamydomonas shows low autofluorescence at the wavelengths where it is detected (Rasala et al. 2013).

# 1. Developing a high throughput screening method: fusion of a selectable marker gene and the protein of interest

## 1.1. Use of polycistronic mRNAs: analysis of two different viral 2A sequences

### 1.1.1. Use of minimal FMDV 2A peptide for processing of a dicistronic vector

The first proposal was based on the fusion strategy of Ble$^R$2AGOI that had previously resulted in a yield of 0.25% total soluble protein (TSP) in Chlamydomonas nucleus (Rasala et al. 2012). Ble$^R$ resistance gene, that confers resistance to the bleomycin family of antibiotics (Stevens et al. 1996), was cloned in frame with the Foot-and-mouth disease virus 2A (FMDV 2A) peptide sequence and the coding region of the Chlamydomonas Thioredoxin h1 (referred to as TRXH in this study), mCherry or human Epidermal growth factor (referred to as hEGF in this study). Ble$^R$ protein (referred to as ShBle in this study) binds to zeocin (bleomycin derivative antibiotic used in the present study) in a 1:1 ratio to prevent DNA double strand breaks induced by the antibiotic (Dumas et al. 1994). Unlike enzymatic resistance genes, high levels of ShBle are needed to survive zeocin selection. Therefore, this strategy is designed to exclusively select transgenic lines that express high levels of shBle and consequently high levels of the protein of interest (Rasala et al. 2012). Since Ble$^R$ gene is cloned in frame with a GOI mediated by the 2A peptide sequence, it is expected that when the ribosome reaches the 2A peptide it will skip the bond formation rendering two independent proteins: ShBle and protein of interest.

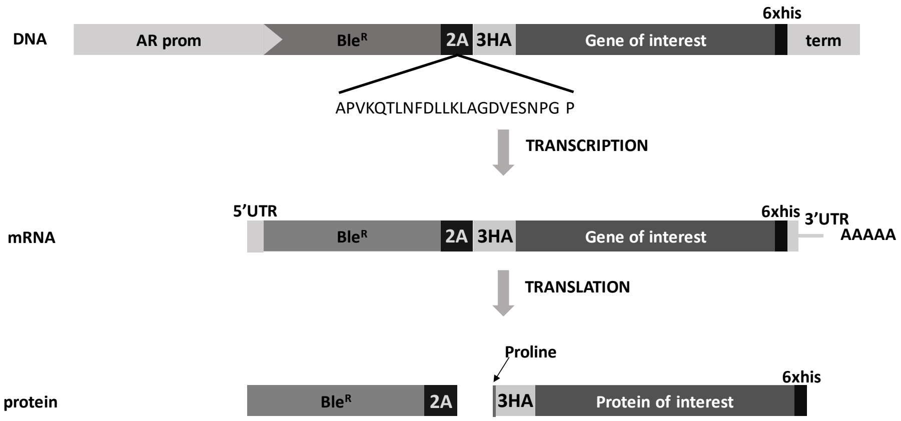

DNA
AR prom
Ble$^R$
2A
3HA
Gene of interest
6xhis
term
APVKQTLNFDLLKLAGDVESNPG P
TRANSCRIPTION
mRNA
5'UTR
Ble$^R$
2A
3HA
Gene of interest
6xhis
3'UTR
AAAAA
TRANSLATION
protein
Ble$^R$
2A
Proline
3HA
Protein of interest
6xhis

**Figure 1. Schematics of FMDV 2A function.** The dicistronic 2A-containing construct used to assess Ble2AGOI strategy for Chlamydomonas nuclear expression would transcribe as one single mRNA and would be translated into two separate peptides (shBle and protein of interest) because when the ribosome reaches the 2A peptide the bond formation is skipped.

To test this strategy three cassettes were designed: pARshBle2ATRXH, pARshBle2AmCherry and pARshBle2AhEGF (see figure 2). Expression was driven by AR promoter (*HSP70-RBCS2* chimeric promoter including *RBCS2* first intron downstream) and 3'UTR *RBCS2*. Ble$^R$ contained *RBCS2* intron 1 that has been previously reported to confer higher expression (Lumbreras et al. 1998). Furthermore, excepting the case of the cassette carrying mCherry gene, three copies of haemagglutinin epitope tag (3HA) (cloned N-terminal of the GOI) and a 6xhis tag (cloned C-terminal of the GOI) were introduced for further detection and downstream processing. The 2A sequence used corresponded to the minimal FMDV 2A sequence (see figure 1) (Rasala et al. 2012, 2013). All the coding region was adapted to the codon usage and GC content of

Chlamydomonas nuclear genome except for the 3HA tag and the mCherry gene.

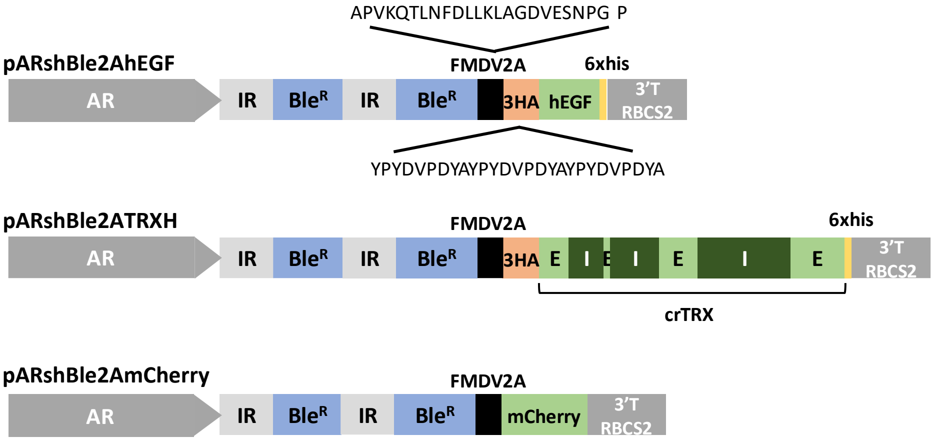

APVKQTLNFDLLKLAGDVESNPG P
pARshBle2AhEGF
FMDV2A
6xhis
AR
IR
Ble$^R$
IR
Ble$^R$
3HA
hEGF
3'T RBCS2
YPYDVPDYAYPYDVPDYAYPYDVPDYA
pARshBle2ATRXH
FMDV2A
6xhis
AR
IR
Ble$^R$
IR
Ble$^R$
3HA
E
I
E
I
E
I
E
3'T RBCS2
crTRX
pARshBle2AmCherry
FMDV2A
AR
IR
Ble$^R$
IR
Ble$^R$
mCherry
3'T RBCS2

**Figure 2. Schematics of pARshBle2AhEGF, pARshBle2ATRXH and pARshBle2AmCherry constructs.** AR: *HSP70-RBCS2* chimeric promoter, 3'T RBCS2: 3' region of *RBCS2*, IR: intron 1 *RBCS2*, 3XHA: triple hemagglutinin tag, brackets represent cDNA of *TRXH* (gene Cre09.g391900), E: TRXH exon, I: TRXH intron.

The constructs pARshBle2ATRXH and pARshBle2AhEGF were transformed into the cell-walled Chlamydomonas strain CC-124 and initially selected on TAP-zeocin 15 µg/ml (considered low zeocin). 96 independent transformants were picked and retested for growth on increasing concentrations of zeocin to assess relative expression levels of shBle.

Independent transformants able to grow at the maximum zeocin concentration tested (100 µg/ml) were selected for further analysis: 6 for TRXH and 8 for hEGF (figure 3).

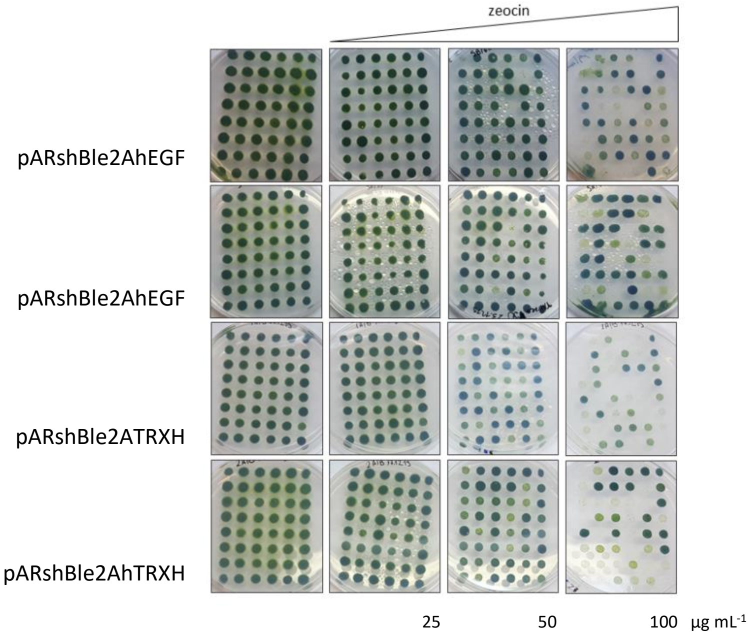


**Figure 3. Zeocin sensitivity test of pARshBle2AhEGF and pARshBle2ATRXH transformants.** Transformants were spotted onto TAP agar plates containing increasing concentrations of zeocin. Selection of high resistant clones was assessed by growth at increasing zeocin concentrations.

An immunoblot analysis was performed to corroborate that high zeocin resistance is correlated with high transgene expression as well as to check the correct expression and processing of the both peptides (figure 4). We used two antibodies to detect both processed and unprocessed forms: a polyclonal antibody against shBle protein to detect full length protein and shBle processed protein and a monoclonal antibody against HA tag to detect full length protein and processed C-terminal protein. Unexpectedly, when transformants were analysed with the antibody against shBle, proteins of varied sizes were observed between different transformants. When transformants were analysed with the antibody against HA tag only one out of 8 transformants showed expression of a protein with a size that corresponds to the expected size of the full length protein (34.87 kDa in the case of shBle:2A:3HA:TRXH and 28.95 kDa in the case of shBle:2A:3HA:hEGF). Importantly, no band or signal was detected in CC-124 negative control with any of the antibodies used. We expected that all clones express shBle since they are resistant to zeocin, but in 4 cases (out of 15) we were not able to detect the protein by immunoblot against shBle. The more likely explanation is expression of shBle under the antibody detection limit although silencing (after the screening on zeocin selection and before the immunoblot) is another possible explanation.

The transformants that showed presence of a protein at the expected molecular weight for the fusion protein were selected: pARshBle2ATRXH#8 and pARshBle2AhEGF#3. According to the results (figure 4) there was no detectable processing of the 2A peptide in the transformant pARshBle2ATRXH#8. Instead, some processing may exist in the transformant pARshBle2AhEGF#3. In this latter transformant, despite not detecting 3HA:hEGF, three specific bands with shBle antibody were observed: the full length protein, a higher band and a lower band. The lower band may correspond to shBle:2A (15.6 kDa) and induces us to think that some processing may exist in this transformant.

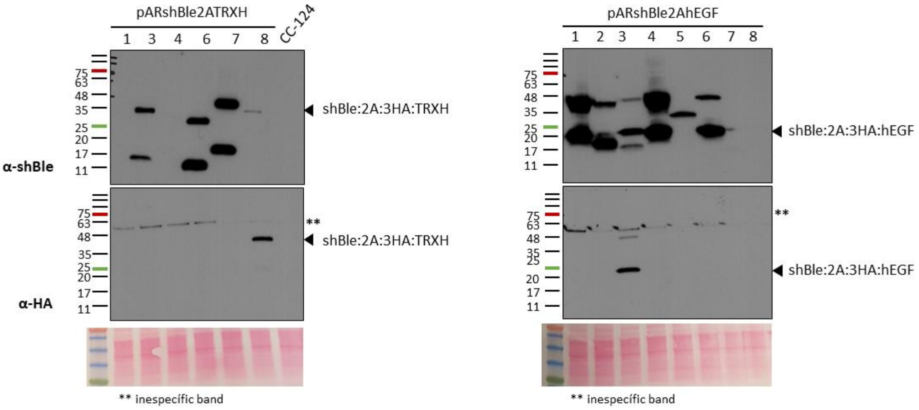
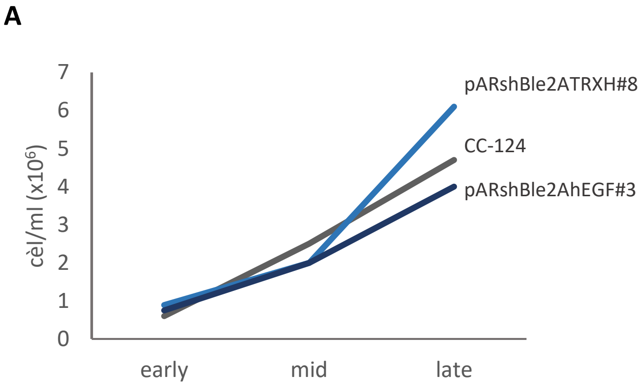

pARshBle2ATRXH
1 3 4 6 7 8 CC-124
75
63
48
35
25
20
17
11
α-shBle
α-HA
shBle:2A:3HA:TRXH
**
shBle:2A:3HA:TRXH
** inespecific band
pARshBle2AhEGF
1 2 3 4 5 6 7 8
shBle:2A:3HA:hEGF
**
shBle:2A:3HA:hEGF
** inespecific band

**Figure 4. Immunoblot of pARshBle2AhEGF and pARshBle2ATRXH high resistant transformants.** 50 µg TSP loaded in each lane. Upper panels correspond to immunoblot against shBle protein. Lower panels correspond to immunoblot against HA tag. Ponceau stained membranes showed as a loading control.

To determine if processing efficiency varies depending on culture phase, we followed the growth by cell density and took samples for immunoblot analysis at early, mid and late phase of growth for the two transformants selected and the wild type strain. Results are presented in figure 5.



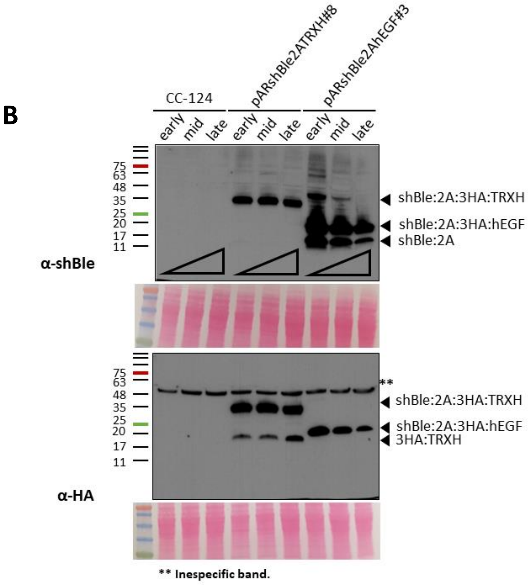

B
CC-124
pARshBle2ATRXH#8
pARshBle2AhEGF#3
early
mid
late
early
mid
late
early
mid
late
75
63
48
35
25
20
17
11
shBle:2A:3HA:TRXH
shBle:2A:3HA:hEGF
shBle:2A
α-shBle
**
shBle:2A:3HA:TRXH
shBle:2A:3HA:hEGF
3HA:TRXH
α-HA
** Inespecific band.

**Figure 5. Immunoblot of pARshBle2AhEGF and pARshBle2ATRXH selected transformants at different culture stages. A)** growth followed by cell counting. **B)** immunoblot analysis. Upper panel corresponds to immunoblot against shBle protein. Lower panel corresponds to immunoblot against HA tag. 50 µg TSP were loaded in each lane. Ponceau stained membranes showed as a loading control.

Whereas no differences were observed in the expression level of shBle:2A:3HA:TRXH at different growth stages, shBle:2A:3HA:hEGF was more expressed at early/mid compared to late log phase of culture. When analysing the transformant pARshBle2ATRXH#8, apart from the full-length protein, we detected a band that may correspond to TRXH:3HA (15.5 kDa) but we did not detect shBle:2A (15.6 kDa). In contrast, when analysing the transformant pAshBle2AhEGF#3, apart from the full-length protein, we detected a band that may correspond to shBle:2A (15.6 kDa) but we did not detect 3HA:hEGF (9.9 kDa). An explanation to this is that shBle:2A would be expressed under the detection level of the shBle antibody in the case of pARshBle2ATRXH#8 and 3HA:hEGF would be degraded in the case of pARshBle2AhEGF#3. Thus, there was processing but was indeed 20-30% (calculated shBle:2A compared to shBle:2A:3HA:hEGF and 3HA:TRXH compared to shBle:2A:3HA:TRXH). Independently of growth phase and transformant the main band always corresponded to the full-length protein. Therefore, processing was constant and lower than reported previously (Rasala et al. 2013).

To further corroborate the results, pARshBle2AmCherry was transformed into CC-124 and 96 independent transformants (initially selected at 15 µg/ml zeocin) were picked and retested for growth on increasing concentrations of zeocin to assess relative expression levels of shBle. Independent transformants able to grow at the maximum zeocin concentration tested (100 µg/ml) were selected as high expressing transformants, independent transformants able to grow at 50 µg/ml of zeocin were selected as medium expressing transformants and independent transformants able to grow in the presence of only 15 µg/ml of zeocin were selected as low expressing transformants. An immunoblot analysis was performed to test if zeocin resistance is correlated with transgene expression and to determine if the fusion protein was correctly expressed and cleaved in the selected transformants (figure 6).

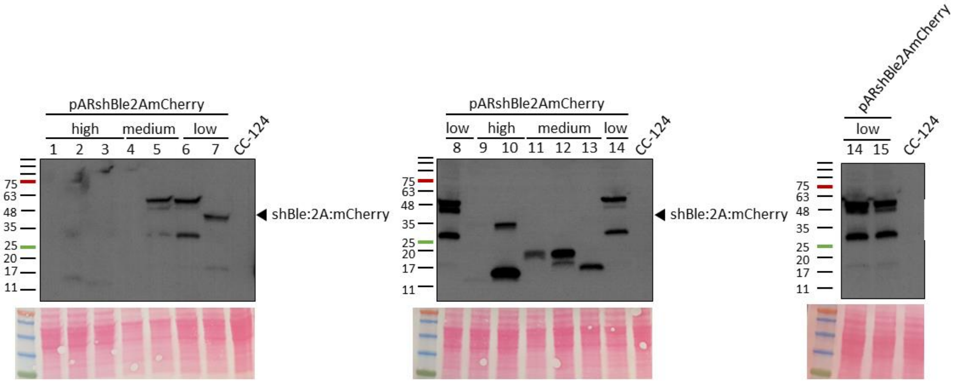

pARshBle2AmCherry
high
medium
low
1 2 3 4 5 6 7
CC-124
75
63
48
35
25
20
17
11
shBle:2A:mCherry
pARshBle2AmCherry
low
high
medium
low
8 9 10 11 12 13 14
CC-124
75
63
48
35
25
20
17
11
shBle:2A:mCherry
pARshBle2AmCherry
low
14 15
CC-124
75
63
48
35
25
20
17
11

**Figure 6. Immunoblot of pARshBlemCherry selected transformants.** 60 µg TSP loaded in each lane, primary antibody against shBle. Ponceau stained membranes are shown as a loading control.

The results were similar to the results obtained with pARshBle2AhEGF and pARshBle2ATRXH: proteins of varied sizes were observed between independent transformants. As all the selected clones are resistant to zeocin, all necessarily express the shBle protein, nevertheless we were not able to detect shBle protein in five out of ten high and medium expressing transformants. Besides, we were able to detect shBle protein in all the low expressing transformants. An explanation is poor processing and/or intermediate fusion proteins that may result in poor efficiency of shBle protein by preventing its biological function when compared to monomeric shBle: less monomeric shBle would confer resistance to high concentrations of zeocin.

To summarize, according to our results the minimal FMDV 2A peptide does not behave as expected (processing efficiency is lower than previously reported) and high expressing clones do not correlate with high expression of the processed protein.

### 1.1.2. Use of extended FMDV 2A peptide for processing of a dicistronic vector

The 2A peptide used in the first strategy presented was the one reported in initial studies to enhance protein expression in Chlamydomonas (Rasala et al.

2012, 2013). It is based on the minimal FMDV 2A peptide, which has been shown to be 45-50% less efficient than the extended FMDV 2A in allowing expression of full-length protein from the gene downstream of the 2A sequence (Donnelly et al. 2001). In Chlamydomonas, extended FMDV 2A peptide (39 amino acids) has been successfully used as a highly efficient system for the production of two proteins from a dicistronic gene (Plucinak et al. 2015). On the base of the low processing efficiency obtained with the short version of FMDV 2A peptide, we decided to test the same strategy but with the extended version of FMDV 2A ($2^E$). To this end the cassettes pARshBle2A$^E$TRXH and pARshBle2A$^E$hEGF (figure 7) were constructed by replacing minimal FMDV 2A for the extended version (2A$^E$) in the cassettes pARshBle2AhEGF and pARshBle2ATRXH.

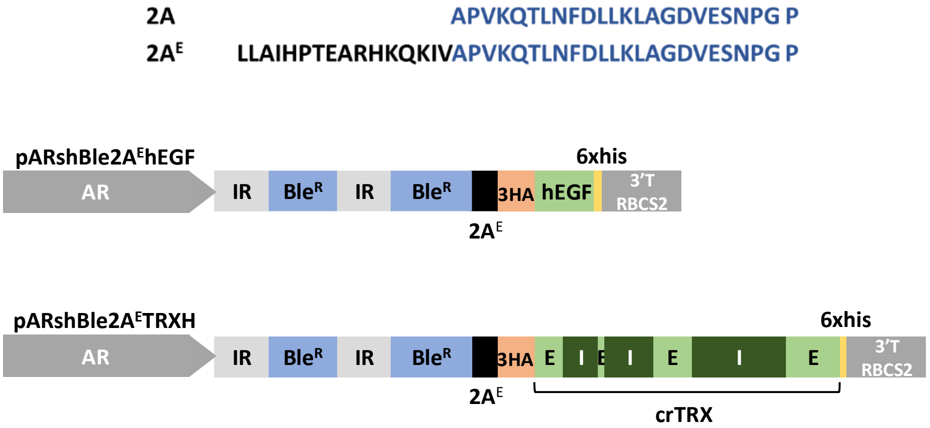

2A
APVKQTLNFDLLKLAGDVESNPG P
2A$^E$
LLAIHPTEARHKQKIVAPVKQTLNFDLLKLAGDVESNPG P
pARshBle2A$^E$hEGF
6xhis
AR
IR
Ble$^R$
IR
Ble$^R$
3HA
hEGF
3’T RBCS2
2A$^E$
pARshBle2A$^E$TRXH
6xhis
AR
IR
Ble$^R$
IR
Ble$^R$
3HA
E
I
E
I
E
I
E
3’T RBCS2
2A$^E$
crTRX

**Figure 7. Schematic comparison of minimal FMDV 2A (2A) and extended FMDV 2A (2A$^E$) sequences and schematics of pARshBle2A$^E$hEGF and pARshBle2A$^E$TRXH constructs.** AR: *HSP70-RBCS2* chimeric promoter, 3’T: 3’ region of *RBCS2*, IR: intron 1 *RBCS2*, I: intron, E: exon.

The constructs were transformed into the cell-walled Chlamydomonas strain CC-124 and an initial selection on TAP-zeocin 15 µg/ml was performed (considered low zeocin). 96 independent transformants were picked and retested for growth on increasing concentrations of zeocin to assess relative expression levels of shBle (figure 8). Independent transformants were selected: those able to grow up to 25 µg/ml zeocin were considered low resistant clones (3 for pARshBle2A$^E$hEGF and 6 for pARshBle2A$^E$TRXH), those able to grow up to 75 µg/ml zeocin were considered medium resistant clones (6 for pARshBle2A$^E$TRXH), and those able to grow up to 300 µg/ml zeocin were considered high resistant clones (3 for pARshBle2A$^E$hEGF and 6 for pARshBle2ATRXH).

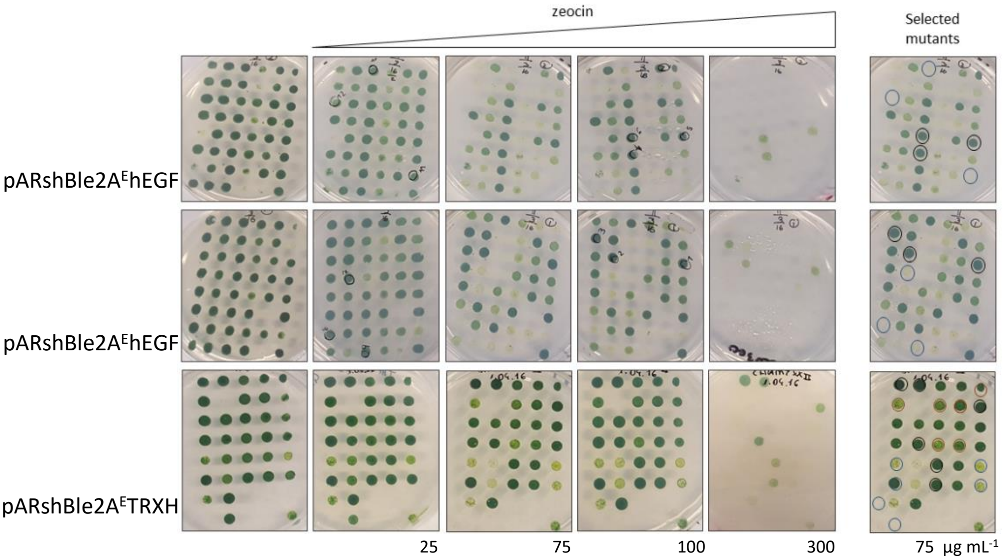


**Figure 8. Zeocin sensitivity test of pARshBle2A$^E$hEGF and pARshBle2A$^E$TRXH transformants.** Initial transformants were spotted onto TAP agar plates containing increasing concentrations of zeocin. The selection of transformants was assessed by growth: Blue marked transformants (low) grew up to 25 µg/ml of zeo, brown marked transformants (medium) grew up to 75 µg/ml of zeo and black marked transformants (high) grew at 300 µg/ml of zeo.

Once transformants had been selected, two immunoblots were performed using antibodies anti-shBle and anti-HA tag as previously to corroborate that high zeocin resistance is correlated with high transgene expression and to determine the correct expression and processing of the both peptides (figure 9). Thus, if the processing mediated by the extended 2A peptide is complete, the expected results would be a single band corresponding to 3HA:TRXH (15.5 kDa) or 3HA:hEGF (9.92 kDa) and a single band corresponding to shBle:2A$^E$ (17.4 kDa). The expected molecular sizes for the full length fusion proteins shBle:2A$^E$:3HA:TRXH and shBle:2A$^E$:3HA:hEGF were 36.7 kDa and 30.8 kDa respectively.

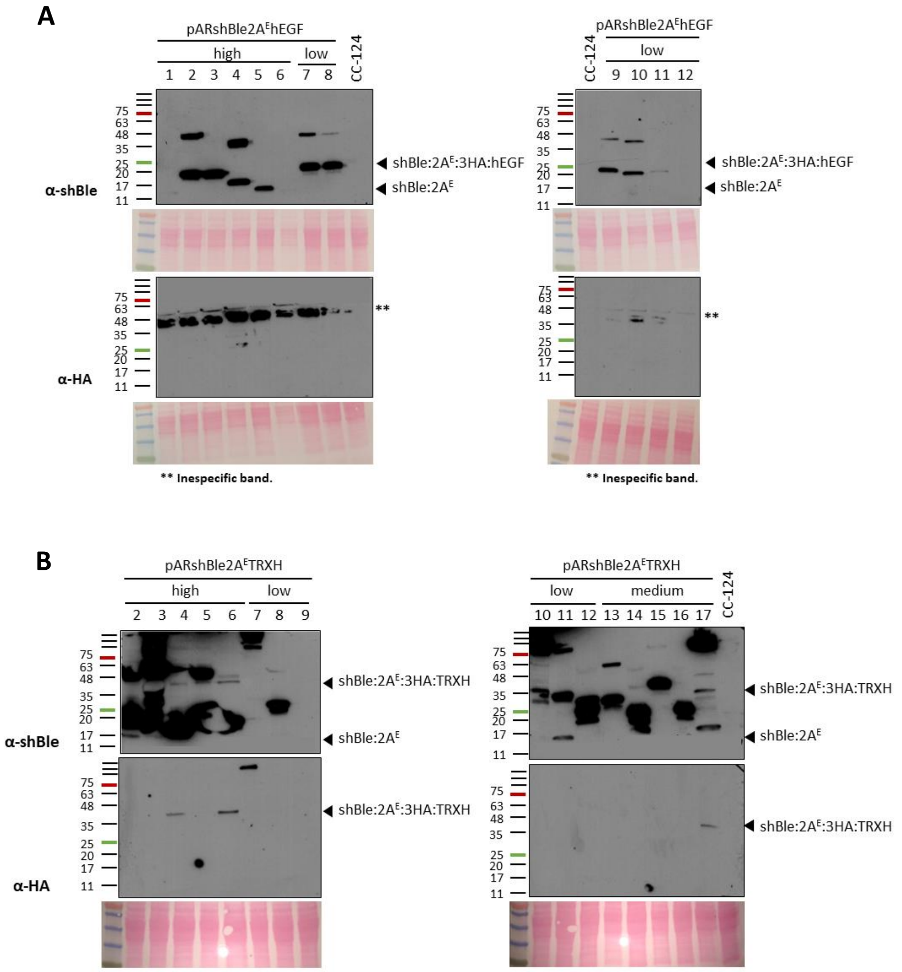

A
pARshBle2A$^E$hEGF
high
low
1 2 3 4 5 6 7 8
CC-124
75
63
48
35
25
20
17
11
α-shBle
shBle:2A$^E$:3HA:hEGF
shBle:2A$^E$
α-HA
**
** Inespecific band.
pARshBle2A$^E$hEGF
low
CC-124
9 10 11 12
B
pARshBle2A$^E$TRXH
high
low
2 3 4 5 6 7 8 9
shBle:2A$^E$:3HA:TRXH
shBle:2A$^E$
pARshBle2A$^E$TRXH
low
medium
10 11 12 13 14 15 16 17
CC-124

**Figure 9. Immunoblot of pARshBle2A$^E$hEGF and pARshBle2A$^E$TRXH selected transformants.** Upper panels correspond to immunoblot against shBle protein. Lower panels correspond to immunoblot against HA tag. Membrane stained with Ponceau showed as a loading control. **A)** pARshBle2A$^E$hEGF transformants immunoblot. 50 µg TSP were loaded in each lane. **B)** pARshBle2A$^E$crTRX transformants immunoblot. 100 µg TSP were loaded in each lane.

As in the previous strategy, we found specific bands with varied sizes between independent transformants. An HA specific band at the expected molecular size for the full-length protein was only present in 2 out of 16 independent transformants of pARshBle2A$^E$TRXH (36.7 kDa) and no HA specific band was found in pARshBle2A$^E$hEGF transformants. Importantly, no band or signal was detected in CC-124 negative control with any of the antibodies used. We expected that all the selected transformants express shBle since they are resistant to zeocin, however in 5 cases (out of 28) we were not able to detect the protein by immunoblot. The range of products or variability in transgene expression makes it difficult to establish a correlation between protein expression and resistance to zeocin. This may be due to poor processing and/or intermediate fusion proteins that may result in poor efficiency of shBle fusion protein in sequestering zeocin and conferring resistance to the antibiotic when compared to monomeric shBle.

To summarize, despite previous studies had reported the efficient processing of FMDV2A peptide sequences in Chlamydomonas (Rasala et al. 2012, 2014; Plucinak et al. 2015) we could not validate this strategy as a high-throughput screening for selecting high transgene expressing transformants. The host strain or growth conditions prior to lysate preparation used in our study may contribute to the variability in processing versus previous ones, but this remains to be determined. The protein fused to Ble2A may also influence processing of the fusion protein.

### 1.2. Use of a fusion protein consisting of a resistance gene and the protein of interest linked by TEV protease recognition sequence

In an analogous way than when testing FMDV 2A peptides, the approach consisting on the fusion comprising Ble$^R$ and GOI was further investigated with a TEV protease recognition sequence instead of a 2A peptide between the ShBle and protein of interest (figure 10). TEV protease (EC 3.4.22.44, Tobacco Etch Virus nuclear-inclusion-a endopeptidase) is a highly sequence-specific cysteine protease from Tobacco Etch Virus (TEV). Due to its high sequence specificity, it is frequently used for the controlled cleavage of fusion proteins *in vitro* and *in vivo* (Dougherty et al. 1989). The cassette included a 3HA tag (three copies of the haemagglutinin epitope tag, cloned C terminal of Ble$^R$), a 6xhis tag (cloned N terminal of Ble$^R$) and a TEV protease recognition sequence. All the coding region was adapted to the GC content and to the codon usage of the nuclear genome of Chlamydomonas. The coding region of the cassette was driven by *RPL23* (Cre04.g211800) cis regulatory elements (5' and 3'). We used these regulatory elements because it has been reported that they outperform the AR chimeric promoter and the promoter-terminator from the PSAD gene in terms of RP expression yields (Li et al. 2016; López-Paz et al. 2017).

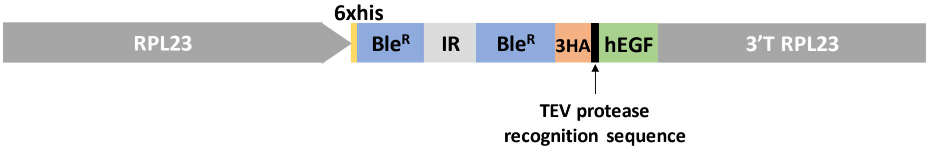


**Figure 10. Schematics of pPLshBleTEVhEGF construct.** RPL23: 5' *RPL23* cis regulatory elements including 5'UTR and promoter, 3'T RPL23: 3'UTR and terminator of *RPL23*, IR: intron 1 RCBS2.

The construct was transformed by electroporation into the cell-walled Chlamydomonas strain CC-124 and 45 transformants surviving on TAP agar 15 µg/ml zeocin plates (considered low zeocin) were picked and screened by their ability to grow at increasing concentrations of zeocin (figure 11). The eight clones able to grow at higher zeocin concentrations were selected.

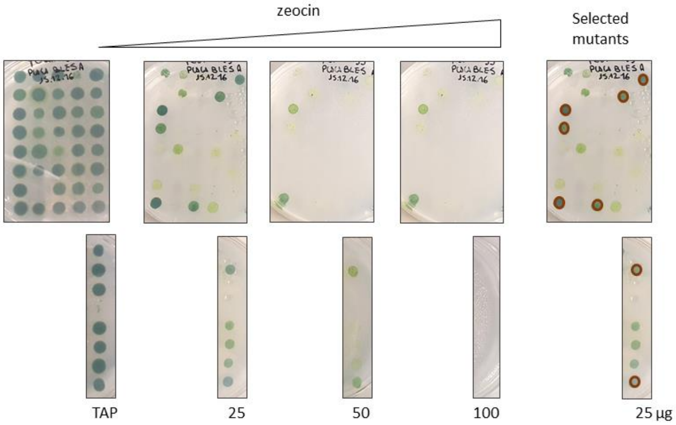


**Figure 11. Zeocin sensitivity test of pPLshBleTEVhEGF transformants.** Transformants were spotted onto TAP agar plates containing increasing concentrations of zeocin. Selection of high resistant clones was assessed by growth at increasing zeocin concentrations.

Once transformants had been selected, an immunoblot analysis was performed as explained previously to corroborate if high zeocin resistance is correlated with high transgene expression and to determine whether the fusion protein was expressed in our transformants (figure 12).

Since all the transformants are resistant to zeocin, we expected that all express shBle and that was corroborated by immunoblot. In total, only one out of six transformants presented a band at the expected molecular weight for the full-length protein (26.5 kDa): pPLshBleTEVhEGF#8. Apart from the transformant that presented a band at the expected molecular weight for the full-length protein (26.5 kDa), the bands detected are higher than this expected molecular weight. Besides, no band was detected in transformants #5 and #6 when hybridised against HA tag. One explanation is partial insertion of the DNA cassette: that would give a protein including, at least partially, our fusion protein and the product of a part of genomic DNA. Another explanation is the formation of protein aggregates. The latter hypothesis was discarded by repetition of the same immunoblots treating the samples with different reducing agents and boiling conditions resulting in the same results obtained before (results not shown).

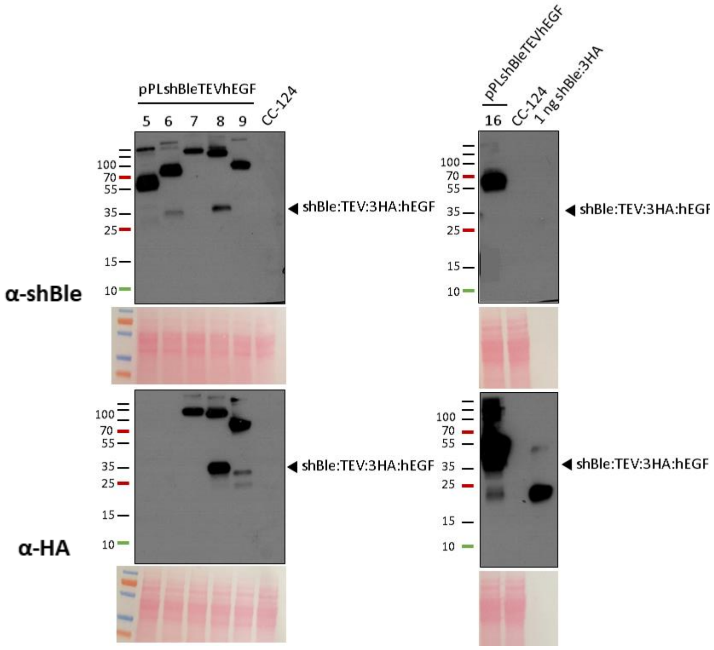

pPLshBleTEVhEGF
5
6
7
8
9
CC-124
100
70
55
35
25
15
10
shBle:TEV:3HA:hEGF
α-shBle
α-HA
pPLshBleTEVhEGF
16
CC-124
1 ng shBle:3HA
shBle:TEV:3HA:hEGF

**Figure 12. Immunoblot of pPLshBleTEVhEGF selected high expressing transformants.** 50 µg TSP were loaded in each lane. Upper panels correspond to immunoblot against shBle protein. Lower panels correspond to immunoblot against HA tag. Membranes stained with Ponceau are shown as a loading control.

To rule out the possibility of different processing due to culture phase we analysed RP expression throughout culture stages. We selected the transformant pPLshBlehEGF#8 for further analysis because it showed a band at the expected molecular weight for the full-length protein (26.5 kDa). Although molecular weight of the band detected in transformant pPLshBlehEGF#16 did not correspond to the expected molecular weight, we decided to further study this transformant because it expresses high yields of recombinant protein. We followed the growth by cell density and analysed five points from initial exponential phase to stationary phase of culture. No significant differences in the expression pattern were observed (figure 13).

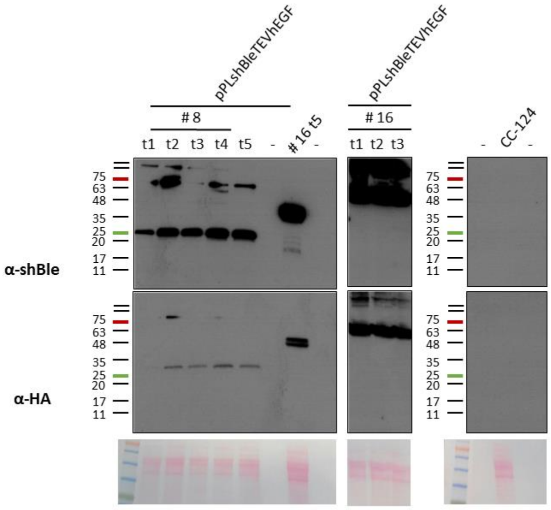

pPLshBleTEVhEGF
# 8
t1 t2 t3 t4 t5 -
# 16 t5
-
pPLshBleTEVhEGF
# 16
t1 t2 t3
-
CC-124
-
75
63
48
35
25
20
17
11
α-shBle
α-HA

**Figure. 13 Immunoblot of pPLshBleTEVhEGF selected transformants at different growth stages.** 50 µg TSP were loaded in each lane. Primary antibody against shBle (minutes of exposure) and against 3HA tag (seconds of exposure). Membranes stained with Ponceau are shown as a loading control.

To determine if partial insertions led to multiple fusion proteins we decided to analyse the integration of the cassette by PCR. We designed several oligonucleotides pairs to be able to determine if insertion of cassette was indeed full or partial. The results are shown in figure 14.

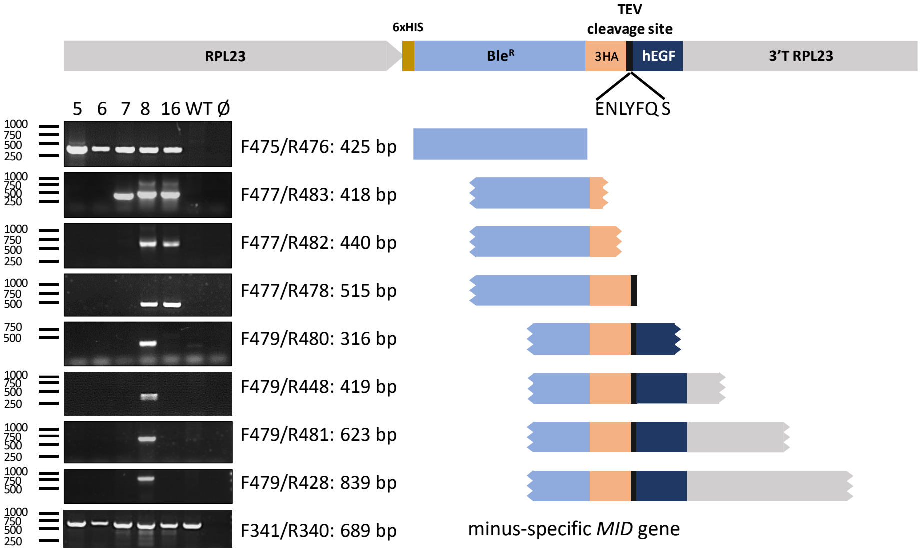
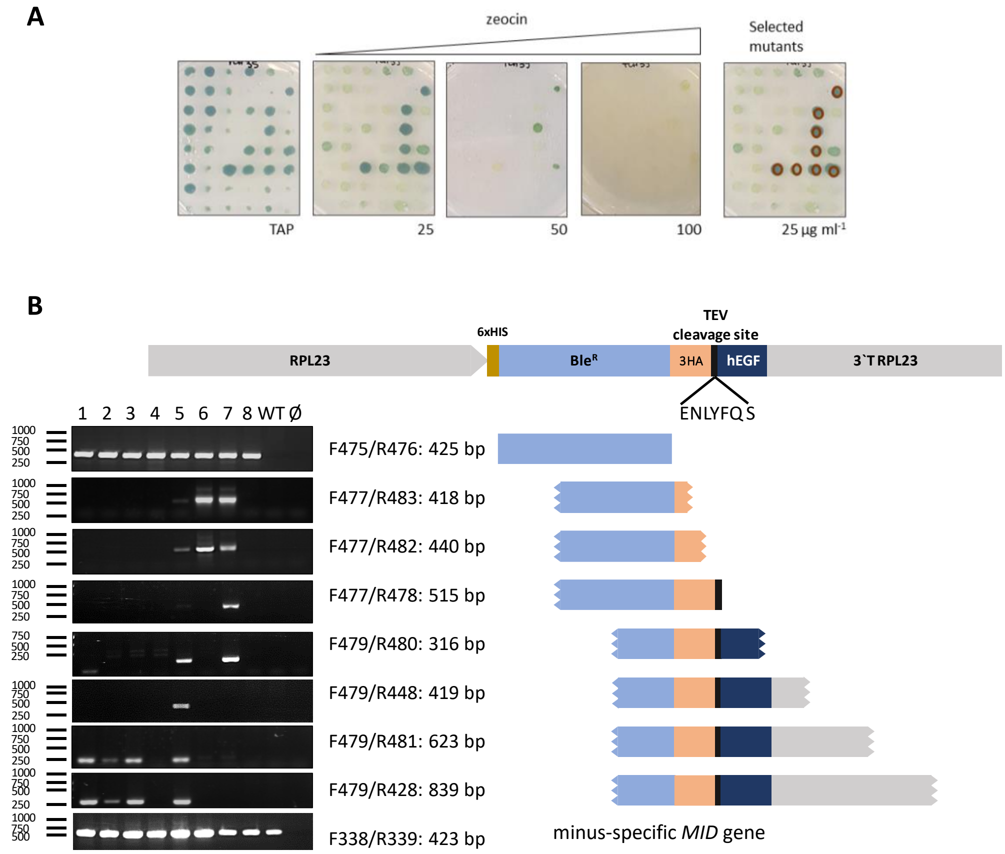


**Figure 14. PCR analysis of the integration of the cassette in selected high resistant pPLshBleTEVhEGF transformants.** Right panel: schematic representation of cassette and combination of oligos to analyse its integration. RPL23: 5' *RPL23* cis regulatory elements including 5'UTR and promoter, 3'T RPL23: 3'UTR and terminator of *RPL23*. Left panel: PCR results. The last image of amplification panel corresponds to the amplification of the mating type minus specific *MID* gene as a control of genomic DNA.

Since all the clones are resistant to zeocin and present bands by immunoblot against shBle, we expected insertion of Ble$^R$ in all of them and this was confirmed by PCR amplification. With the next three couples of oligos it was determined that all the clones in which we were able to detect HA specific signal by immunoblot have inserted at least one sequence encoding for one HA copy. Amplification by the next oligos (F479/R480) gives information of which clones have inserted from Ble$^R$ to the GOI (hEGF). There was correct amplification only in transformant pPLshBleTEVhEGF#8. Finally, only transformant pPLshBleTEVhEGF#8 showed the expected sizes in the amplification of three different oligonucleotides pairs for the insertion of 3'UTR.

Additionally, we transformed pPLshBleTEVhEGF into the alternative wild type strain 21gr to corroborate the results. In this case, 48 independent transformants (initially selected on TAP-zeocin 15 µg/ml) were picked and screened by its ability to grow at increasing concentrations of zeocin (figure 15-A) and the eight clones surviving to higher zeocin concentrations were selected. As expected, when analysed by immunoblot against shBle, proteins of varied sizes were observed between transformants (figure 15-B). Again, we observed specific signal corresponding to diverse low molecular weights. However, only one out of eight transformants presented a specific HA band at the expected molecular size for the full-length protein (26.5 kDa): pPLshBleTEVhEGF#7. Moreover, this transformant also presented a lower band when hybridised against shBle. Apart from this transformant, we also detected specific HA bands in transformants #1, #6 and #8. In the transformant #1 the band corresponded to a smaller molecular weight than the expected for the full-length protein. In the transformants #6 and #8 the bands corresponded to a higher molecular weight than the expected for the full-length protein. Given the results obtained with CC-124 strain the more likely explanation to these results is the partial insertion of the cassette. The translation of adjacent genomic sequences to the partial insertion may explain the detected proteins with higher molecular weights. To further confirm these results, we analysed the integration of the cassette by PCR (figure 15-C). As expected, all the transformants had inserted the complete Ble$^R$ gene. Transformants #6, #7 and #8 were detected by immunoblot against HA so they necessarily had inserted at least one sequence encoding for one HA copy. However, that was corroborated only in transformants #6 and #7, not in transformant #8. Only transformants #5 and #7 had inserted cassette from Ble$^R$ to hEGF. Finally, the obtained PCR products between Ble$^R$ and 3'UTR had lower sizes than expected for transformants #1, #2, #3 and #5, indicating a deletion of part of the cassette. Genomic DNA of transformant #1 was sequenced and it showed a deletion between Ble$^R$ and 3'UTR. As transformants #1, #2, #3 and #5 shared the size of the PCR products between Ble$^R$ and 3'UTR it is expected that the deletion introduced would be the same.


A
zeocin
Selected mutants
TAP
25
50
100
25 µg ml-1
B
TEV cleavage site
6xHIS
RPL23
BleR
3HA
hEGF
3`T RPL23
ENLYFQ S
1 2 3 4 5 6 7 8 WT Ø
1000
750
500
250
F475/R476: 425 bp
F477/R483: 418 bp
F477/R482: 440 bp
F477/R478: 515 bp
F479/R480: 316 bp
F479/R448: 419 bp
F479/R481: 623 bp
F479/R428: 839 bp
F338/R339: 423 bp
minus-specific MID gene

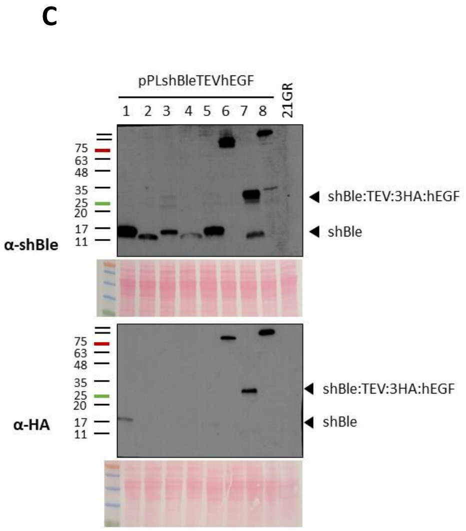

C
pPLshBleTEVhEGF
21GR
1 2 3 4 5 6 7 8
75
63
48
35
25
20
17
11
α-shBle
shBle:TEV:3HA:hEGF
shBle
α-HA
shBle:TEV:3HA:hEGF
shBle

**Figure 15. Analysis of pPLshBleTEVhEGF 21gr transformants. A Zeocin sensitivity test.** Transformants were spotted onto TAP agar plates containing increasing concentrations of zeocin. Selection of high resistant clones was assessed by growth at increasing zeocin concentrations. **B PCR analysis of the integration of the cassette in selected high resistant transformants of pPLshBleTEVhEGF.** Right panel: schematic representation of cassette and combination of oligos to analyse its integration. RPL23: 5' *RPL23* cis regulatory elements including 5'UTR and promoter, 3'T RPL23: 3'UTR and terminator of *RPL23*. Left panel: PCR results. The last image of amplification panel corresponds to the amplification of the mating type minus specific *MID* gene as a control of genomic DNA. **C Immunoblot of pPLshBleTEVhEGF selected high expressing transformants.** 50 µg TSP were loaded. Upper panel correspond to immunoblot against shBle protein. Lower panel correspond to immunoblot against HA tag. Membranes stained with Ponceau are shown as a loading control.

To summarize, as a result of this strategy we obtained the clone pPLshBleTEVhEGF#8 (strain CC-124) and the maximum yield achieved with this transformant was quantified (figure 16). The expression in transformant 21gr-pPLshBleTEVhEGF#7 was also quantified because it has inserted the cassette from Ble$^R$ to hEGF and thus it expreses the complete fusion protein. Despite not having inserted the complete cassette, transformant CC-124-pPLshBleTEVhEGF#16 was included due to its high yields. We had previously seen that the maintenance of transformants in the presence of zeocin leads to a higher expression of shBle (results not shown) and thus we included each transformant maintained with and without zeocin for approximately 2 months to quantify the highest yields obtained. Expression yields were estimated by densitometry based on immunoblots (figure 16, table 5).

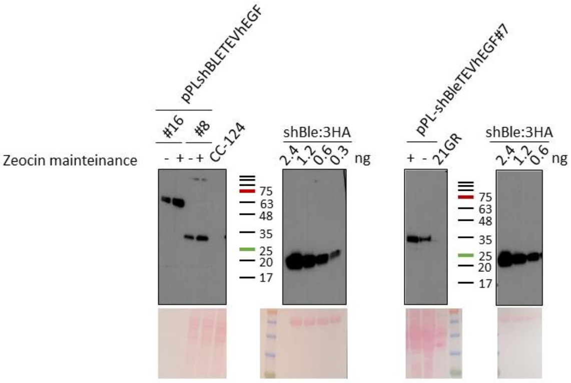


**Figure 16. Immunoblot used to estimate expression levels by densitometry in selected transformants.** Immunoblot of #16: 1 µg TSP, #8: 25 µg TSP, #7: 50 µg TSP. -: maintained on TAP agar plates, +: maintained on TAP zeocin agar plates; primary antibody against HA tag. Membrane stained with Ponceau showed as a loading control.

**Table 5. Estimation of expression in selected transformants.** Densitometry based on figure 16. -: maintained on TAP agar plates, +: maintained on TAP zeocin agar plates.

| Transf. | ng detected | µg RP/L culture | % TSP |
|---|---|---|---|
| **#8 +** | 0.605 | 1.83 | 0.0024 |
| **#8 -** | 0.361 | 1.02 | 0.0014 |
| **#16 +** | 1.837 | 85.76 | 0.1837 |
| **#16 -** | 0.888 | 39.45 | 0.0888 |
| **#7 +** | 0.632 | 0.91 | 0.0013 |
| **#7 -** | 0.409 | 0.76 | 0.0008 |

We corroborated that maintenance in presence of zeocin for two months led to duplication of the expression level. RP yield of our selected transformant pPLshBleTEVhEGF#8 is approximately 1 µg/L culture. This expression level is low and hampers the downstream processing being too low for any industrial process. Finally, the expression level of the chimeric fusion protein expressed by transformant pPLshBleTEVhEGF#16 has been estimated in 80 µg/L culture.

To confirm expression of hEGF in transformant pPLshBleTEVhEGF#8 a polyclonal antibody against hEGF was used (figure 17).

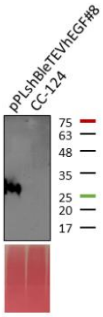


**Figure 17. Immunoblot of hEGF fusion protein obtained transformant.** Primary antibody against hEGF. 12 µl of total soluble protein (100X) were loaded.

### 1.2.1. Identification of a highly expressed chimeric protein resulting from the partial insertion of exogenous DNA cassette.

The expression level of the transformant pPLshBleTEVhEGF#16 has been estimated in 80 µg/L culture. This transformant was selected with the aim of analysing if it higher yield could be due to some stability domain, regulatory region or other characteristics that could be used in an expression strategy. For example, a successful strategy to enhance stability of recombinant proteins has been demonstrated based on $(SP)_n$ glycomodules (Shpak et al. 1999; Xu et al. 2007; Ramos-Martinez et al. 2017). A monoclonal anti-HA antibody was used to immunoprecipitate protein that was further purified from SDS PAGE gel and subjected to mass spectrometry analysis after digestion with trypsin. The result of the immunoprecipitation is shown in figure 18.

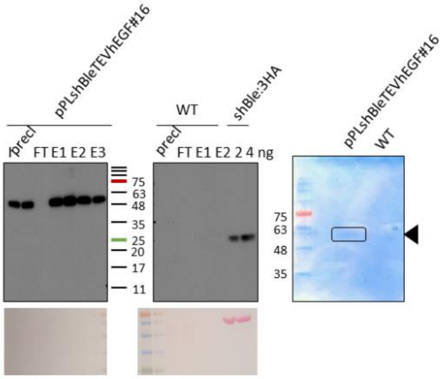

pPLshBleTEVhEGF#16
I
precl
FT E1 E2 E3
75
63
48
35
25
20
17
11
WT
shBle:3HA
precl
FT E1 E2 2 4 ng
pPLshBleTEVhEGF#16
WT
75
63
48
35

**Figure 18. Immunoprecipitation of shBle-X chimeric protein.** A monoclonal antibody against HA tag was used to immunoprecipitate shBle-X chimeric protein from total extract. Left panel: Analysis of immunoprecipitation by immunoblot. WT: wild type CC-124 strain, I: input, precl: input after preclearing step (incubation with resin without coupled antibody), FT: flow through, E1→E4: sequential elutions. Membrane stained with Ponceau are shown as a loading control. Right panel: SDS PAGE of IP product, marked band excised for sequencing.

The result of protein analysis by mass spectrometry was a list of 54 identified proteins from which 20 proteins that corresponded to contaminations were discarded. Regarding our fusion protein, peptides from shBle to TEV protease recognition sequence were detected confirming the PCR analysis and immunoblot results. This coverage corresponds approximately to 19.6 kDa of the designed fusion protein (shBle:TEV:3HA:hEGF) and since the immunoprecipitated protein analysed (chimeric protein from now on) has a molecular weight of 50 kDa, the protein fragment to complete it should correspond to 30 kDa (figure 19). It is important to remark that the partial insertion of the cassette may not be in frame with a coding region and thereby it wouldn't translate for any protein.

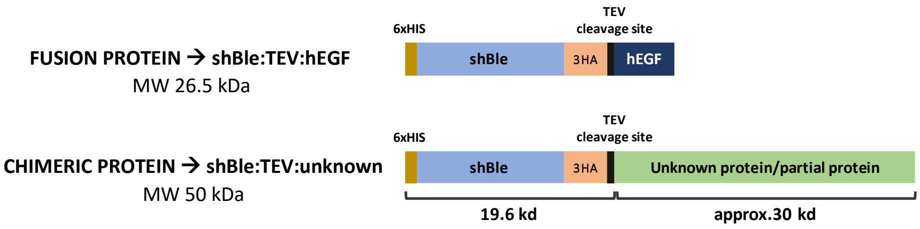

FUSION PROTEIN → shBle:TEV:hEGF
MW 26.5 kDa
6xHIS
TEV
cleavage site
shBle
3HA
hEGF
CHIMERIC PROTEIN → shBle:TEV:unknown
MW 50 kDa
6xHIS
TEV
cleavage site
shBle
3HA
Unknown protein/partial protein
19.6 kd
approx.30 kd

**Figure 19. Schematics of fusion protein (designed) and chimeric protein found in pPLshBleTEVhEGF#16.**

One of the characteristics described on the sequencing results are PSM (peptide spectrum matches) that displays the total number of identified peptide sequences for a determined protein, including those redundantly identified. According to that, the two parts of the chimeric protein were expected to be equally represented in the results and thus present similar PSM: as the known part of the fusion protein corresponds to 19.6 kDa and has 17 PSM, the other part of the fusion protein (that represents approximately 30 kDa) was expected to present a similar but slightly bigger PSM number. On this basis, those proteins with a PSM similar to 17 were further studied. It is necessary to highlight that we are accepting errors with this assumption because detection of peptides is sequence dependent: depending on the cutting sites of Trypsin for a determined sequence more or less information will be lost because too small or too big peptides won't be detected. This criterion resulted in a list of five proteins (see table 6).

**Table 6. Candidates to be the unknown part of the chimeric protein expressed by the transformant pPLshBleTEVhEGF#16.**

| Protein | Uniprot code | PSM | Mw | Coverage (%) |
|---|---|---|---|---|
| **Ribulose bisphosphate carboxylase large chain** | RBL_CHLRE | 23 | 52.5 | 31.16 |
| **Flagellar associated protein** | A8I7W0_CHLRE | 27 | 56.7 | 43.39 |
| **Predicted protein (Fragment)** | A8ICI8_CHLRE | 21 | 27 | 54.29 |
| **Eukaryotic translation elongation factor 1 alpha 1** | A8HX38_CHLRE | 21 | 50.8 | 38.01 |
| **Adenosylhomocysteinase** | A8IXE0_CHLRE | 17 | 52.7 | 30.02 |

The distribution of identified peptides was analysed in the five selected proteins (figure 20).

>sp|P00877|RBL_CHLRE Ribulose bisphosphate carboxylase large chain
MVPQTETKAGAGFKAGVKDYRLTYYTPDYVVRDTDILAAFRMTPQLGVPPEECGAAVAAESSTGTWTTVWTDGLTSL
DRYKGRCYDIEPVPGEDNQYIAYVAYPIDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPAYVKTFVGPPHGIQVE
RDKLNKYGRGLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDRFLFVAEAIYKAQAETGEVK
GHYLNATAGTCEEMMKRAVCAKELGVPIIMHDYLTGGFTANTSLAIYCRDNGLLLHIHRAMHAVIDRQRNHGIHFRVL
AKALRMSGGDHLHSGTVVGKLEGEREVTLGFVDLMRDDYVEKDRSRGIYFTQDWCSMPGVMPVASGGIHVWHMP
ALVEIFGDDACLQFGGGTLGHPWGNAPGAAANRVALEACTQARNEGRDLAREGGDVIRSACKWSPELAAACEVWKE
IKFEFDTIDKL

>tr|A8I7W0|A8I7W0_CHLRE Flagellar associated protein
MASRGLFAALCAACLAVASAAPLRNSADCVNVDLLKVTGVNVFPTTRQISSSSVQFLQDRTTVSVAQNFEVSYYSTFKV
VTNLYTFAKETYVLYQCGTAAPNPADYGLPAATKVLAVPLTAVSVQDTTVLNFMKILNVLDRVAYVTPYAVEPCMQALA
GGSCNRTSDNAVADKIDGKFVGFSVKPSDPQAITFTATADPGPLNRAEYIKYVATFFNREAAANTAFNAIVSAYNANKQ
LAATYLSKNSAVSPLVAWIYYQPKYDSATYKAPEQVVLYFNAYRRQYVQDAGGRLHDLAALNAQYGNNNTITVSGSQL
VFTDLPNTLPVLRTILASVDVVIDESFYDGKNPTQITLDDFRALYGLTDAAALTDLKFLANQRLLRLDGSANPGGTTAWFE
EAVARPDIVGPTRVDDVLHDLVSAFLPNALLSIGDTVPAKLVRNMFASNPGRVVLSSAAQCPAAQATCGAPASPVCPA
VYLDCQGRLTAATDSQPCAPSCSASPSNNAPNGAGGAASRAAAQPLLLVALLVAALAVLGWGL

>tr|A8ICI8|A8ICI8_CHLRE Predicted protein (Fragment)
MSEVTPPEPGASDYEIYKEGFHDRLSRYWPEPREQLHAGGWQRTNQMLGDIRHLAAVHGCGQIKQALDLCCGEGAT
AVYVAAKEGWEVTGVEIVPTATKVARARASAYGSVVRVRIAPHGTDGAVQLGLSYTAVEDLVRFAAASIHYMPLPSNSY
DLVYGQDPDGLANEHRIHAFKEVLRVLRPGGLFYFWHHWIPGPGWPAALLDEYGADPVTGSPRLSYEQYVEDLQSTLV
TSRQVHGHRGFL

>tr|A8HX38|A8HX38_CHLRE Eukaryotic translation elongation factor 1 alpha 1
MSEGKEHLSIVICGHVDSGKSTTTGRLLFELGGIPERELEKLKEEAAALGKSSFAFAFYMDRAKEERERGVTIACTTKEFFT
DRWHYTIIDAPGHRDFIKNMISGAAQADVCLLMVPADGNFTTAIQKGDHKAGEIQGQTRQHARLINLLGVKQLIVGVN
KMDSDTAGYKKERYDEIANEMRHMLVRVGWKDDFVNKSVPILPISGWLGDNLITKSTNMTWYSGQEVVNLKGEKIQ
VHTLLDALNSFVVVPERKTDAPLRLPISGAYKIKGVGDVLAGRVEQGVVKPGDEVIFLPTHTTANPCTGKVFTVEMHHK
RVDKAGPGDNVGMNIKGLDKGNMPRTGDVMILKSDQTLKIVKDFTAQIQTLDIPGEVKKGYSPIGFVRCGRSACRISGI
NWKVGKETGGKKMENPVGLKANEMAEVVYEPTQPLIVDSFKNCEGLSRIAFLDGNTAVMLGKVVSVTHK

>tr|A8IXE0|A8IXE0_CHLRE Adenosylhomocysteinase
MALSVDHVNGREYKVKDIAEADFGRLEIDLAEAEMPGLMACRSEFGPAQPFKGAKITGSLHMTIQTAVLIETLTALGAE
VRWCSCNIFSTQDHAAAAIARDSAAVFAWKGETLEEYWWCTEQALTWAGCDGPDLLVDDGGDATLLIHEGLKAEEQ
YAKDGTLPNPDSTDNAEFKIVLGLIRDSIQKDATKWTRMSKKVMGVSEETTTGVKRLYEMQANGSLLFPAINVNDSVTKS
KFDNVYGCRHSLPDGIMRATDVMIAGKTAFIAGYGDVGKGCASAMKAAGARTIVSEIDPICALQAAMEGYQVQPLEDCL
ETADIFITTTGNKDIIMAEHMAKMKNNAIVGNIGHFDNEVDMAGLYAWPGIKRQNIKPQVDRFIFPDGHGVIVLAEGRLL
NLGCATGHPSFVMSCSFTNQVIAQLELWNERSTGRYEKKVYVLPKHLDEKVAALHLPKLGVKLTKLSADQAAYINVPVDG
PYKPAHYRY

**Figure 20. Analysis of the distribution of identified peptides along the five Ble:unknown putative candidates.** Identified peptides in the assay are highlighted in yellow. In the case of Adenosylhomocysteinase the suspected part of the protein present in the results is highlighted in bold blue letters.

As shown in figure 20, all the proteins but Adenosylhomocystenaise had detected peptides covering all its sequence. The partial Adenosylhomocystenaise (highlighted in bold blue letters in figure 20) would correspond to 35.5 kDa and would meet the expected size to be a candidate for the unknown part of the chimeric protein. From the remaining four proteins, all but A8ICI8_CHLRE Predicted protein (Fragment) are too big to represent the unknown part of the chimeric protein. However, the

A8ICI8_CHLRE Predicted protein (Fragment) meets the approximate molecular weight to be a candidate for the unknown part of the fusion protein. Thus, the list was reduced to two candidates: A8ICI8_CHLRE Predicted protein (Fragment) and Adenosylhomocysteinase. We designed reverse oligos that will amplify for the sequence that transcribes for the first peptide (5’) detected of both candidates to perform a PCR analysis (see figure 21).

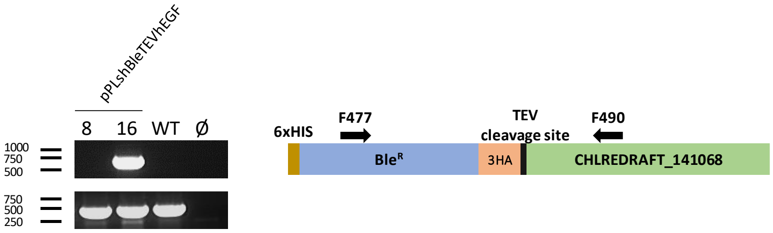


**Figure 21. PCR analysis to determine the sequence transcribing for the chimeric protein expressed in clone pPLshBleTEVhEGF#16.** Upper panel corresponds to PCR result for F477/R490. Lower panel corresponds to amplification of minus-specific *MID* gene (F341/R340) as a control for genomic DNA.

As seen in figure 21 the amplification of A8ICI8_CHLRE predicted protein gave a specific PCR product only in transformant pPLshBleTEVhEGF#16. Thus, the chimeric fusion protein is integrated by shBle:3HA:TEV:A8ICI8 (predicted protein). Amplification with the reverse oligo for Adenosylhomocysteinase (F497) did not give a specific PCR product (results not shown).

It is possible that the native protein region increases the stability or solubility of the fusion protein, but it remains to be determined. On the other hand, the gene encoding for this protein has six introns, and one is 997 bp length. Given that the average for introns length in Chlamydomonas is 373 bp (Merchant et al. 2007) this intron represents an interesting regulatory region to test. We did not further pursue the study of any domain or regulatory region, but the analysis of this chimeric protein corroborates that the partial insertion of the cassette leads to different transgenes and prevents the selection of high expressing transformants of the fusion protein.

## 2. Developing a high throughput screening method: fusion of luciferase reporter to the protein of interest

### 2.1. Fusion of 3XHA tagged Gaussia Luciferase reporter to the protein of interest to enable screening of transformants

In the light of the levels reported when RP is expressed from Chlamydomonas nucleus and targeted to medium (1.5 mg gLuc/L in wild type cells (Lauersen et al. 2013a, b) and 12 mg gLuc/L in UVM4 strain (Lauersen et al. 2015a)) in addition to the advantages that secretion offers (reduced exposure to proteases and easy recovery from medium), we decided to test an strategy based on secretion. As previously introduced, the high variability in the expression of the same cassette from independent clones of Chlamydomonas makes it necessary to screen a high number of transformants to select those which express RP to high levels. We need a marker or a reporter gene and in the present approach a fusion protein comprising a reporter gene and a GOI was tested. Gaussia luciferase (gLuc) catalyses the oxidation of the substrate coelenterazine resulting in light emission and was selected as a reporter gene due to the fact that it has been previously shown to have potential for strong bioluminescence and it can be secreted when expressed in the nucleus of Chlamydomonas (Ruecker et al. 2008; Shao and Bock 2008). With the aim of developing an *in vivo* assay to screen transformants in a high throughput format, a cassette was designed for the expression of a fusion protein consisting of gLuc and hEGF that would be secreted to medium or periplasmic space.

The designed plasmid pAR-ARSss-glucTEVhEGF-HygB$^R$ (see figure 22) carries two cassettes: i) the gene *AphVII'* conferring resistance to hygromycin B driven by β-tubulin cis regulatory elements (5' and 3') and ii) the cassette ARSss–gLuc-3HA tag-TEV-hEGF-6xhis tag where gLuc means Gaussia Luciferase, ARSss means Chlamydomonas Arylsulfatase 1 secretion signal and TEV means

Tobacco Etch Virus protease recognition sequence, the whole cassette driven by AR promoter and *RBCS2* 3'UTR. In place of the predicted secretion signal of gLuc, it carries Chlamydomonas arylsulfatase 1 secretion signal (ARSss), which has been successfully used to secrete RP from the nucleus of Chlamydomonas (Rasala et al. 2012). The expression of the former cassette would allow an initial selection of colonies by resistance to antibiotic. The expression of the latter cassette would result in a fusion protein directed to the periplasmic space or to medium (depending on if a cell walled strain or cell wall-less strain is used) that would allow a high-throughput *in vivo* screening of transformants: high gLuc activity would correspond to high expression of the whole fusion protein. Moreover, this fusion protein would be easily purifiable from medium thanks to a 6xhis tag and further downstream processing would be performed by TEV protease cleavage to obtain hEGF.

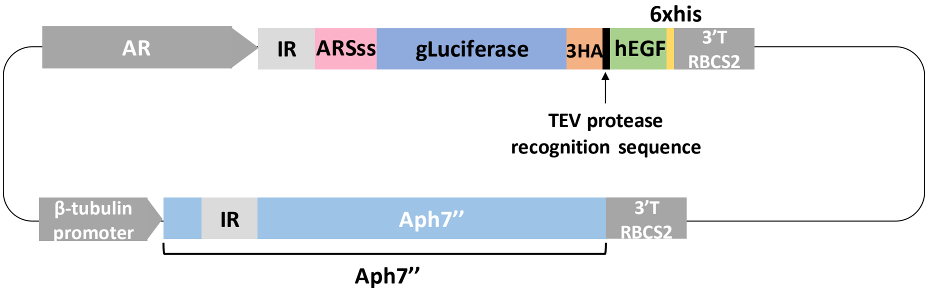


**Figure 22. Schematics of pAR-ARSss-glucTEVhEGF-HygB$^R$ construct.** AR: *HSP70-RBCS2* chimeric promoter, ARSss: arylsulfatase 1 secretion signal. IR: intron 1 *RBCS2*, 3'T RBCS2: 3' region of *RBCS2*.

The cell walled strain CC-124 was transformed with the plasmid pAR-ARSss-glucTEVhEGF-HygB$^R$ and 192 initial transformants (resistant to 60 µg hygromycin) were tested for gLuc activity. Positive transformants were divided in high expressing (named H) or low expressing (named L). Results are shown in figure 23.

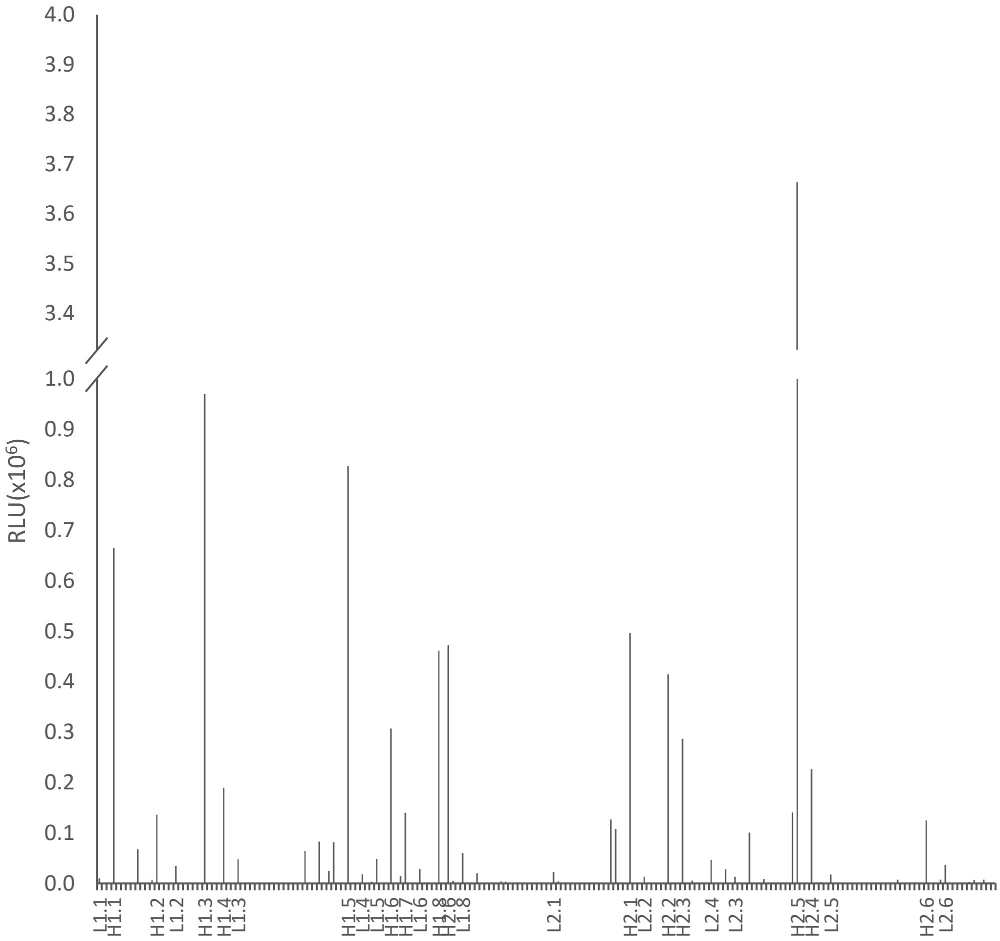


**Figure 23. gLuc activity test results of pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ initial transformants.** The assay was performed with 100 µl of culture (including both cells and culture medium) at log phase of growth. Only transformants selected for further analysis are identified. L:low expressing transformants, H: high expressing transformants. RLU: relative luminescence units.

To confirm the expression of initially obtained transformants, in a subsequent analysis cells were separated from culture medium by centrifugation to examine the activity present either in medium or cells. The gLuc activity results of the positive transformants are shown in figure 24.

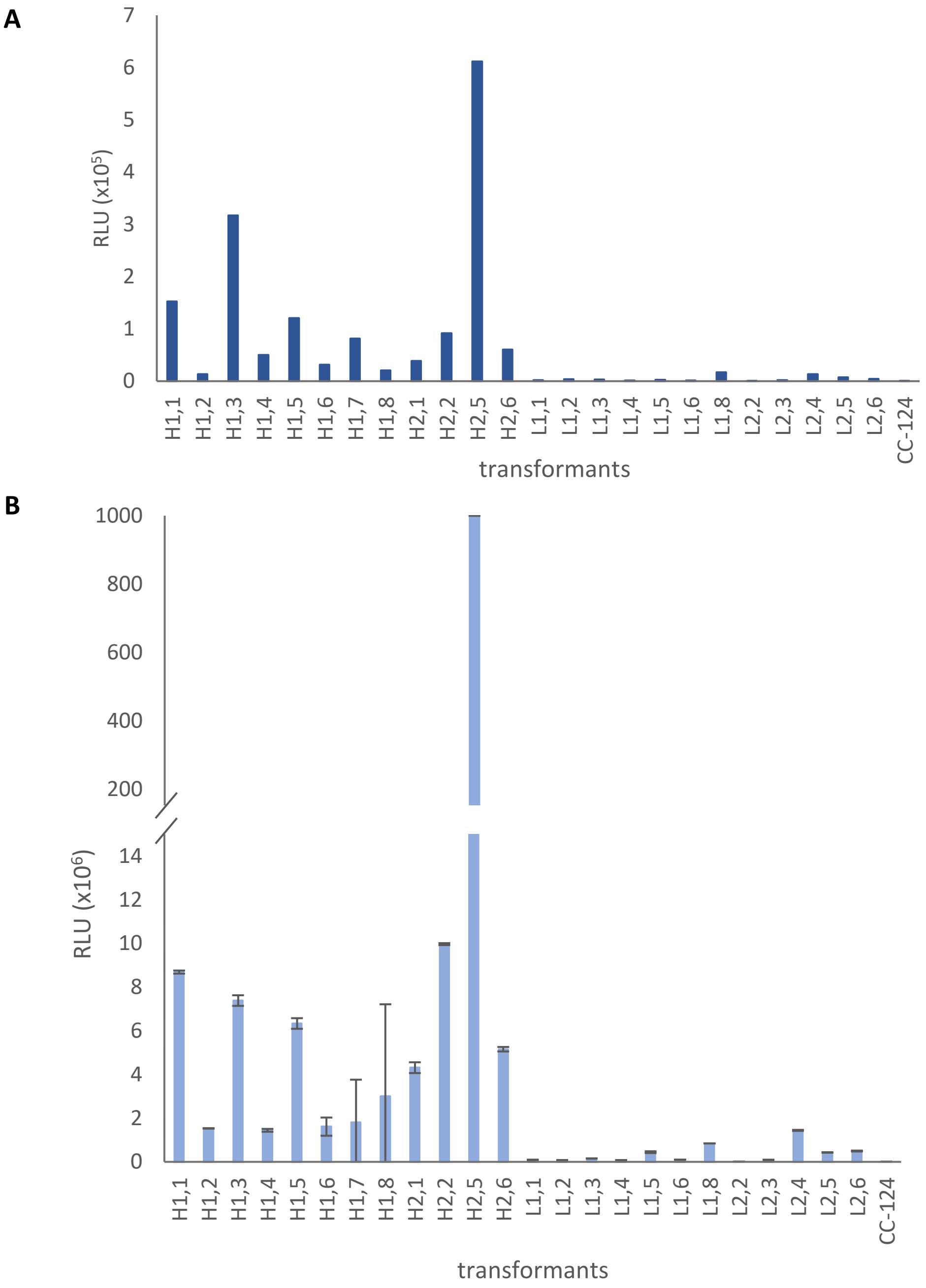


**Figure 24. gLuc activity assay results for the selected pAR-ARSss-glucTEVhEGF-HygB$^R$ transformants**. H: high, L: low. **A)** Intracellular gLuc activity. Assay was performed with cell extract corresponding to 100 µl of culture, **B)** Extracellular gLuc activity. Assay was performed with 100 µl of culture supernatant, error bars show the standard deviation of two technical replicates.

Given that CC-124 is a cell walled strain, fusion protein should be directed to periplasmic space thanks to ARSss and gLuc activity mainly found in cells. However, as it can be seen in figure 24, gLuc activity was present in medium rather than in cells in all positive transformants. It is likely that the fusion protein accumulates in the periplasm and then it is released to culture medium when cells divide to complete its asexual reproduction cycle.

As previously exposed, the expression strategy is based on the expression of a fusion protein (gLuc:3HA:TEV:hEGF:6xhis) and thus we expected that the gLuc activity would be directly related to fusion protein expressed. To corroborate this hypothesis, positive transformants were analysed by immunoblot (figure 25).

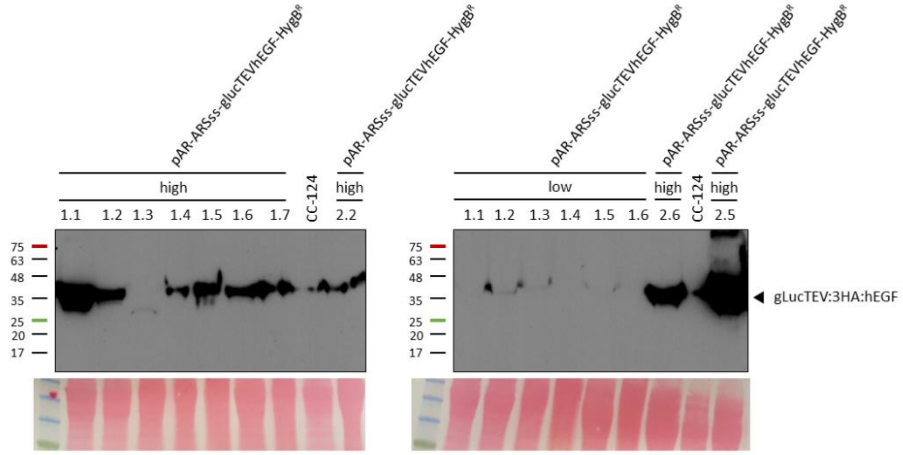

pAR-ARSss-glucTEVhEGF-HygB$^R$
pAR-ARSss-glucTEVhEGF-HygB$^R$
high
CC-124
high
1.1 1.2 1.3 1.4 1.5 1.6 1.7 2.2
75
63
48
35
25
20
17
pAR-ARSss-glucTEVhEGF-HygB$^R$
pAR-ARSss-glucTEVhEGF-HygB$^R$
pAR-ARSss-glucTEVhEGF-HygB$^R$
low
high
CC-124
high
1.1 1.2 1.3 1.4 1.5 1.6 2.6 2.5
gLucTEV:3HA:hEGF

**Figure 25. Expression analysis of pAR-ARSss-glucTEVhEGF-HygB$^R$ selected transformants by immunoblot.** Primary antibody anti HA-tag, 24 µl of TSP (100X relative to culture) were loaded in each lane. Ponceau stained membranes are shown below exposure as a loading control.

Focusing in the immunoblot results of soluble extracts (figure 25), a direct relation between gLuc activity and fusion protein detected by immunoblot was confirmed, being H2.5 the highest expressing transformant. Contrary to the observed with the fusion protein comprising shBle, when using gLuciferase as a reporter gene the band detected corresponded to the expected size for the fusion protein (34.2 kDa) in all the transformants but one (H1.3). There was a slight cross-contamination in CC-124 negative control but we have (previously and later) made immunoblots against this strain and no band or signal had been detected with any of the antibodies used. Nevertheless, despite containing most of the gLuc activity, no bands could be detected by immunoblot in culture supernatant samples (results not shown).

Since #H2.5 is the highest expressing transformant, its intracellular RP expression level was estimated by densitometry based on an immunoblot against HA tag. The expression level is 1.66 µg RP/L culture in cells. Results are shown in figure 26 and table 7.

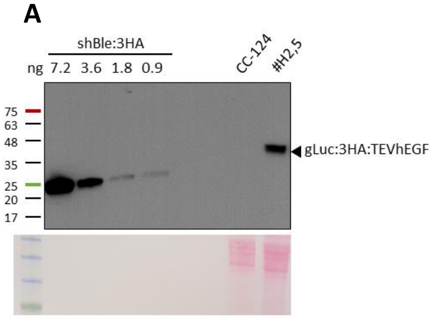

A
shBle:3HA
ng 7.2 3.6 1.8 0.9
CC-124
#H2,5
75
63
48
35
25
20
17
gLuc:3HA:TEVhEGF

**Table 7. pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ H2.5 transformant expression level estimation by densitometry.**

| sample | ng detected | ng corrected | culture loaded (L) | µg RP/L culture |
|---|---|---|---|---|
| **#H2,5 EXT** | 2.96 | 4.14 | 0.0025 | 1.66 |

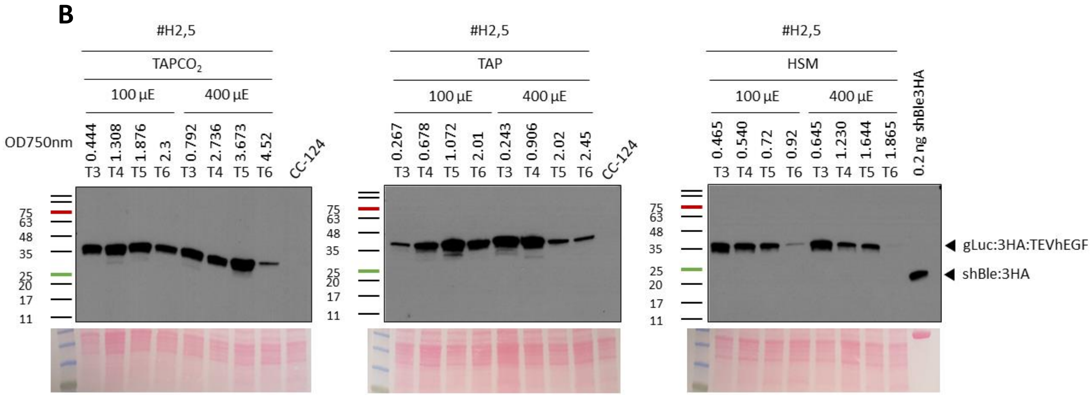

B
#H2,5
TAPCO$_2$
100 µE
400 µE
OD750nm
0.444 T3
1.308 T4
1.876 T5
2.3 T6
0.792 T3
2.736 T4
3.673 T5
4.52 T6
CC-124
#H2,5
TAP
100 µE
400 µE
0.267 T3
0.678 T4
1.072 T5
2.01 T6
0.243 T3
0.906 T4
2.02 T5
2.45 T6
CC-124
#H2,5
HSM
100 µE
400 µE
0.465 T3
0.540 T4
0.72 T5
0.92 T6
0.645 T3
1.230 T4
1.644 T5
1.865 T6
0.2 ng shBle3HA
75
63
48
35
25
20
17
11
gLuc:3HA:TEVhEGF
shBle:3HA

**Figure 26. Immunoblots of intracellular fraction of transformant pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ #H2.5.** Immunoblots against HA tag. **A** Densitometry based quantification of expression from a late-log phase of growth; 5 µl of TSP (500X relative to culture) were loaded. **B** Immunoblot at different growth stages and growth conditions, 25 µg TSP were loaded. Ponceau stained membranes shown as a loading control.

Transformant #H2.5 was grown under different conditions to check if there was an increase in RP yields. As it can be seen in figure 26-B, RP expression was independent of growth medium and its expression decreases as culture reaches stationary phase, being this more accentuated in HSM medium.

Growth of transformant H2.5 was followed by optical density (OD) and expression of RP was evaluated by gLuc activity assay in cells and in culture medium. Results shown in figure 27 revealed that activity found in medium, expectedly corresponding to complete fusion protein, represents only 35% at initial stages of growth and achieves 80-90% at the latter stages of growth. Thereby, it seems that the fusion protein accumulates in medium.

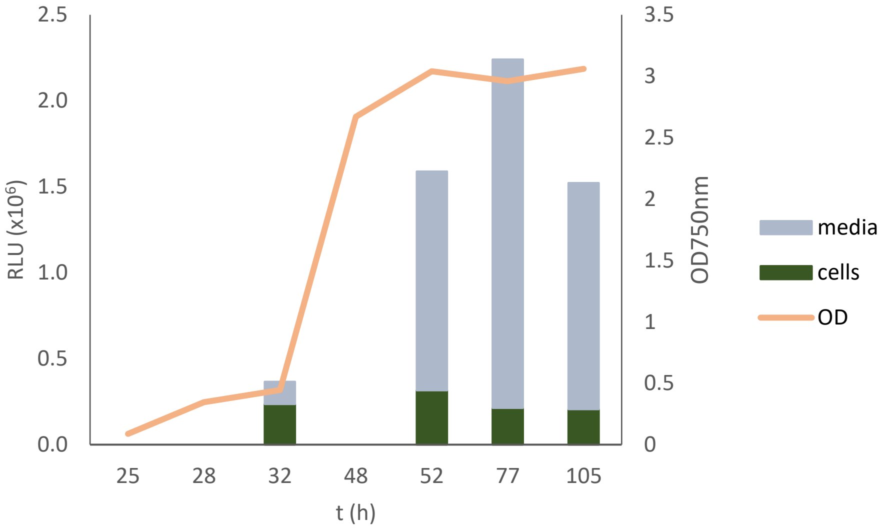


**Figure 27. Expression analysis of pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ #H2.5 selected transformant by gLuc activity test at different growth stages.** Growth followed by optical density at 750 nm is represented as a line. GLuc activity assay results are presented as accumulated bars: intracellular luciferase activity is presented in green and extracellular luciferase activity is presented in blue. gLuc activity assay was performed with supernatant or cell extract corresponding to 50 µl of culture.

The expression strategy allows an *in vivo* high-throughput screening that leads to the selection of the highest expressing strains of the fusion protein.

Although there is a limit on RP expression (1.6 µg RP/L) lower than reported for other proteins (Lauersen et al. 2013a), that may be due to hEGF instability. However, it is important to remark that we quantified the expression of the fusion protein in cells rather than in medium (where gLuc activity is mainly detected). According to luciferase activity the expression levels in medium are much higher than in cells (figure 24) but we could not further analyse it because we were unable to detect the fusion protein in medium by immunoblot against HA tag not even concentrating it a hundred times. At this point, two steps became necessary: i) to detect protein in medium by other means than gLuc assay and ii) to further improve expression levels.

Extracellular and intracellular fusion protein was detected though a primary antibody against gLuc. As it is shown in figure 28 (where intracellular fraction is approximately three times more concentrated than extracellular fraction relative to culture volume), there is a higher yield of fusion protein found in culture medium. Also, expression of a fusion protein comprising hEGF was corroborated by immunoblot against hEGF (figure 28).

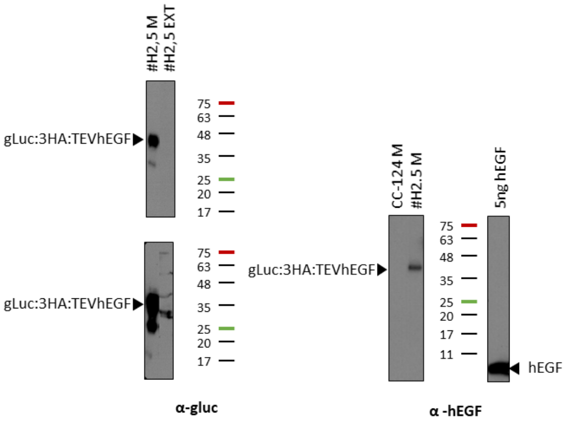


**Figure 28. Immunoblot of pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ #H2.5 transformant.** Left panel corresponds to an immunoblot with a primary antibody against gLuciferase. 8 µl of culture medium (50X relative to culture volume) or total cell extract (166X relative to culture volume) were loaded. Right panel corresponds to an immunoblot with a primary antibody against hEGF. 12.5 µl of culture medium (25X) were loaded. Upper panel corresponds to an exposure of seconds and lower panel corresponds to an exposure of minutes. M: culture medium, EXT: total soluble protein.

Interestingly, despite detecting intracellular and extracellular fusion protein by immunoblot against gLuciferase, only intracellular fusion protein is detected by immunoblot against HA tag. It is likely that a PTM prevents the detection of the 3HA tag from secreted proteins in Chlamydomonas.

### 2.2. Comparison of the efficiency of two different cis regulatory regions to drive transgene expression

Cis regulatory regions (5’ and 3’) of Chlamydomonas Ribosomal protein L23 (*RPL23*, Cre04.g211800) (López-Paz et al. 2017) were tested and compared to AR chimeric promoter and 3’UTR *RBCS2*. To that end the vector pPL-ARSss-glucTEVhEGF-HygB$^R$ (figure 29) was designed. This vector is pAR-ARSss-glucTEVhEGF-HygB$^R$ (figure 22, page 60) with a few changes in the cassette that codes for the fusion protein: AR chimeric promoter and 3’UTR *RBCS2* cis regulatory regions are substituted for *RPL23* cis regulatory regions and a 6xhis tag is cloned N-terminal of gLuc gene instead of C-terminal of hEGF. It is important to highlight that *RPL23* gene has naturally occurring introns in its 5’ and 3’ untranslated regions (López-Paz et al. 2017). The 5’cis regulatory region (1004 bp) contains the predicted promoter and 5’ UTR that is predicted to contain an intron. For the 3’ cis regulatory region the 3’UTR is also predicted to contain an intron and adjacent DNA is used (total of 725 bp).

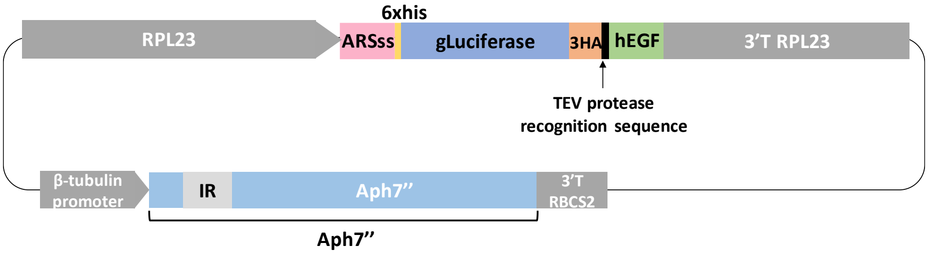


**Figure 29. Schematics of pPL-ARSss-gLucTEVhEGF-HygB$^R$ construct**. RPL23: 5’ *RPL23* cis regulatory elements including 5’UTR and promoter, 3’T RPL23: 3’UTR and terminator of *RPL23*, ARSss: Chlamydomonas arylsulfatase 1 secretion signal, IR: intron 1 *RBCS2*, 3’T RBCS2: 3’ region of *RBCS2*.

Cell walled strain CC-124 was transformed with vectors pPL-ARSss-gLucTEVhEGF-HygB$^R$ (figure 29) and pAR-ARSss-gLuc-3HA-TEV-hEGF-6his-HygB$^R$ (figure 22) and 192 initial transformants (resistant to 60 µg/ml of Hygromycin B) of each vector were selected for further screening by gLuc activity assay. gLuc activity was assayed directly with culture (including both cells and medium) (figure 30). Two metrics were used to compare performances between populations (transformants of each vector). The first metric was the proportion of positive transformants (luciferase activity 10-fold above background: >1 x $10^3$ RLUs) that was statistically analysed by Fisher's exact test. The second metric was the distribution of positive RLU values binned by expression level and was statistically analysed by Wilcoxon-Mann-Whitney rank sum test. Differences were considered significant when $p < 0.05$. We could not find a significantly higher proportion of positive transformants with vector pPL-ARSss-glucTEVhEGF-HygB$^R$ than with vector pAR-ARSss-glucTEVhEGF-HygB$^R$, determined by Fisher's exact test ($p < 0.05$). However, we found a significant increase in expression levels for positive transformants of pPL-ARSss-glucTEVhEGF-HygB$^R$ compared with the control construct determined by a two-tailed Mann-Whitney U test ($p < 0.05$).

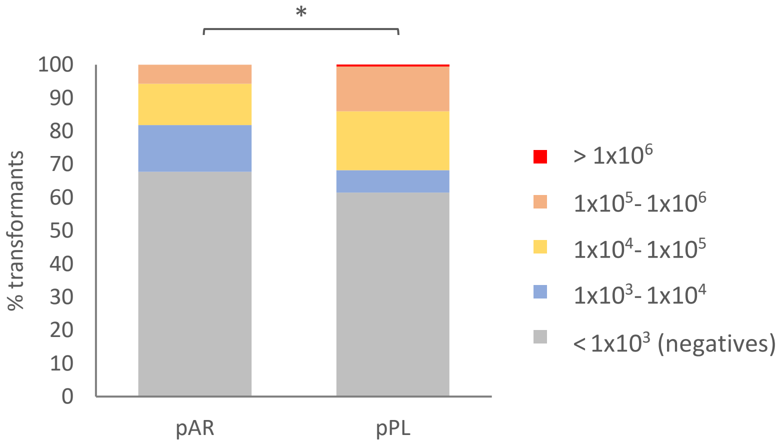


**Figure 30. gLuc activity results for pAR-ARSss-gLucTEVhEGF (pAR) and pPL-ARss-gLucTEVhEGF (pPL) transformants**. Assay was performed with 100 µl of culture supernatant. * Indicates a significant increase in expression levels for positive transformants compared with the control construct determined by a two-tailed Mann-Whitney U test ($p<0.05$).

The highest expressing transformants were selected and analysed by immunoblot (figure 31). In all cases, the band detected corresponds to the protein with expected size (34.2 kDa) in all the transformants. No band or signal was detected in CC-124 negative control with the antibody used.

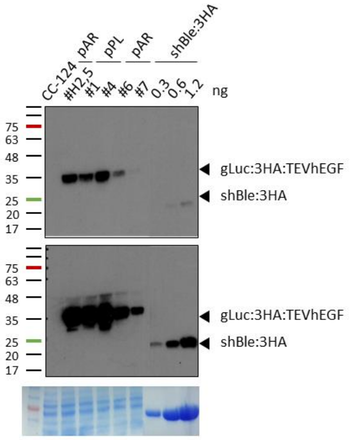


**Figure 31. Immunoblot of pAR-ARSss-gLucTEVhEGF-hygB$^R$ (pAR) and pPL-ARSss-gLucTEVhEGF-hygB$^R$ (pPL) highest expressing transformants of CC-124 strain.** Primary antibody anti HA tag. 50 µg of TSP were loaded. Two exposures are shown: the one showed above corresponds to seconds of exposure and the one showed below corresponds to hours of exposition. Coomassie brilliant blue stained membrane is shown as a loading control. ShBle:3HA is used as a positive control.

Since a significant increase in gLuciferase activity was found for positive transformants of pPL-ARSss-glucTEVhEGF-HygB$^R$ when compared to pAR-ARSss-glucTEVhEGF-HygB$^R$ transformants, further work was performed with *RPL23* cis-regulatory regions.

### 2.3. Comparison of the effect of several secretion peptides on expression yields

It has been shown that the use of alternative signal peptides affects the efficiency of protein secretion (Knappskog et al. 2007; Kober et al. 2013) and thereby signal peptide selection affects the overall yield of secreted proteins in an expression system. Regarding Chlamydomonas, a recent study tested the effect of 10 signal peptides on secretion of mCherry fluorescent protein showing that the efficiency of RP secretion and the RP yield are deeply influenced by the signal peptide used (Molino et al. 2018): a variation of 8-fold in the fluorescence detected among secreting transformants was detected.

With the aim to improve expression levels, two secretion sequences were tested and compared: arylsulfatase 1 secretion sequence (ARSss) and metalloprotease gametolysin secretion sequence (GAMss). Both secretion sequences had been previously used to secrete nuclear RP from Chlamydomonas nucleus (Rasala et al. 2012; Ramos-Martinez et al. 2017). To that end we designed the vectors pPL-ARSss-glucTEVhEGF-paro$^R$ and pPL-GAMss-glucTEVhEGF-paro$^R$ (figure 32). Both carry two cassettes: one that carries the resistance to paromomycin and one that encodes the fusion protein 6xhis:gLuc:3HA:TEV:hEGF. The two vectors differ only in the secretion sequence.

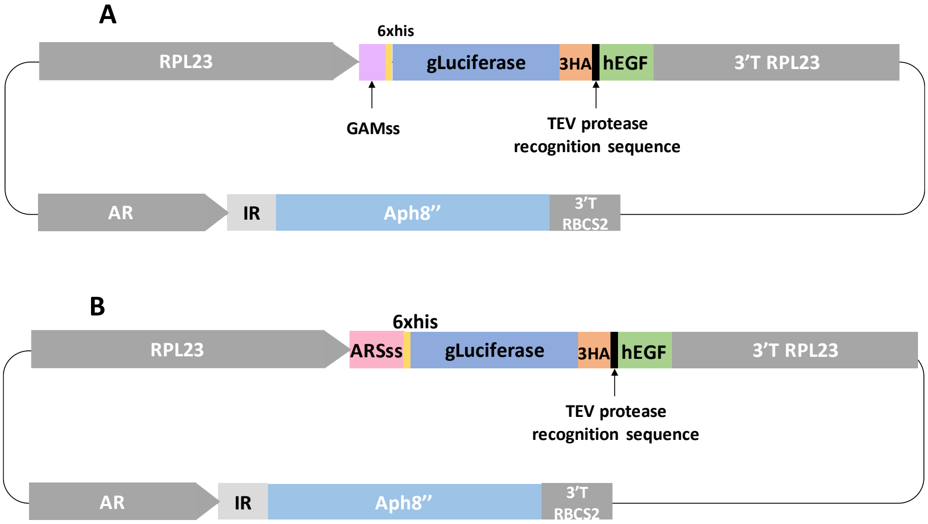

A
RPL23
6xhis
gLuciferase
3HA
hEGF
3'T RPL23
GAMss
TEV protease recognition sequence
AR
IR
Aph8''
3'T RBCS2
B
RPL23
6xhis
ARSss
gLuciferase
3HA
hEGF
3'T RPL23
TEV protease recognition sequence
AR
IR
Aph8''
3'T RBCS2

**Figure 32. Schematics of pPL-GAMss-gLucTEVhEGF-paro$^R$ (A) and pPL-ARSss-gLucTEVhEGF-paro$^R$ (B) vectors.** RPL23: 5' *RPL23* cis regulatory elements including 5'UTR and promoter, 3'T RPL23: 3'UTR and terminator of *RPL23*, ARSss: arylsulfatase 1 secretion signal, GAMss: metalloprotease gametolysin secretion signal, AR: HSP70-RBCS2 chimeric promoter, IR: intron 1 *RBCS2*, 3'T RBCS2: 3' region of *RBCS2*.

The cell walled strain CC-124 and the cell wall-less mutant strain UVM4 (a mutant strain that was isolated because it has an elevated expression of transgenes (Neupert et al. 2009)) were transformed with both plasmids. The expression level of 48 initial transformants (resistant to 25 µg paromomycin) of each vector and each strain was tested by luciferase activity assay.

Luminescence values were normalized by culture density (OD 750 nm) and expressed as relative luciferase units (RLUs). Transformants that showed luciferase activity 100-fold above background ($>1 \times 10^4$ RLUs) were considered positives and were binned by expression level in groups: $1 \times 10^4$ - $1 \times 10^5$ RLUs, $1 \times 10^5$ - $5 \times 10^5$ RLUs, $5 \times 10^5$ - $1 \times 10^6$ RLUs and higher than $1 \times 10^6$ RLU. We used two metrics to assess performance in each population (transformants of each vector) compared among the same strain. The first metric was the proportion of positive transformants (luciferase activity above background: $>1 \times 10^4$ RLUs) that was statistically analysed by Fisher's exact test. Considering significant differences when $p < 0.05$, proportion of positives was not significantly different. The second metric was the distribution of positive RLU values binned by expression level and was statistically analysed by Wilcoxon-Mann-Whitney rank sum test. Differences were considered significant when $p < 0.05$. Both strains gave statistically significant differences among the two populations (transformants of each vector) resulting GAMss in a higher yield of RP in medium (figure 33).

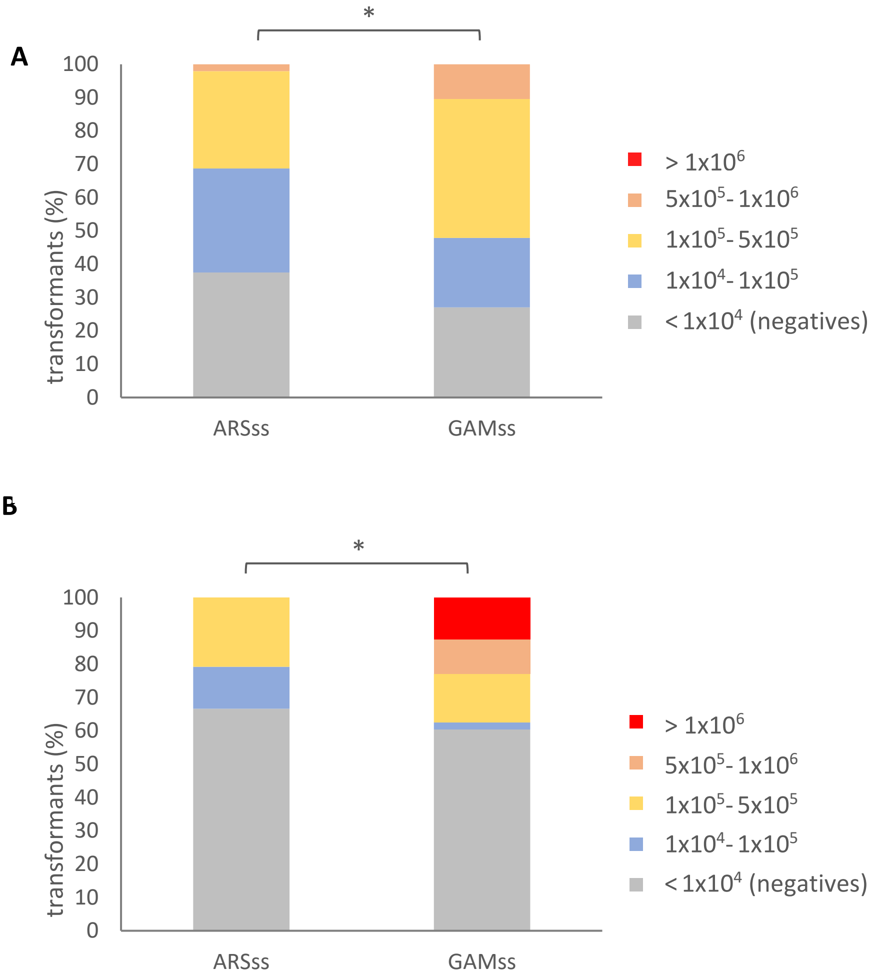


**Figure 33. Distribution of normalized luciferase expression (RLU) values from Chlamydomonas strains UVM4 (A) and CC-124 (B) transformed with vectors pPL-GAMss-gLucTEVhEGF (GAMss) and pPL-ARSss-gLucTEVhEGF (ARSss).** Assay was performed with 25 µl of culture supernatant. * Indicates a significant increase in expression levels for positive transformants compared with the control construct determined by a two-tailed Mann-Whitney U test ($p<0.05$).

Chlamydomonas native extracellular Carbonic Anhydrase 1 secretion sequence (CAHss) has been also successfully used to secrete RP from Chlamydomonas nucleus achieving an expression level of 1.5 mg gLuc/L in wild type cells (Lauersen et al. 2013a) and 12 mg gLuc/L in UVM4 strain (Lauersen et al. 2015a). We cloned the vector pPL-CAHss-glucTEVhEGF-paro$^R$ (figure 34) in order to test CAHss in comparison to GAMss and ARSss (figure 32).

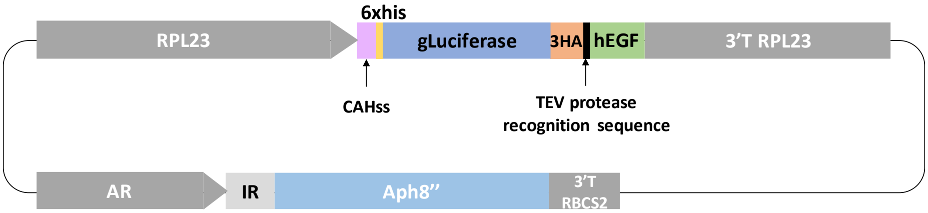


**Figure 34. Schematics of pPL-CAHss-gLucTEVhEGF-paro$^R$ vector.** RPL23: 5' *RPL23* cis regulatory elements including 5'UTR and promoter, 3'T RPL23: 3'UTR and terminator of *RPL23*, CAHss: Chlamydomonas carbonic anhydrase 1 secretion signal, AR: *HSP70-RBCS2* chimeric promoter, IR: intron 1 *RBCS2*, 3'T RBCS2: 3' region of *RBCS2*.

The three vectors (that differ only in the secretion sequence) were transformed into the strain CC-124 and the expression level of 48 initial transformants (resistant to 25 µg paromomycin) of each vector were tested by luciferase activity assay. Luminescence values were normalized by culture density (OD 750 nm) and expressed as relative luciferase units (RLUs). Transformants were binned by expression level in groups as explained previously. The two metrics used to assess performance in each population (transformants of each vector) were the same explained previously and differences were considered significant when $p < 0.05$. Proportion of positives analysed by Fisher's exact test was not significantly increased. Analysed by Wilcoxon-Mann-Whitney rank sum test, expression level resulted statistically different between CAHss and the other two populations (GAMss and ARSss transformants).

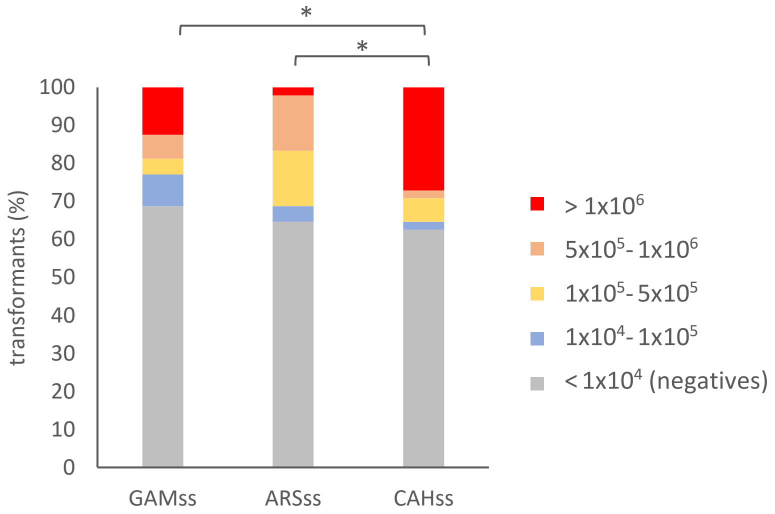
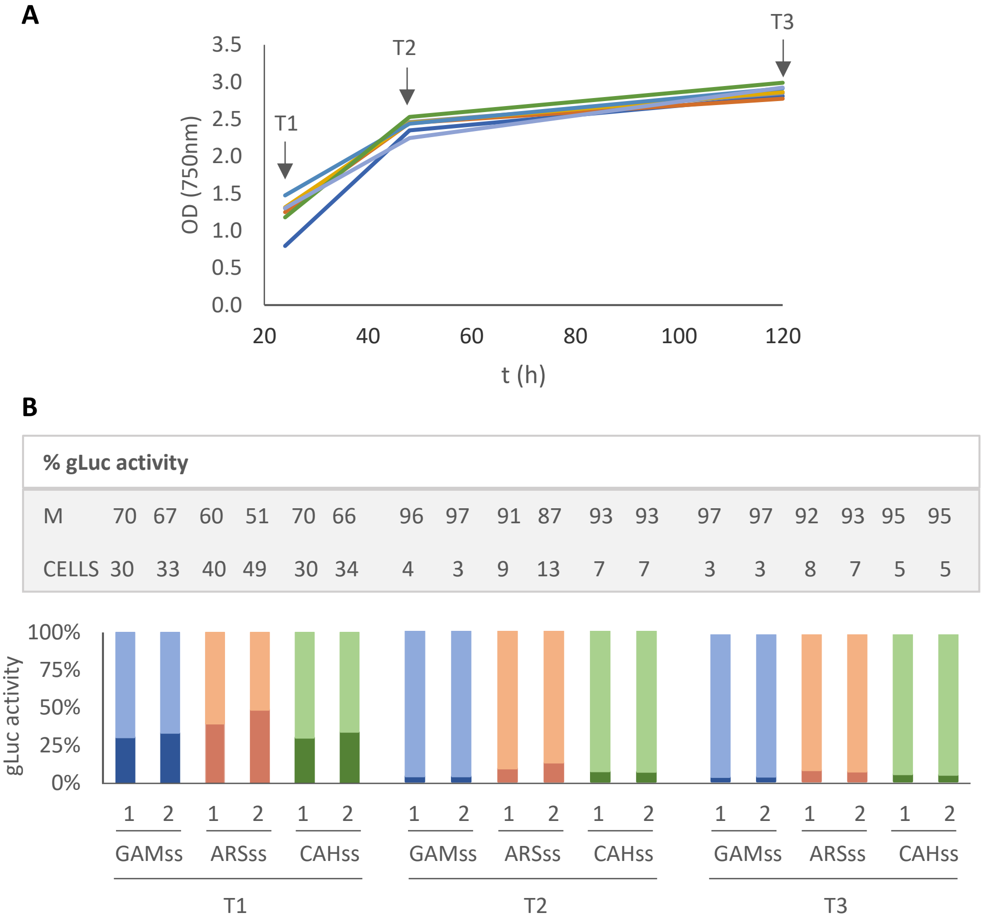


**Figure 35. Distribution of normalized luciferase expression (RLU) values from CC-124 Chlamydomonas strain transformed with vectors pPL-GAMss-gLucTEVhEGF (GAMss), pPL-ARSss-gLucTEVhEGF (ARSss) and pPL-CAHss-gLucTEVhEGF (CAHss).** Assay was performed with 5ul of culture supernatant. * Indicates a significant increase in expression levels for positive transformants compared with the control construct determined by a two-tailed Mann-Whitney U test ($p<0.05$).

Use of CAHss significantly increased expression levels when compared with ARSss and GAMss. To gain further insight onto secretion efficiency of each secretion signal growth of two independent transformants of each vector was followed by optical density (OD) and expression of RP was evaluated by gLuc activity assay in cells and in culture supernatant. As it can be seen in figure 36, it seems that ARSss has a lower secretion efficiency compared to CAHss and GAMss. Results shown in figure 36 confirmed previously obtained results (*2.1. Fusion of 3XHA tagged Gaussia Luciferase reporter to the protein of interest to enable screening of transformants - figure 27*): activity found in medium represents only 30-50 % at initial stages of growth and achieves more than 90% at the latter stages of growth. Thereby, it seems that the fusion protein accumulates in medium.



**% gLuc activity**

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M | 70 | 67 | 60 | 51 | 70 | 66 | 96 | 97 | 91 | 87 | 93 | 93 | 97 | 97 | 92 | 93 | 95 | 95 |
| CELLS | 30 | 33 | 40 | 49 | 30 | 34 | 4 | 3 | 9 | 13 | 7 | 7 | 3 | 3 | 8 | 7 | 5 | 5 |

**Figure 36. Secretion efficiency analysis of GAMss, ARSss and CAHss secretion signals by gluciferase activity. A)** Growth followed by optical density at 750 nm. **B)** Secretion efficiency analysed by gLuc activity. Two transformants of each vector were analysed. Intracellular gLuc activity (CELLS, represented in dark colour) and extracellular gLuc activity (M, represented in light colour) were assayed with two technical replicates. GAMss: pPL-GAMss-gLucTEVhEGF, ARSss: pPL-ARSss-gLucTEVhEGF, CAHss: pPL-CAHss-gLucTEVhEGF.

### 2.4. Use of glycomodules to enhance expression and secretion of nuclear transgenes

As introduced previously, a successful strategy to enhance yields and stability of secreted RP in plant cell cultures has been demonstrated based on (Ser-Pro)n glycomodules (Shpak et al. 1999; Xu et al. 2007). Concerning Chlamydomonas, this glycomodules conferred an enhanced protein stability and allowed accumulation of RP in culture medium. Due to that, yields of a (Ser-Pro)n-fused fluorescent protein increased up to 12 fold compared to fluorescent protein without glycomodules reaching a maximum yield of 15 mg/L (Ramos-Martinez et al. 2017). To our knowledge, this represents the highest yield of RP expression in Chlamydomonas nucleus reported until today. Those glycomodules consisted in tandem serine (Ser) and proline (Pro) repeats of 10 and 20 units [hereafter $(SP)_n$, wherein n=10 or 20] and are the same tested and patented sequences in RP expression from plants.

In order to determine if expression of our reporter-fusion proteins could be improved by addition of SP glycomodules as has been previously reported with other proteins, and to further investigate the high-expressing strains UVM4 and UVM11 (Neupert et al. 2009) as hosts for RP expression, 3 plasmids were transformed into UVM4 and UVM11 mutant strains and into the cell wall-less strain CC-1883: pPL-ARSss-glucTEVhEGF-hygB$^R$, pPL-ARSss-glucTEVhEGF-SP$_{10}$-hygB$^R$ and pPL-ARSss-glucTEVhEGF-SP$_{20}$-hygB$^R$ (figure 38). These vectors are the same tested previously (pPL-ARSss-glucTEVhEGF-hygB$^R$) with the addition of SP$_{10}$ or SP$_{20}$ glycomodules cloned C-terminal of the fusion protein. These vectors carry the expression of two different cassetes: i) resistance to hygromycin B and ii) the cassette coding for the fusion protein ARSss–6xhis tag–gLuc-3HA tag-TEV–hEGF-glycomodule. The expression level of 48 initial transformants (resistant to 60 µg hygromycin B) of each vector and each strain were tested by luciferase activity assay. Luminescence values were normalized by culture density (OD 750 nm) and expressed as relative luciferase units (RLUs). Transformants that showed luciferase activity 100-fold above background (>1 x $10^4$ RLUs) were considered positives and were binned by expression level in groups as previously explained (*2.3 Comparison of the effect of several secretion peptides on expression yields*). We used two metrics to assess performance in each population (transformants of each vector) compared among the same strain. The first metric was the proportion of positive transformants (luciferase activity above background: >1 x $10^4$ RLUs) that was statistically analysed by Fisher's exact test. Only strain UVM11 transformed with pPL-ARSss-glucTEVhEGF-$SP_{10}$-$hygB^R$ showed a significantly higher proportion of positive transformants than the control construct, pPL-ARSss-glucTEVhEGF-$hygB^R$, determined by Fisher's exact test ($p < 0.05$). The second metric was the distribution of positive RLU values binned by expression level and was statistically analysed by Wilcoxon-Mann-Whitney rank sum test. Differences were considered significant when $p < 0.05$. Strain CC-1883 performed clearly worse than UVM4 and UVM11 and it did not give positive colonies of pPL-ARSss-glucTEVhEGF-$hygB^R$ vector, reason why it was excluded from the statistical analysis. However, since positive transformants of strain CC-1883 were obtained for both vectors carrying glycomodule sequences, the positive effect on RP yields exerted by glycomodule sequences is destacable in this case. For strains UVM4 and UVM11 there was a significant increase in expression levels for positive transformants of the two vectors tested (pPL-ARSss-glucTEVhEGF-$SP_{10}$-$hygB^R$ and pPL-ARSss-glucTEVhEGF-$SP_{20}$-$hygB^R$) compared with the control construct (pPL-ARSss-glucTEVhEGF-$hygB^R$) determined by a two-tailed Mann-Whitney U test ($p < 0.05$). Thereby, the increase of RP expression from Chlamydomonas nucleus with $SP_{10}$ and $SP_{20}$ glycomodules was corroborated with our reporter-fusion proteins. Furthermore, we found a significant increase in expression levels for positive transformants of the three vectors tested when expressed in strain UVM4 compared with the same construct expressed in strain UVM11 determined by a two-tailed Mann-Whitney U test ($p < 0.05$).

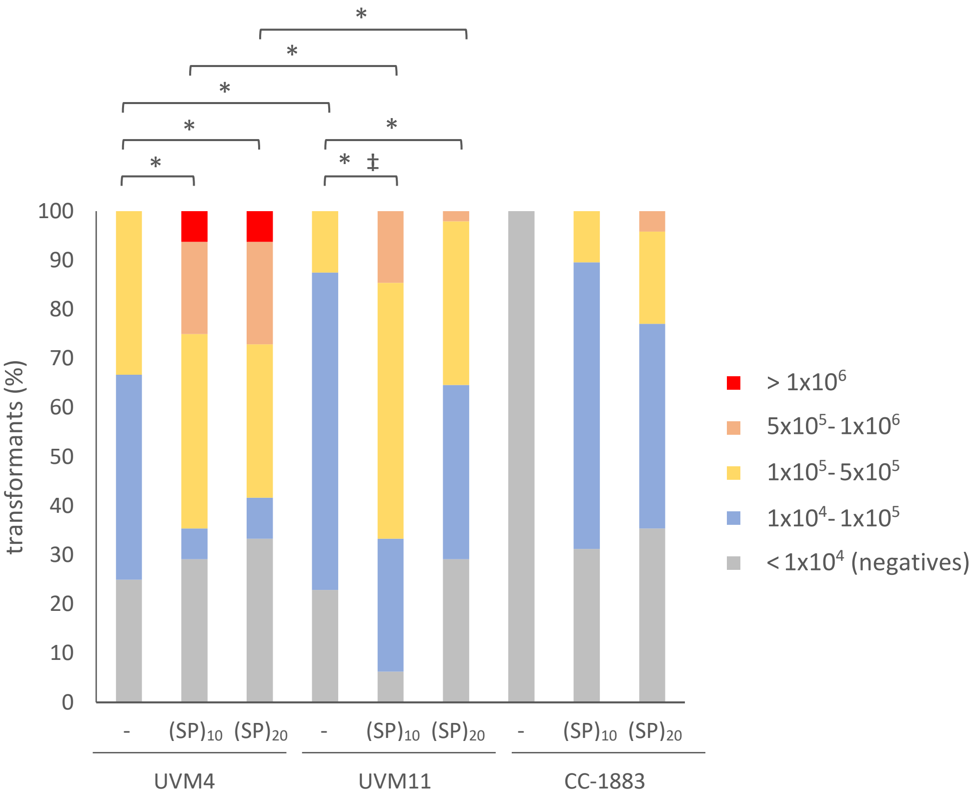


**Figure 37. Distribution of normalized luciferase expression (RLU) values from Chlamydomonas strains UVM4, UVM11 and CC-1883 transformed with constructs pPL-ARSss-glucTEVhEGF-hygB$^R$, pPL-ARSss-glucTEVhEGF-(SP)$_{10}$-hygB$^R$ and pPL-ARSss-glucTEVhEGF-(SP)$_{20}$-hygB$^R$.** 48 transformants were analysed except for CC-1883 strain transformed with pPL-ARSss-glucTEVhEGF-hygB$^R$ (24 transformants analysed). Assay was performed with 2 µl of culture supernatant. ‡ indicates a significantly higher proportion of positive transformants than the control construct, pPL-ARSss-glucTEVhEGF-hygB$^R$, determined by Fisher's exact test ($p < 0.05$).* indicates a significant increase in expression levels for positive transformants compared with the control construct determined by a two-tailed Mann-Whitney U test ($p < 0.05$). -: pPL-ARSss-glucTEVhEGF-hygB$^R$, (SP)$_{10}$: pPL-ARSss-glucTEVhEGF-(SP)$_{10}$-hygB$^R$, (SP)$_{20}$: pPL-ARSss-glucTEVhEGF-(SP)$_{20}$-hygB$^R$.

Given that fusion proteins comprising SP$_{10}$ and SP$_{20}$ glycomodules showed an enhanced stability allowing accumulation in medium of RPs, we investigated the existence of native glycomodules from Chlamydomonas in its more abundant secreted proteins that would confer higher stability to such proteins and thus may be used to enhance RP yields. For that purpose we took advantage of a Chlamydomonas proteome study of the most abundant Chlamydomonas proteins present in medium under different conditions (Baba et al. 2011) and we studied the presence of putative glycomodules in these proteins. We identified possible glycomodules as repetitive sequences, not necessarily with $(SP)_n$ motif, presenting several combinations and lengths of repetitive sequences. We tested five sequences that represent different glycomodules (table 8).

**Table 8. Chlamydomonas native glycomodules selected to test its effect on RP expression.**

| Protein of origin | JGI protein ID | Glycomodule name | Sequence |
|---|---|---|---|
| LCL5 | 196466 | **LCL** | SSSPAPASSAPAPARSSSASPPPPPRRPPPRP |
| GP1 | 34358 | **GP1** | PPSPAPPSPAPPSPAPPSPAPPSPAPPSPGPPS |
| GP2 | 195768 | **GP2** | PPSPPPPTPPSPPPPSPPPPVLSPPPSPPPPSPPP<br>PAPPPPSPPPPVPPPPSPPPPSPPPPSPPPAAAS<br>PPPSPPPPPPPSPPPPVARLP |
| PHC21 | 93464 | **PHC21A** | MPPPSPPPPPPPSPPPPSPPPPSPKPPSPPPPSP<br>PPPSPPPPSPKPPSPKPPSPPPPSPPPP |
|  |  | **PHC21B** | PPSPPPSPPPPPPPPPPPPPPPAPPPPPP<br>PPPPPPPPPSPPPPPPPPPPPPPPPPPFP<br>PPPPPPPPPPPSPPPP |

$(SP)_{10}$ sequence was substituted in pPL-ARSss-glucTEVhEGF-$(SP)_{10}$-hyg$B^R$ for LCL, GP1, GP2, PHC21A and PHC21B (table 8) generating a battery of vectors to test the selected sequences. Schematics of DNA cassettes are shown in figure 38.

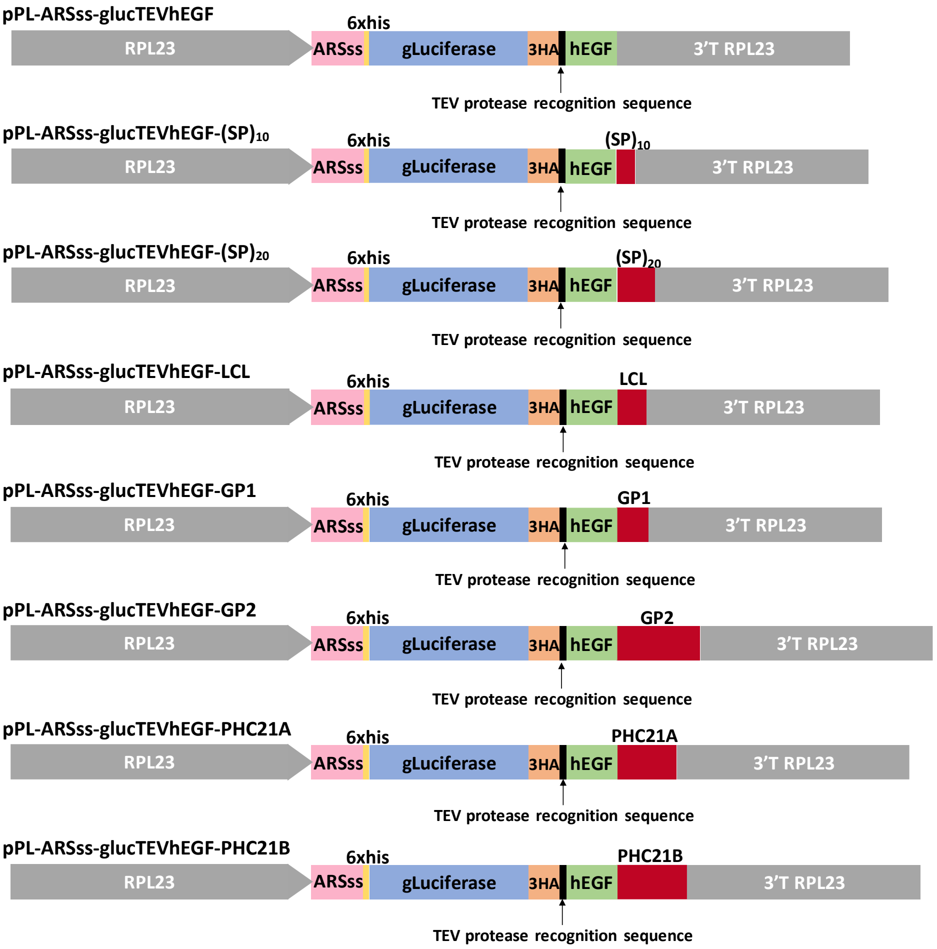

pPL-ARSss-glucTEVhEGF
pPL-ARSss-glucTEVhEGF-(SP)10
pPL-ARSss-glucTEVhEGF-(SP)20
pPL-ARSss-glucTEVhEGF-LCL
pPL-ARSss-glucTEVhEGF-GP1
pPL-ARSss-glucTEVhEGF-GP2
pPL-ARSss-glucTEVhEGF-PHC21A
pPL-ARSss-glucTEVhEGF-PHC21B
RPL23
6xhis
ARSss
gLuciferase
3HA
hEGF
(SP)10
(SP)20
LCL
GP1
GP2
PHC21A
PHC21B
3'T RPL23
TEV protease recognition sequence

**Figure 38. Schematics of cassettes used to test the selected glycomodules.** RPL23: 5' *RPL23* cis regulatory elements including 5'UTR and promoter, 3'T RPL23: 3'UTR and terminator of *RPL23*, ARSss: arylsulfatase 1 secretion signal, IR: intron 1 *RBCS2*. Refer to table 8 to see the detail of the sequences identified as LCL, GP1, GP2, PHC21A, PHC21B. In all cases the complete vectors include a cassette carrying the gene *AphVII'* conferring resistance to hygromycin B driven by β-tubulin 5' cis regulatory elements and 3' region of *RBCS2* that is not included in this figure.

Since UVM4 was the best performing cell wall-less strain tested, this strain and the cell walled strain CC-124 were transformed with the 8 designed plasmids (figure 38). The expression level of 48 transformants (resistant to 60 µg hygromycin B) of each vector and each strain were tested by luciferase activity assay. For the strain UVM4 only 21 transformants of pPL-ARSss-glucTEVhEGF-

$(SP)_{20}$-hygB$^R$ and 12 transformants of pPL-ARSss-glucTEVhEGF-PHC21B-hygB$^R$ were obtained and analysed. Luminescence values were normalized by culture density (OD 750 nm) and expressed as relative luciferase units (RLUs) (figure 39). The assay was performed twice: first with cultures at late log phase of grow and then results were corroborated with cultures at stationary phase of growth. Transformants that showed luciferase activity 100-fold above background (>1 x $10^4$ RLUs) were considered positives and were binned by expression level in groups as explained. The metrics were the same than previously described and differences were considered significant when $p < 0.05$. Proportion of positives (analysed by Fisher's exact test) was not increased with any of the vectors tested. Determined by a two-tailed Mann-Whitney U test ($p < 0.05$) there was a significant increase in expression levels with the control construct for three of the vectors tested in both UVM4 and CC-124 strains: pPL-ARSss-glucTEVhEGF-$(SP)_{10}$-hygB$^R$, pPL-ARSss-glucTEVhEGF-$(SP)_{20}$-hygB$^R$ and pPL-ARSss-glucTEVhEGF-PHC21A-hygB$^R$. Thus, fusion proteins carrying $(SP)_{10}$ and $(SP)_{20}$ glycomodules or the native PHC21A glycomodule resulted in significantly higher expression levels than the fusion protein without glycomodules in both strains tested. However, fusion proteins carrying $(SP)_{10}$ and $(SP)_{20}$ glycomodules resulted in significantly higher expression levels than fusion proteins carrying PHC21A glycomocule in strain UVM4. Also, in CC-124 strain fusion proteins carrying $(SP)_{10}$ glycomodule resulted in significantly higher expression levels than fusion proteins carrying PHC21A. Summarizing, the proposed native glycomodule PHC21A significantly enhanced RP expression from Chlamydomonas nucleus but it did not surpass the performance of previously described $(SP)_{10}$ and $(SP)_{20}$ sequences.

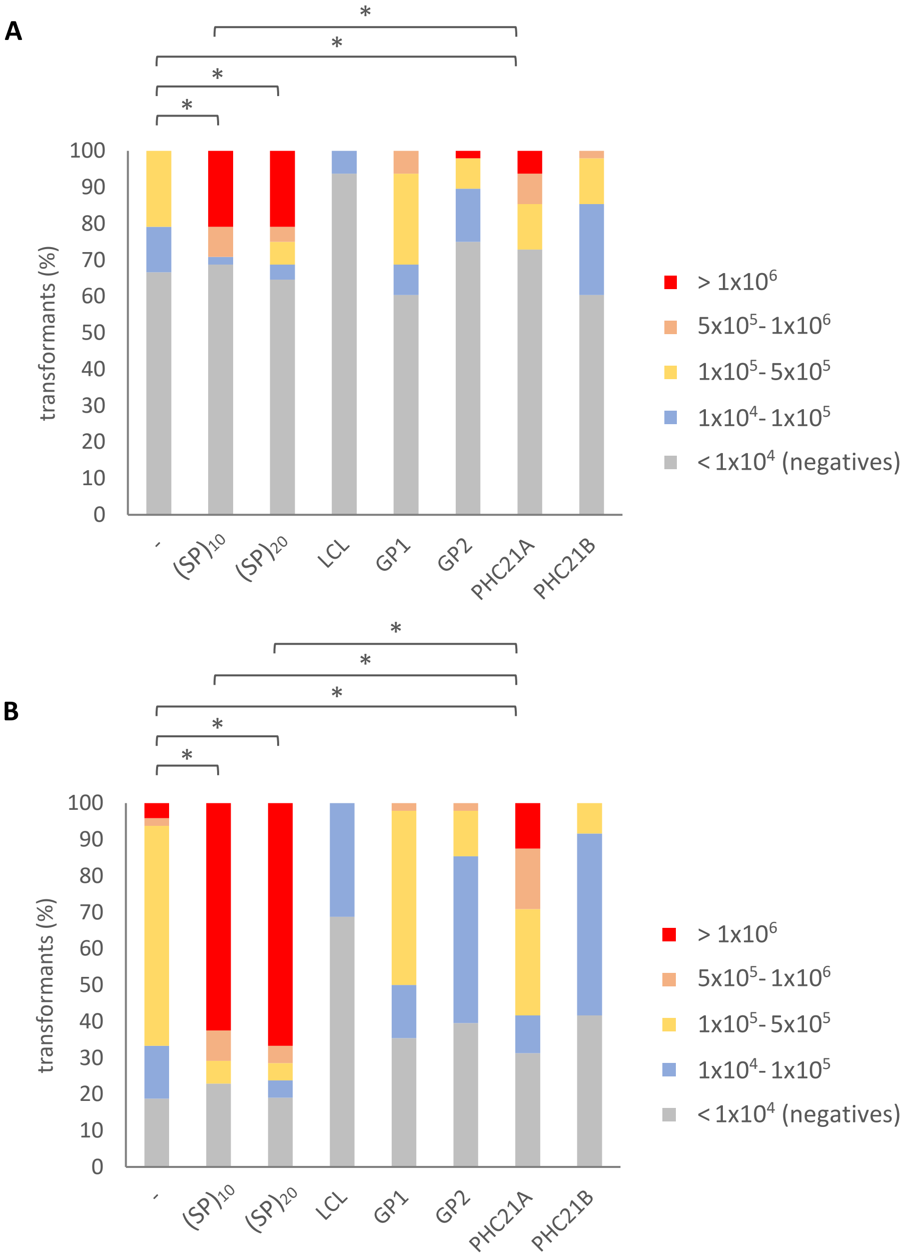


**Figure 39. Distribution of normalized luciferase expression (RLU) values from Chlamydomonas strains CC-124 (A) and UVM4 (B) transformed with constructs illustrated in figure 38.** 48 transformants were analysed except for UVM4 strain transformed with pPL-ARSss-glucTEVhEGF-SP20-hygB$^R$ (21 transformants analysed) and pPL-ARSss-glucTEVhEGF-PHC21B-hygB$^R$ (12 transformants analysed). Assay was performed with 25 µl of supernatant from cultures at late log phase of growth. * indicates significant increase in expression levels for positive transformants compared with the control construct determined by a two-tailed Mann-Whitney U test ($p < 0.05$). -: pPL-ARSss-glucTEVhEGF -hygB$^R$, SP10: pPL-ARSss-glucTEVhEGF-SP10-hygB$^R$, SP20: pPL-ARSss-glucTEVhEGF-SP20-hygB$^R$, LCL: pPL-ARSss-glucTEVhEGF-LCL-hygB$^R$, GP1: pPL-ARSss-glucTEVhEGF-GP1-hygB$^R$, GP2: pPL-ARSss-glucTEVhEGF-GP2-hygB$^R$, PHC21A: pPL-ARSss-glucTEVhEGF-PHC21A-hygB$^R$, PHC21B: pPL-ARSss-glucTEVhEGF-PHC21B-hygB$^R$.

The three highest expressing transformants of each strain and vector were selected and an immunoblot analysis of concentrated culture supernatant was performed to determine whether the fusion proteins were correctly expressed and if expression levels correlated with luciferase activity (figure 40). The immunoblot analysis was performed with cultures at stationary phase of growth as it has been it has been observed that this is the stage in which higher concentration of RP is present in the medium (*2.1. Fusion of 3XHA tagged Gaussia Luciferase reporter to the protein of interest to enable screening of transformants).*

Fusion proteins carrying $(SP)_{10}$ and $(SP)_{20}$ glycomodules are correctly expressed in the 12 transformants analysed by immunoblot: detected bands had higher molecular weight than predicted by sequence (36.1 kDa for gLuc:3HA:TEV:hEGF:$(SP)_{10}$ and 37.9 kDa for gLuc:3HA:hEGF:$(SP)_{20}$) but this is suggested to be O-glycosylation as previously reported (Shpak et al. 1999).
Of the fusion proteins carrying native glycomodules expressed in CC-124 transformants, only fusion proteins carrying GP1 or PHC21A sequences were detected, concurring to what has been observed with luciferase activity. As it is the case of proteins carrying $(SP)_{10}$ and $(SP)_{20}$ glycomodules, fusion proteins are detected at higher molecular weights than predicted by sequence (36.9 kDa for gLuc:3HA:TEV:hEGF:GP1 and 40.1 kDa for gLuc:3HA:hEGF:PHC21A) presumably due to glycosylation. The molecular weight at which are detected is also consistent between independent transformants of the same construct. Expression level was lower than the observed for proteins carrying $(SP)_{10}$ and $(SP)_{20}$ glycomodules. In the transformant GP2#20 there is detection of recombinant protein at a molecular weight too low to correspond to a fusion protein carrying a glycomodule, thereby it may correspond to a partial insertion of the cassette.

Regarding UVM4 transformants, use of native glycomodules resulted in detectable expression of RP only in one transformant expressing a fusion protein carrying the PHC21A sequence. As in fusion proteins carrying $(SP)_{10}$ and $(SP)_{20}$ glycomodules and in the same manner as it happens in CC-124 transformants fusion protein is detected at a higher molecular weight than predicted by sequence (40.1 kDa) suggesting the presence of O-glycosylation as mentioned.

Strikingly, more degradation was observed in UVM4 transformants. Although there is degradation of RP in the CC-124 transformants, the proportion of intact fusion protein is higher than the proportion of degraded protein. On the contrary, in UVM4 transformants there is a higher proportion of degraded protein than intact fusion protein in all the transformants analysed (those which present detectable RP). It is important to highlight that transformant UVM4#9, that expresses a fusion protein carrying a $(SP)_{10}$ glycomodule, presented less degradation than the other UVM4 transformants expressing the same RP (#7, #8) and is the one less grown of the three (it had not reached stationary phase when harvested).

Immunoblot analysis results support the statistical analysis based on luciferase assay results: increase of secreted RP yields from Chlamydomonas nucleus with fusion of $(SP)_{10}$ and $(SP)_{20}$ glycomodules was corroborated and none of the native proposed sequences surpassed the performance of $(SP)_{10}$ and $(SP)_{20}$ previously described sequences.

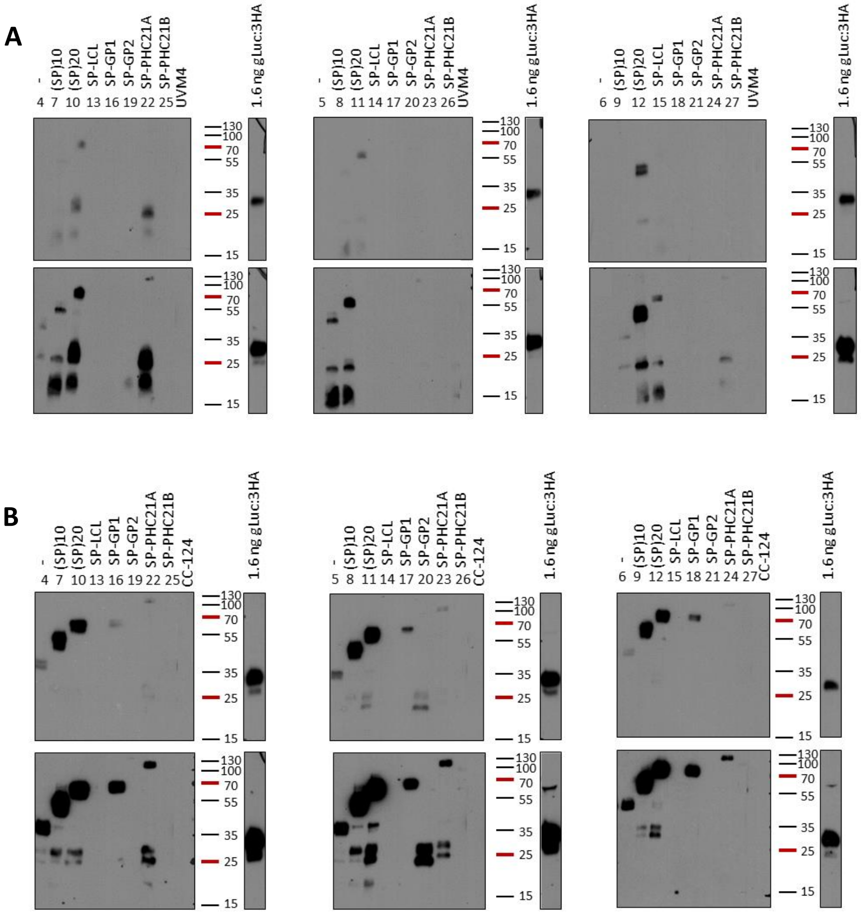

A
B
-
(SP)10
(SP)20
SP-LCL
SP-GP1
SP-GP2
SP-PHC21A
SP-PHC21B
UVM4
CC-124
4 7 10 13 16 19 22 25
5 8 11 14 17 20 23 26
6 9 12 15 18 21 24 27
1.6 ng gLuc:3HA
130
100
70
55
35
25
15

**Figure 40. Immunoblot of selected high expressing transformants of fusion proteins comprising a glycomodule sequence.** 8 µl of culture medium concentrated 24X were loaded in each lane. Primary antibody against gLuciferase. Two exposures are shown for each immunoblot: the one showed above corresponds to seconds of exposure and the one showed below corresponds to minutes of exposition. **A UVM4 transformants immunoblot. B CC-124 transformants immunoblot.** -: pPL-ARSss-glucTEVhEGF -hygB$^R$, (SP)10: pPL-ARSss-glucTEVhEGF-(SP)$_{10}$-hygB$^R$, (SP)20: pPL-ARSss-glucTEVhEGF-(SP)$_{20}$-hygB$^R$, LCL: pPL-ARSss-glucTEVhEGF-LCL-hygB$^R$, GP1: pPL-ARSss-glucTEVhEGF-GP1-hygB$^R$, GP2: pPL-ARSss-glucTEVhEGF-GP2-hygB$^R$, PHC21A: pPL-ARSss-glucTEVhEGF-PHC21A-hygB$^R$, PHC21B: pPL-ARSss-glucTEVhEGF-PHC21B-hygB$^R$.

## 2.5. Introduction of an intronic sequence within a reporter gene to enhance nuclear expression

Inclusion of intronic sequences within genes has been successfully used to improve expression levels of RP from the nucleus of Chlamydomonas (Lumbreras et al. 1998). Specifically, *RBCS2* intron 1 (Sizova et al. 2001) has been successfully used to enhance expression of genes and the vector pPL-ARSss-gLucINT-$(SP)_{20}$-hEGF was ideated to test the effect of this intronic region on the expression of a fusion protein by inserting it within the reporter gene gLuciferase (figure 41). Besides, we ideated this vector with $(SP)_{20}$ glycomodule sequence cloned after the gLuc gene and just before the TEV protease recognition sequence. This was performed with the aim of taking advantage of the increase in overall expression yield caused by $(SP)_{20}$ glycomodule (that confers stability) while allowing the obtainment of hEGF after the cleavage (without glycomodules). Since it was previously demonstrated that the 3HA was useless for secreted protein (*2.1. Fusion of 3XHA tagged Gaussia Luciferase reporter to the protein of interest to enable screening of transformants*), this tag was not included.

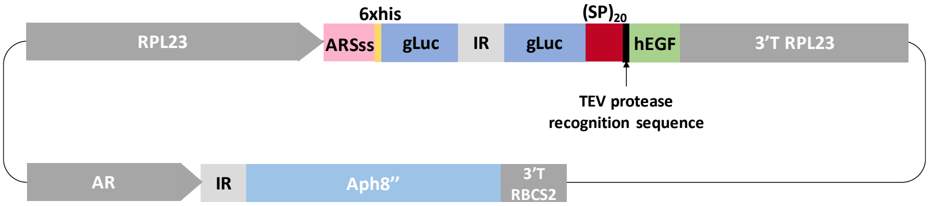


**Figure 41. Schematics of pPL-ARSss-gLucINTRON-$(SP)_{20}$-TEV-hEGF vector.** RPL23: 5' *RPL23* cis regulatory elements including 5'UTR and promoter, 3'T RPL23: 3'UTR and terminator of *RPL23*. ARSss: arylsulfatase 1 secretion signal, AR: *HSP70-RBCS2* chimeric promoter, IR: intron 1 *RBCS2*, 3'T RBCS2: 3' region of *RBCS2*.

Strains CC-124 and UVM4 were transformed with vector pPL-ARSss-gLucINTRON-$(SP)_{20}$-TEV-hEGF and 48 initial transformants (resistant to 60 µg hygromycin) were tested for gLuc activity. Results are shown in figure 42 as distribution of normalized luciferase expression in CC-124 and UVM4 transformants. This fusion protein gave RLU values significantly lower than the other fusion proteins comprising gLuc tested. This fact prevents the statistical comparison of the expression of this fusion protein to other fusion proteins by gLuciferase activity.

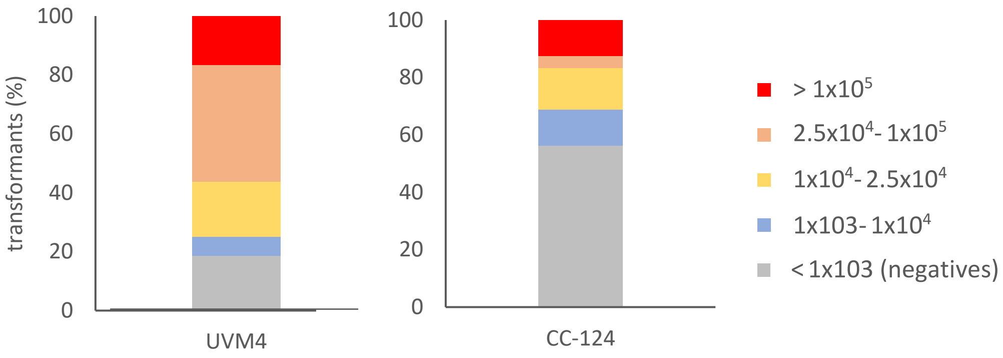


**Figure 42. Distribution of normalized luciferase expression (RLU) values from Chlamydomonas strains CC-124 and UVM4 transformed with pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF vector.** Assay was performed with 20 µl of culture supernatant. 48 transformants of each strain were analysed.

Expression of the full-length fusion protein was corroborated by immunoblot in the three highest expressing transformants obtained of each strain (figure 43).

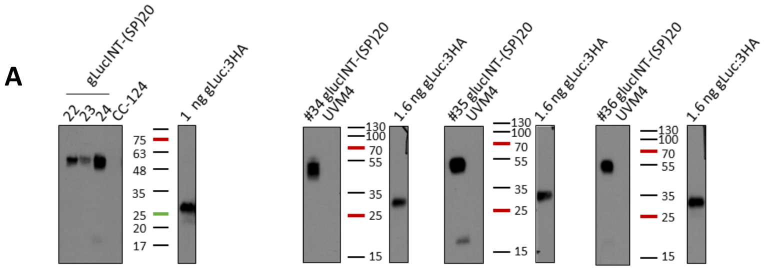
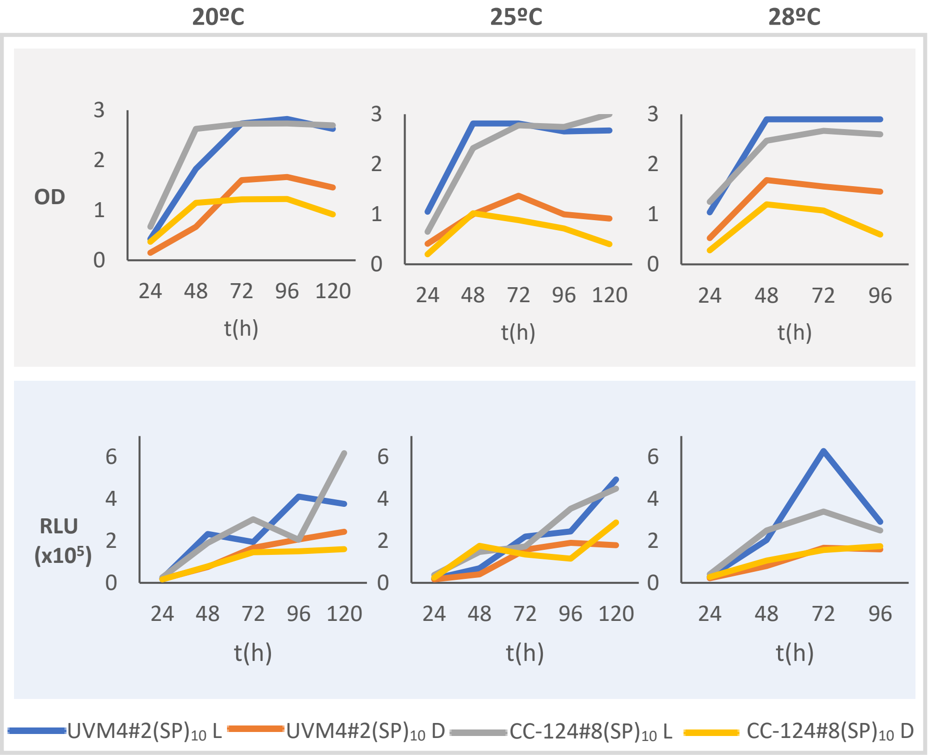
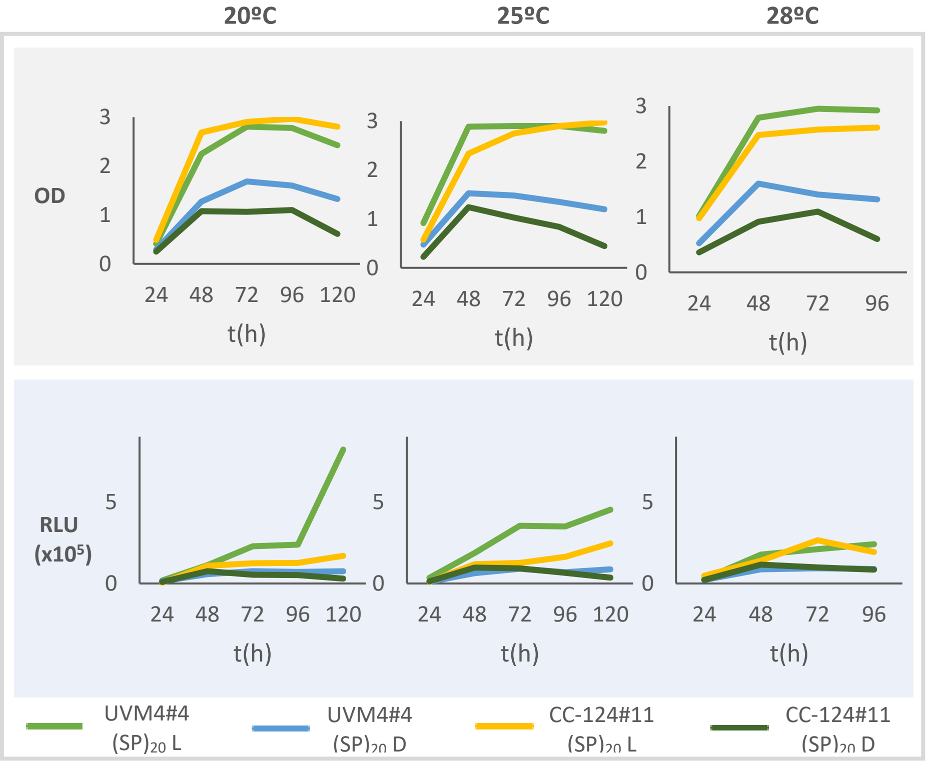

A
gLucINT-(SP)20
22
23
24
CC-124
1 ng gLuc:3HA
75
63
48
35
25
20
17
#34 glucINT-(SP)20
UVM4
1.6 ng gLuc:3HA
#35 glucINT-(SP)20
UVM4
1.6 ng gLuc:3HA
#36 glucINT-(SP)20
UVM4
1.6 ng gLuc:3HA
130
100
70
55
35
25
15

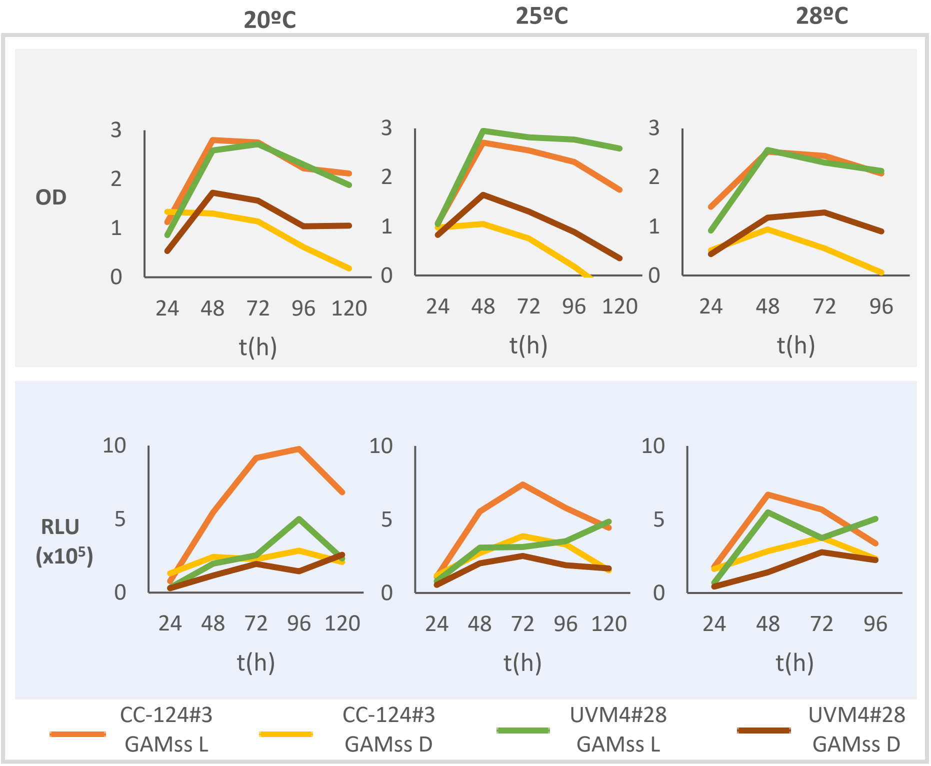
**Figure 43. Immunoblot of pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF highest expressing transformants. A) Transformants from CC-124 strain.** 12 µl of 20X concentrated culture supernatant were loaded. **B) Transformants from UVM4 strain.** 8 µl of 24X concentrated culture supernatant were loaded. gLucINT-(SP)20: pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF.

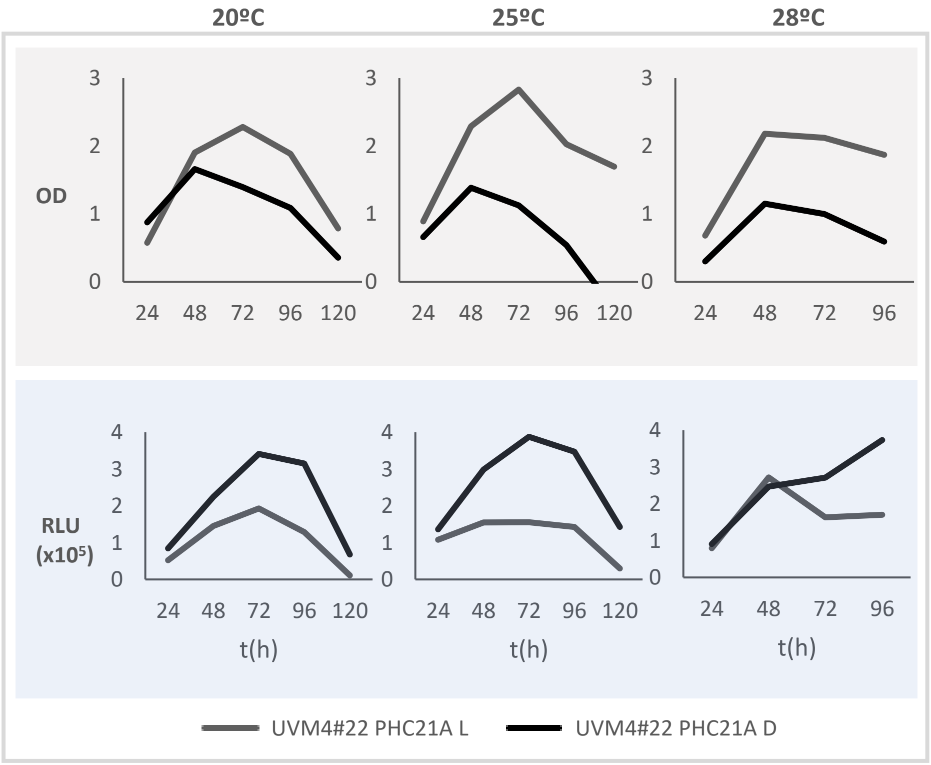
## 2.6. Characterization and quantification of transgene expression during cell culture growth

In order to analyse the expression of each fusion protein under different conditions, selected transformants were growth at 20ºC, 25ºC and 28ºC both under standard conditions (100 µE of light) and in dark. Growth was followed by OD750nm and RP expression was analysed by luciferase activity assay (figure 44). The transformants included in the assay were the highest transformant of CC-124 and UVM4 strains obtained for vectors pPL-GAMss-glucTEVhEGF-hygB$^R$, pPL-ARSss-glucTEVhEGF-(SP)10-hygB$^R$ and pPL-ARSss-glucTEVhEGF-(SP)20-hygB$^R$. The highest expressing transformant obtained for vectors pPL-ARSss-glucTEVhEGF-PHC21A-hygB$^R$ and pPL-ARSss-gLucINTRON-(SP)20-TEV-hEGF-hygB$^R$ were also included. No difference in growth was observed among CC124 and UVM4 transformants (results not shown). Significant differences could not be appreciated between growth at different temperatures. Immunoblots against gLuc were performed to complete the analysis of RP expression throughout culture stages (figure 44).


20ºC
25ºC
28ºC
OD
0
1
2
3
24
48
72
96
120
t(h)
RLU (x10$^5$)
0
2
4
6
UVM4#2(SP)$_{10}$ L
UVM4#2(SP)$_{10}$ D
CC-124#8(SP)$_{10}$ L
CC-124#8(SP)$_{10}$ D


20ºC
25ºC
28ºC
OD
0
1
2
3
24
48
72
96
120
t(h)
RLU (x10$^5$)
0
5
UVM4#4 (SP)$_{20}$ L
UVM4#4 (SP)$_{20}$ D
CC-124#11 (SP)$_{20}$ L
CC-124#11 (SP)$_{20}$ D


20ºC
25ºC
28ºC
OD
RLU (x10⁵)
0
1
2
3
5
10
24
48
72
96
120
t(h)
CC-124#3 GAMss L
CC-124#3 GAMss D
UVM4#28 GAMss L
UVM4#28 GAMss D


20ºC
25ºC
28ºC
OD
RLU (x10⁵)
0
1
2
3
4
24
48
72
96
120
t(h)
UVM4#22 PHC21A L
UVM4#22 PHC21A D

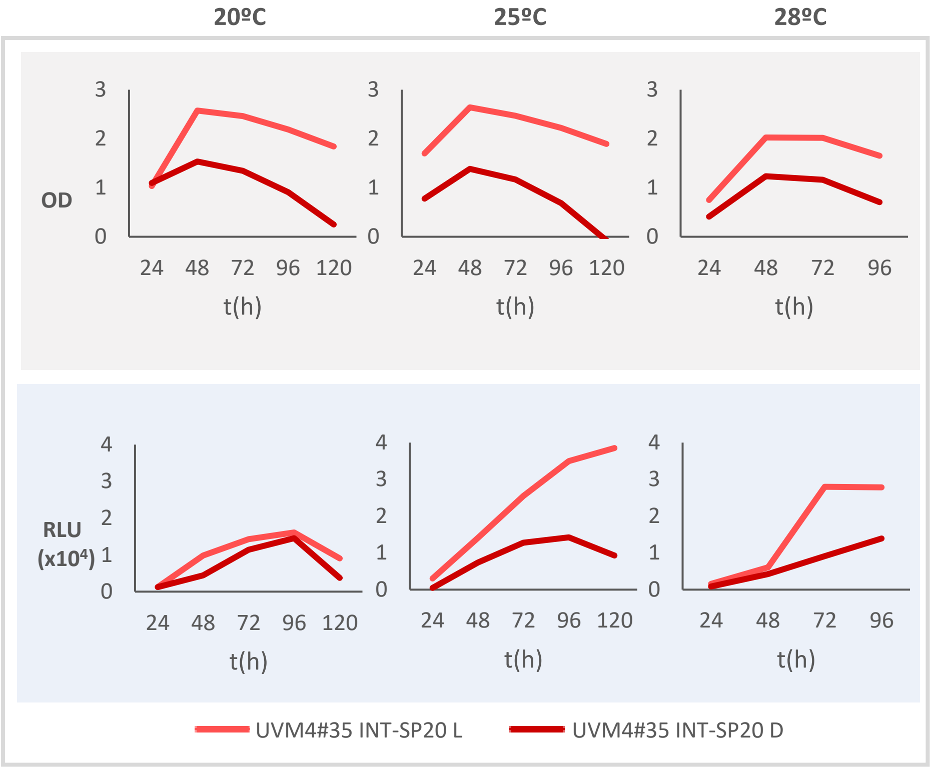


**Figure 44. Expression analysis of selected transformants by gLuc activity test at different growth stages under different growth conditions.** Each panel shows the growth followed by optical density at 750nm in the graphics located above. Each panel shows the gLuc activity assay results in the graphics located above. gLuc activity assay was performed with supernatant corresponding to 0.5 µl of culture medium. D: dark, L: light, GAMss: pPL-GAMss-glucTEVhEGF-hygB$^R$, (SP)$_{10}$: pPL-ARSss-glucTEVhEGF-(SP)$_{10}$-hygB$^R$, (SP)$_{20}$: pPL-ARSss-glucTEVhEGF-(SP)$_{20}$-hygB$^R$, PHC21A: pPL-ARSss-glucTEVhEGF-PHC21A-hygB$^R$, INT-(SP)$_{20}$: pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF-hygB$^R$.

In general terms, growth under standard conditions (light) performed better than growth in the dark both in terms of biomass and RP expression. The transformant UVM4#22 expressing the fusion protein gLuc:TEV:hEGF:PHC21A is the exception since its gLuc activity was higher under dark conditions.

Moreover, it was shown that there is accumulation of RP until stationary phase of culture. By immunoblot analysis it was shown that after the culture reaches the stationary phase there are degradation products of the fusion protein and thus the increase in gLuc activity seen in some cases may be due to such degradation releasing gLuc. The exception is transformant UVM4#35 expressing the fusion protein gLuc:$(SP)_{20}$:TEV:hEGF for which the immunoblot analysis and the gLuc activity analysis are consistent: there is a maximum of expression under light conditions at 96h (48h after reaching stationary phase of culture) and there is no detectable degradation of the fusion protein under any condition by immunoblot.

Finally, the immunoblot analysis confirmed higher degradation of fusion proteins in UVM4 strain transformants than in CC-124 strain transformants as previously seen (*2.4 Use of glycomodules to enhance expression and secretion of nuclear transgenes of Chlamydomonas - figure 40*).

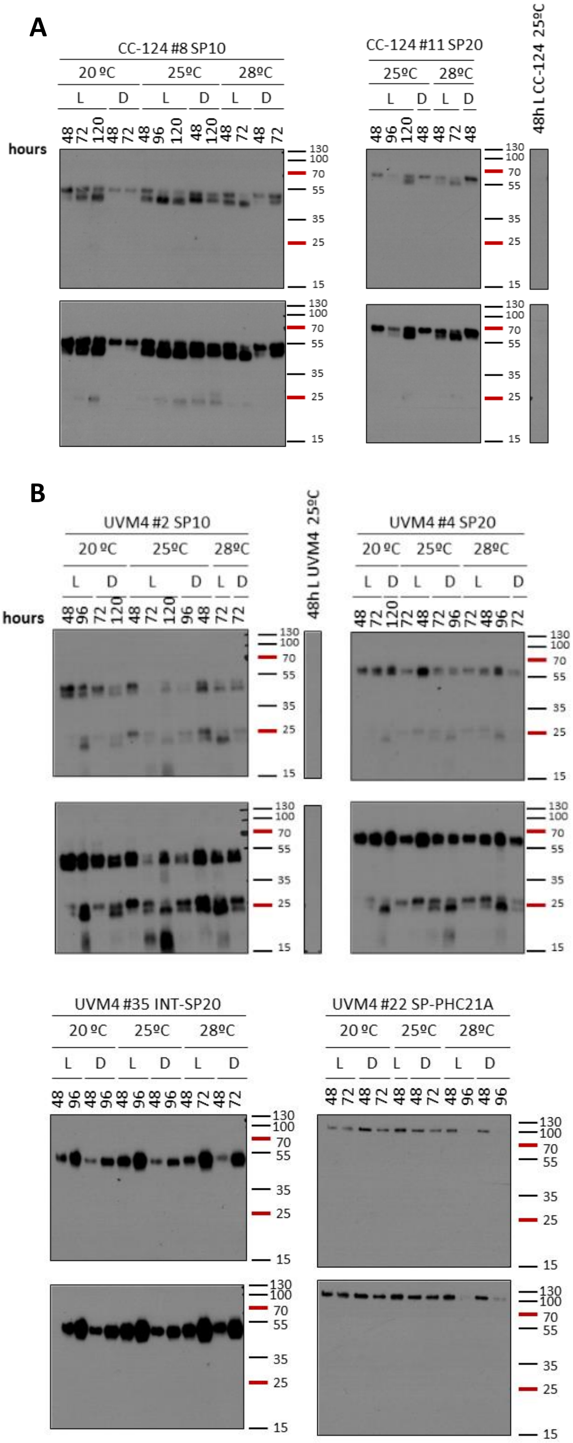
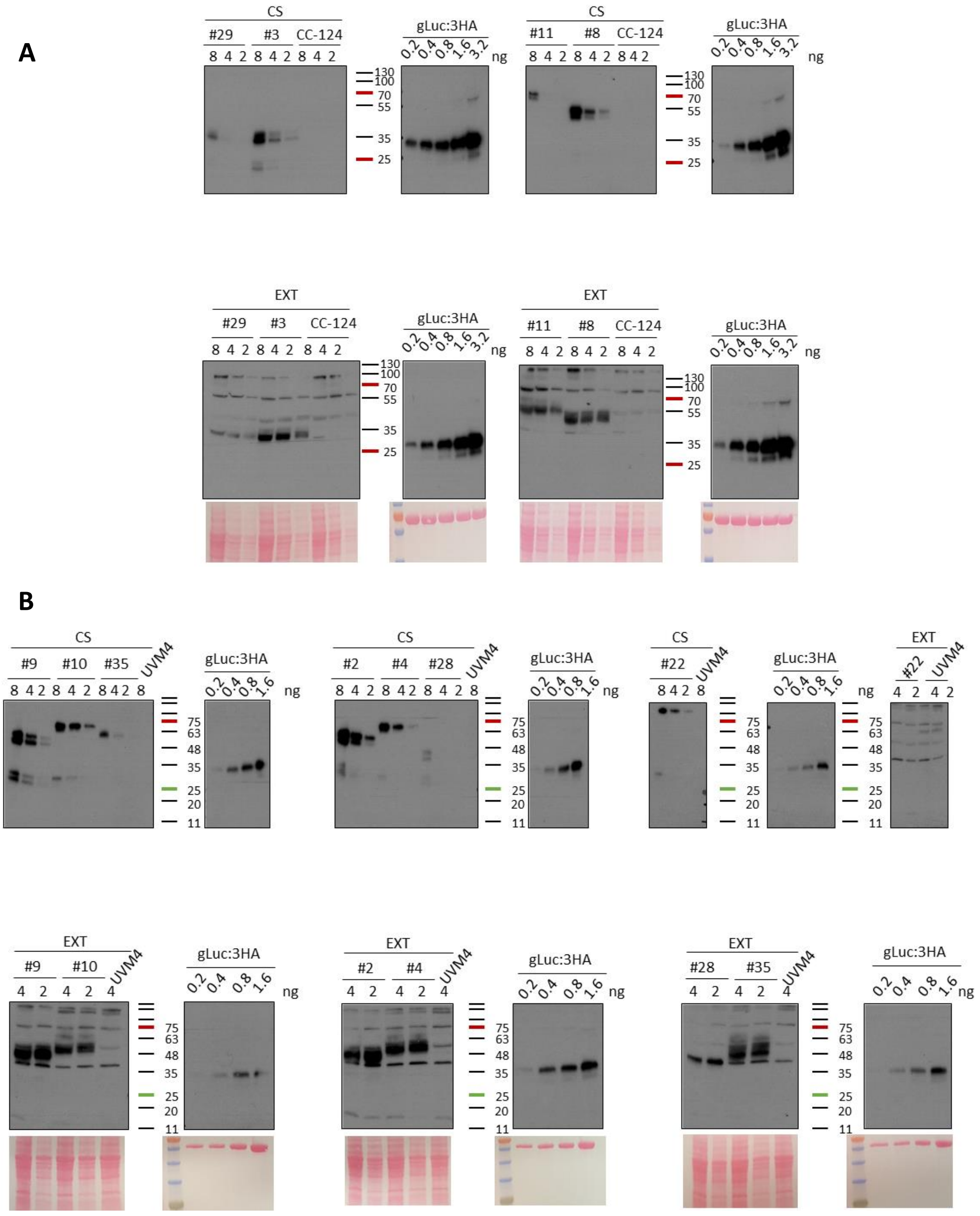

A
CC-124 #8 SP10
20 ºC
25ºC
28ºC
L D L D L D
hours
48 72 120 48 72 48 96 120 48 120 48 72 48 72
130 100 70 55 35 25 15
CC-124 #11 SP20
25ºC 28ºC
L D L D
48 96 120 48 48 72 48
48h L CC-124 25ºC
B
UVM4 #2 SP10
20 ºC 25ºC 28ºC
L D L D L D
48 96 72 120 48 72 120 96 48 72 72
48h L UVM4 25ºC
UVM4 #4 SP20
20 ºC 25ºC 28ºC
L D L D L D
48 72 120 72 48 72 96 72 48 96 72
UVM4 #35 INT-SP20
20 ºC 25ºC 28ºC
L D L D L D
48 96 48 96 48 96 48 96 48 72 48 72
UVM4 #22 SP-PHC21A
20 ºC 25ºC 28ºC
L D L D L D
48 72 48 72 48 72 72 48 96 48 96

**Figure 45. Immunoblot analysis of selected transformants at different growth stages under different growth conditions.** Two exposures are shown for each immunoblot: the one showed above corresponds to seconds of exposure and the one showed below corresponds to minutes of exposure. **A)** Immunoblot analysis of selected transformants from CC-124 strains: CC-124#8 (expressing the fusion protein gLuc:TEV:hEGF:$(SP)_{10}$) and CC-124#11 (expressing the fusion protein gLuc:TEV:hEGF:$(SP)_{20}$). 12 µl of concentrated culture supernatant (18X) were loaded in each lane. **B)** Immunoblot analysis of selected transformants from UVM4 strain: UVM4#2 (expressing the fusion protein gLuc:TEV:hEGF:$(SP)_{10}$), UVM4#4 (expressing the fusion protein gLuc:TEV:hEGF:$(SP)_{20}$), UVM4#22 (expressing the fusion protein gLuc:TEV:hEGF:PHC21A) and UVM4#35 (expressing the fusion protein gLuc:$(SP)_{20}$:TEV:hEGF). 12 µl of concentrated culture supernatant (18X) were loaded in each lane.

Summarizing, the optimal growth stage to harvest cultures for maximum RP expression is just once the culture reaches stationary phase except for the case of fusion protein gLuc:$(SP)_{20}$:TEV:hEGF for which maximum expression is observed 48 hours after have reached stationary. Moreover, no significant differences have been detected between expression at different temperatures. Given these results, selected transformants were grown at 25ºC under standard light conditions (100 µE) and were harvested at stationary phase of growth to estimate maximum expression of RP by densitometry based on immunoblots. It is important to highlight that the densitometry-based quantification of RP expression is estimated based on theoretical protein backbone, without considering the significant molecular weight provided by glycomodules. Densitometry-based quantification results are shown in table 9.

Since the fusion protein gLucINTRON:$(SP)_{20}$:TEV:hEGF accumulates without degradation until 48h after reaching stationary phase (figures 44 and 45) the quantification was also estimated based on a culture harvested 48h after reaching the stationary phase.


A
CS
#29
#3
CC-124
8 4 2 8 4 2 8 4 2
gLuc:3HA
0.2 0.4 0.8 1.6 3.2 ng
130
100
70
55
35
25
CS
#11
#8
CC-124
8 4 2 8 4 2 8 4 2
EXT
#29
#3
CC-124
EXT
#11
#8
CC-124
B
CS
#9
#10
#35
UVM4
8 4 2 8 4 2 8 4 2 8
gLuc:3HA
0.2 0.4 0.8 1.6 ng
75
63
48
35
25
20
11
CS
#2
#4
#28
UVM4
CS
#22
UVM4
8 4 2 8
EXT
#22
UVM4
4 2 4 2
EXT
#9
#10
UVM4
4 2 4 2 4
EXT
#2
#4
UVM4
EXT
#28
#35
UVM4

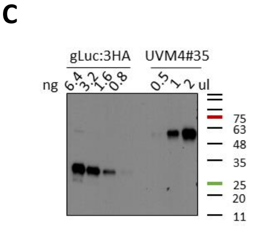

C
gLuc:3HA
UVM4#35
ng
6.4 3.2 1.6 0.8
0.5 1 2 ul
75
63
48
35
25
20
11

**Figure 46. Immunoblots used to estimate quantification of expression from CC-124 (A), UVM4 (B) transformants and (C) UVM4#35 transformant (culture supernatant harvested 48h after reaching stationary phase of culture).** Just above each immunoblot there is indicated the microliters of cell extract (EXT; 100X relative to culture) or culture supernatant (CS, 24X relative to culture) that were loaded in each lane. Ponceau stained membranes are shown only for cell extracts as a loading control because concentration of protein in concentrated culture supernatant is too low to be detected by ponceau staining.

**Table 9. Densitometry based quantification results in selected transformants.** There is indicated the optical density (OD750nm) of each culture at the moment of harvest. Quantification of glycomodules-containing protein estimated in reference to protein backbone. NQ: not quantified because signal was out of the range of the standard curve. ND: not determined.

| transformant | Fusion protein expressed | OD750nm | µg/L in CM | µg/L in EXT |
|---|---|---|---|---|
| UVM4#35 | gLucINTRON:$(SP)_{20}$:TEV:hEGF | 2.6 | 1.8 | 4.6 |
| UVM4#35 | gLucINTRON:$(SP)_{20}$:TEV:hEGF | 48 h after reaching stationary phase | 134.3 | ND |
| UVM4#9 | gLuc:3HA:TEV:hEGF:$(SP)_{10}$ | 2.7 | 3.5 | 5.2 |
| UVM4#10 | gLuc:3HA:TEV:hEGF:$(SP)_{20}$ | 2.7 | 6.9 | NQ |
| UVM4#2 | gLuc:3HA:TEV:hEGF:$(SP)_{10}$ | 2.8 | 12.7 | 4.4 |
| UVM4#4 | gLuc:3HA:TEV:hEGF:$(SP)_{20}$ | 2.6 | 29.3 | 8.9 |
| UVM4#28 | gLuc:3HA:TEV:hEGF | 2.8 | 2.1 | 2.5 |
| UVM4#22 | gLuc:3HA:TEV:hEGF:PHC21A | 2.7 | 10.4 | NQ |
| CC-124#8 | gLuc:3HA:TEV:hEGF:$(SP)_{10}$ | 2.3 | 11.0 | 3.2 |
| CC-124#11 | gLuc:3HA:TEV:hEGF:$(SP)_{20}$ | 2.5 | 3.5 | NQ |
| CC-124#3 | gLuc:3HA:TEV:hEGF | 2.5 | 8.2 | 3.2 |

As can be seen in table 9, RP yields from 1 to 100 µg RP /L of culture have been achieved for the different isoforms and the highest expression levels reported have been achieved with UVM4 strain. Also, extracellular expression levels are generally higher than intracellular expression levels.

## *Strategy 2. Study of the effect of transgene copy number on nuclear transgene expression levels*

Several factors such as insertion position greatly influence the expression of a transgene in Chlamydomonas. Due to that, there is a high variability in the expression of the same cassette in independent clones of Chlamydomonas and this makes it difficult to establish a correlation between copy number and expression level of transgene from published studies. Furthermore, to our knowledge, there hasn't been a systematic study to determine the relation between copy number and transgene expression in Chlamydomonas nucleus. To gain further insight into the effect of multicopy transgene on expression levels we designed two different strategies: 1) comparison of RP expression in diploid strains carrying two copies of transgene compared with its haploid counterparts, and 2) analysis of RP expression in haploid strains containing one or two copies of transgene.

### 1. Analysis of recombinant protein expression in diploid cells

Added to the increase of tolerance of diploid strains to various stresses, our interest in investigating Chlamydomonas diploids as RP expressing strains is based on the fact that they grow better in terms of biomass under nitrogen starvation conditions (Kwak et al. 2017). Since nitrogen becomes limiting when culture reaches stationary phase in a Chlamydomonas culture, we hypothesize that if diploid cells grow better under nitrogen starvation they would produce more RP at the latter stages of growth compared to their haploid counterparts. To our knowledge, the study of diploids to produce RP is an innovative approach that has not been previously tested in Chlamydomonas.

To study this approach, we used the cassette pPL-ARSss-glucTEVhEGF (figure 47) that codes for a fusion protein comprising gLuc as a reporter and hEGF as our protein of interest fused by a TEV protease recognition sequence and a 3HA tag, ARSss secretion sequence is placed just before gLuc sequence to direct the fusion protein to the periplasm. We took advantage of the previously used pPL-ARSss-glucTEVhEGF construct and replaced the gene *AphVII* (resistance to hygromycin b) for *AphVIII* gene (resistance to paromomycin).

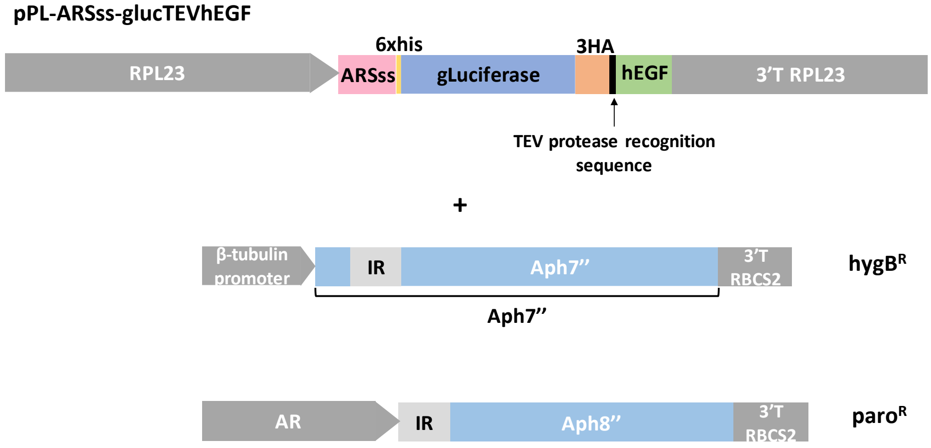

pPL-ARSss-glucTEVhEGF
6xhis
3HA
RPL23
ARSss
gLuciferase
hEGF
3'T RPL23
TEV protease recognition sequence
+
β-tubulin promoter
IR
Aph7''
3'T RBCS2
hygB^R
Aph7''
AR
IR
Aph8''
3'T RBCS2
paro^R

**Figure 47. Schematics of vectors used to study the effect of transgene copy number on Chlamydomonas nuclear transgene expression levels.** AR: *HSP70-RBCS2* chimeric promoter, RPL23: 5' *RPL23* cis regulatory elements including 5'UTR and promoter, 3'T RPL23: 3'UTR and terminator of *RPL23*, ARSss: Chlamydomonas arylsulfatase 1 secretion signal, IR: intron 1 *RBCS2*, brackets represent cDNA of *AphVII''* gene.

Both vectors (carrying cassette pPL-ARSss-gLucTEVhEGF and resistance to paromomycin or hygromycin B) were transformed into two cell walled strains with opposite mating type: CC-124 and CC-125. Then, all the possible transformants were picked to a maximum of 96. Screening through gLuc activity assay allowed the selection of independent transformants with similar expression levels: P1 and P2 were selected as mating type minus carrying resistance to hygromycin, P3 and P4 were selected as mating type plus carrying resistance to paromomycin.

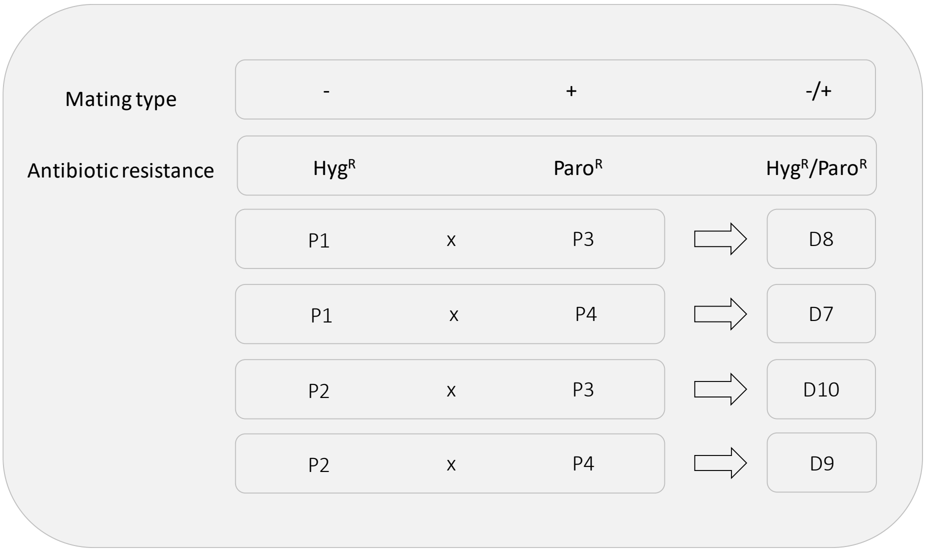

Mating type
-
+
-/+
Antibiotic resistance
Hyg$^R$
Paro$^R$
Hyg$^R$/Paro$^R$
P1 x P3
D8
P1 x P4
D7
P2 x P3
D10
P2 x P4
D9

**Figure 48. Schematics of crossings performed to obtain diploid progeny.** P; parental strain, D; diploid progeny.

The selected transformants were mated between them to obtain a diploid cell: P1 and P2 were both mated with P3 and P4. Eight diploid clones from each crossing were selected by their ability to grow on plates containing both antibiotics. Also, all progeny should be diploids since a maturation step required for zygote maturation was not included in the protocol. Confirmation of diploidy was performed by PCR looking for the presence of the minus-specific *MID* gene (minus dominance) or the plus specific *FUS1* gene (fusion) (Werner and Mergenhagen 1998). The results are shown in figure 49.

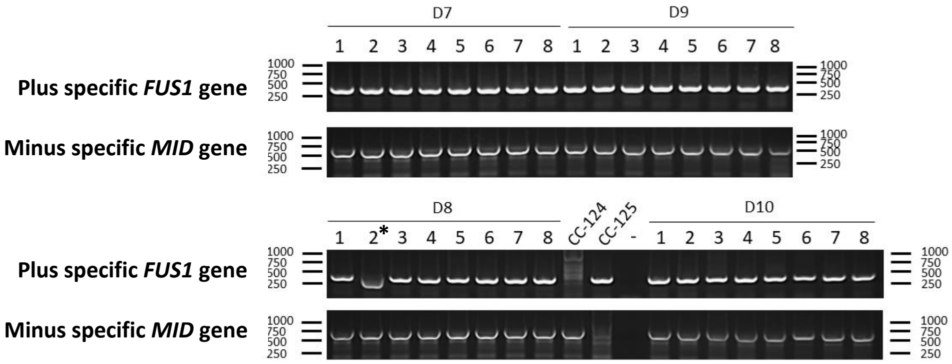


**Figure 49. Confirmation by PCR of the diploid nature of the selected progeny.** *Indicates a problem with the well that affected the electrophoresis of the sample.

Since the PCR analysis confirmed the diploid nature of the selected cells we analysed the expression of three diploid strains resulting from each crossing by luciferase activity assay through culture stages. Luciferase activity in culture supernatant was assayed with two technical replicates for each diploid strain and with two biological replicates and two technical replicates for each parental strain. As it can be seen in figure 50, all cultures were at stationary phase of growth at 48h with the exception of P2, which was detected to be at stationary phase of growth at 72h. Moreover, diploid progeny does not outperform its haploid background strains in terms of gLuc activity. Our initial hypotehsis was that an increase of gLuc activity may be caused by higher biomass accumulation in diploid strains at the end of the culture growth (due to nitrogen starvation). However, a longer exponential growth phase and higher RP expression levels were not detected in the diploid progeny analysed and thus our hypothesis must be discarded.

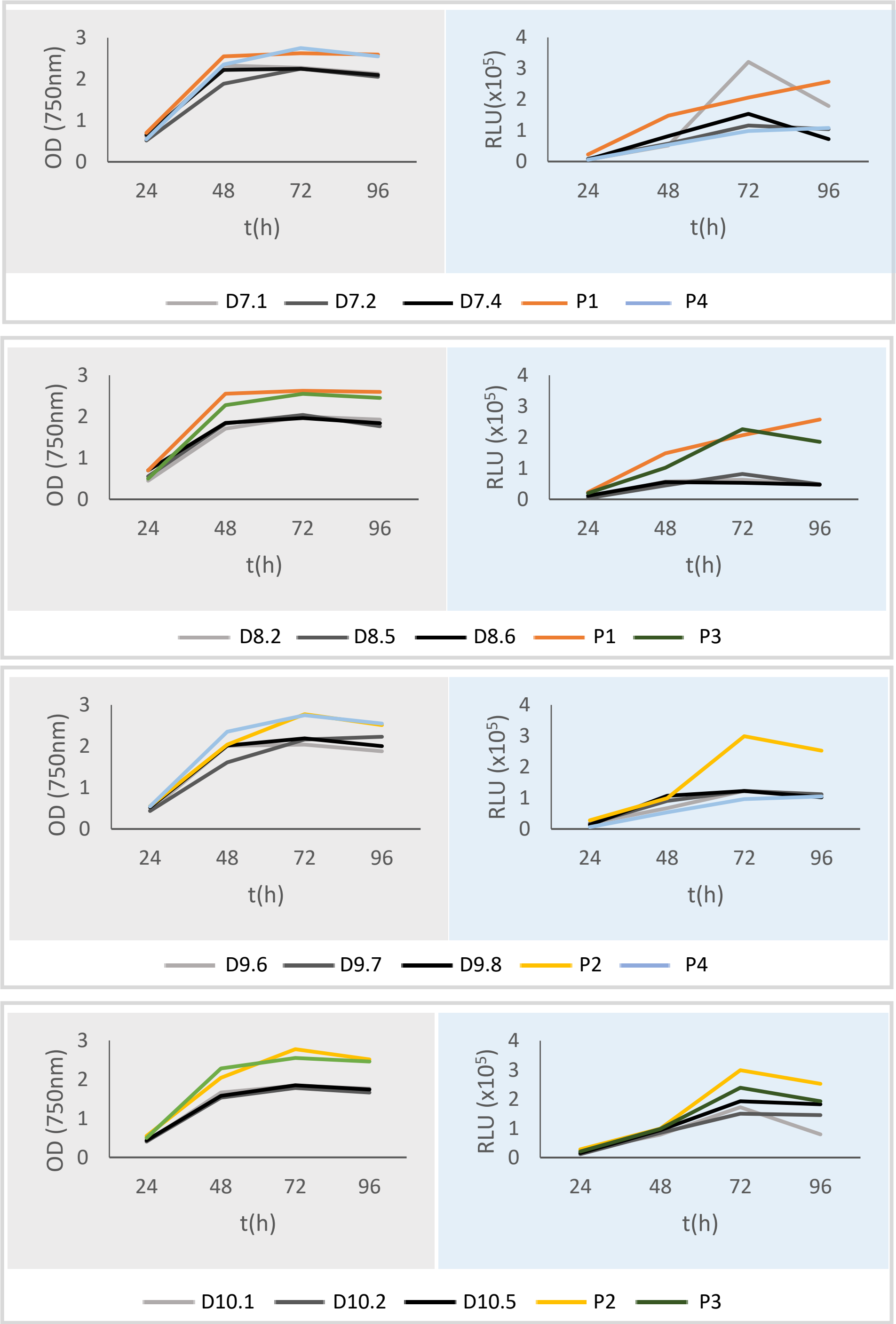


**Figure 50. Expression analysis of diploid cells by gLuc activity at different growth stages.** Left panels represent the growth of culture followed by optical density (OD). Only growth of one biological replicate is shown for each parental strain as there were not significant differences in growth between biological replicates. Right panels represent gLuc activity in relative luminescence units (RLU) assayed in culture supernatant with two technical replicates except for parental strains that were assayed with two biological and two technical replicates. D; diploid strain, P; parental strain.

## 2. Study of the relation between transgene copy number and recombinant protein expression in haploid cells

This approach aimed to investigate the relation between gene copy number and RP yield in the nucleus of haploid Chlamydomonas cells. To that end, our experimental design was based on generating strains containing more than one copy of transgene by crossing. Since parental strains are independent transformants, it is expected that they carry independent insertions because transgene is inserted at different locations.

To develop this strategy, we transformed a vector containing the cassette pPL-ARSss-glucTEVhEGF (figure 47) in combination with a second cassette carrying either *AphVIII* (paromomycin resistance) or *AphVII* gene (hygromycin resistance). Independent clones expressing RP to similar levels were selected as parental strains as indicated previously *(1. Analysis of recombinant protein expression in diploid cells)*. Parental strains (resistant either to hygromycin B or to paromomycin and expressing gLuc to similar levels) were mated to obtain a strain that contains two copies of the transgene as independent insertions. Meiosis product was analysed and, as expected, approximately 25% of the progeny was resistant to both antibiotics, a quarter was resistant only to paromomycin, a quarter was resistant only to hygromycin and a quarter was not able to grow on antibiotic selection. Selection of progeny was performed on the base of selectable markers: progeny resistant to each antibiotic (harbouring one transgene copy) or to both (harbouring two transgene copies) was selected.

Expression levels in culture supernatant were tested by luciferase activity assay with two technical replicates of two biological replicates and luminescence values were normalized by culture density (OD 750 nm). The gLuc activities of parental strains and progeny are shown in figure 51.

As it can be seen in figure 51, progeny harboring one transgene copy number have expression levels similar to the parental strains while progeny harboring two copies of the transgene have an expression level approximately 2-fold relative to parental strains. Summarizing, the preliminary results obtained indicate a direct relation between gene copy number and transgene expression from the nucleus of Chlamydomonas.

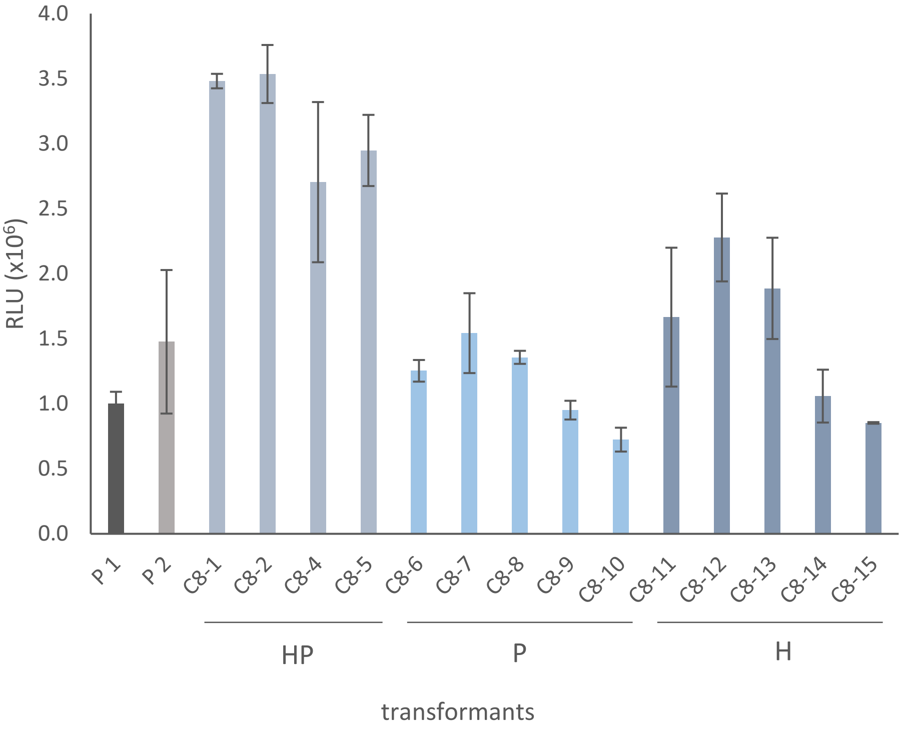


**Figure 51. Luciferase activity assessment of gene copy number effect on RP expression from Chlamydomonas nucleus.** P1: parental transformant mating type minus carrying resistance to hygromycinB; P2: parental transformant mating type plus carrying resistance to paromomycin; HP: transformants showing resistance to paromomycin and hygromycin B; P: transformants showing resistance to paromomycin; H: transformants showing resistance to hygromycinB. Results shown correspond to corrected RLU (RLU/OD750nm) assayed in culture supernatant. Error bars show the standard deviation between two biological replicates assayed with two technical replicates.

## *Strategy 3. Obtaining a Chlamydomonas strain with improved expression of nuclear transgenes*

We ideated a screening to isolate strains with impaired silencing mechanisms. Our experimental design is shown in figure 52.

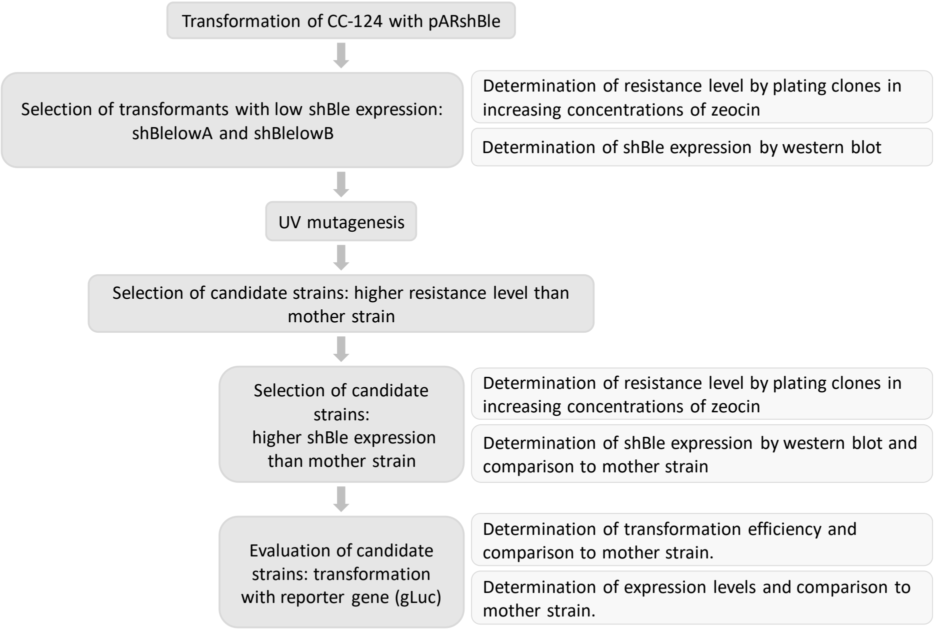

Transformation of CC-124 with pARshBle
Selection of transformants with low shBle expression: shBlelowA and shBlelowB
Determination of resistance level by plating clones in increasing concentrations of zeocin
Determination of shBle expression by western blot
UV mutagenesis
Selection of candidate strains: higher resistance level than mother strain
Selection of candidate strains: higher shBle expression than mother strain
Determination of resistance level by plating clones in increasing concentrations of zeocin
Determination of shBle expression by western blot and comparison to mother strain
Evaluation of candidate strains: transformation with reporter gene (gLuc)
Determination of transformation efficiency and comparison to mother strain.
Determination of expression levels and comparison to mother strain.

Figure 52. Schematic diagram of the experimental proposal to screen high expressing strains.

We took advantage of the fact that there is a high variability in the expression of the same cassette from independent clones of Chlamydomonas by selecting one transformant with low expression levels. Our experimental approach is based on the mutagenesis of this low expressing transformant and a screening to obtain a mutant with improved expression of nuclear transgenes. Our screening tool is shBle$^R$ as it expression is directly proportional to its phenotype (see *1.1.1 Use of minimal FMDV 2A peptide for processing of a dicistronic vector in Chlamydomonas*) thereby allowing a fast and easy screening. Our screening is based on the idea that genetic inactivation (by mutation through UV) of the silencing mechanism would reduce the transcriptional repression of shBle$^R$ gene, thus allowing growth on higher zeocin concentrations (Neupert et al. 2009; Kurniasih et al. 2016). We used a cassette carrying shBle$^R$ driven by AR chimeric promoter and 3'UTR of *RBCS2* (figure 53).

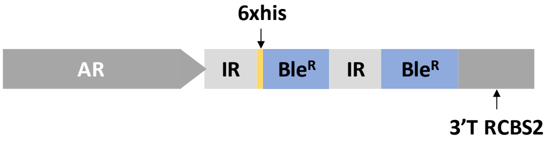
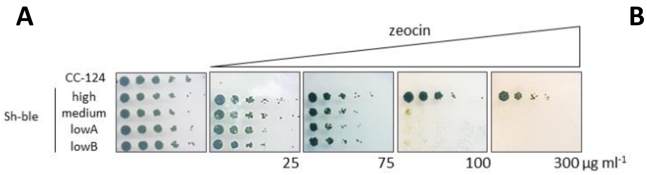


**Figure 53. Schematics of pAR-his-shBle vector.** AR; chimeric promoter *HSP70-RBCS2,* IR; intron 1 *RBCS2*, 3'T RBCS2: 3'UTR and terminator of *RBCS2*.

Once transformed into Chlamydomonas strain CC-124, 96 transformants were initially selected on TAP-zeocin 15 µg/ml (considered low zeocin) and screened by their ability to grow in presence of increasing concentrations of zeocin. The strains shBlelowA and shBlelowB, which grew on 25 µg/ml zeocin but were sensitive to 100 µg/ml zeocin (and thus displayed low expression of shBle) were chosen for the subsequent mutagenesis experiment. Furthermore, from the same screening, we chose one strain that displayed resistance to medium zeocin concentrations (75 µg/ml) and one strain resistant to high zeocin concentrations (300 µg/ml). These two strains were called shBlemedium and shBlehigh respectively and were used as controls. Expression level of ShBle by these strains was corroborated by immunoblot (Figure 54).



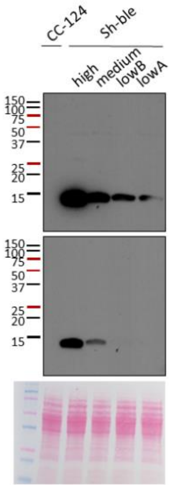


**Figure 54. Zeocin sensitivity test and immunoblot of selected strains for mutagenesis.** ShBlehigh and ShBlemedium were selected as controls, ShBlelowA and ShBlelowB were selected as background strains for UV mutagenesis. **A) Zeocin sensitivity test of selected strains.** Sensitivity test was performed by plating the same number of cells at increasing concentrations of zeocin. **B) Immunoblot of selected strains.** 50 µg TSP were loaded in each lane, primary antibody against shBle. Upper panel corresponds to an exposure of minutes and lower panel corresponds to an exposure of seconds. Ponceau stained membrane is shown as a loading control.

Conditions for UV mutagenesis were optimized to establish an assay that would led to a survival of 10-30% of the exposed cells (figure 55): between 200 x 100 µJ and 300 x 100 µJ of UV exposure.

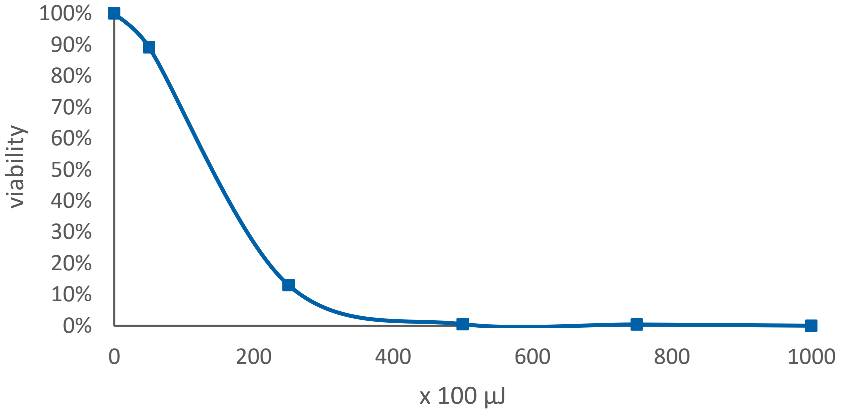


**Figure 55. Viability of mutagenized cells with increasing irradiation.**

Two independent experiments were performed with shBleLowA and ShBleLowB strains. In both cases mutagenized cells were plated on 100 µg/mL of zeocin. We selected all the mutants able to grow on 100 µg/ml zeocin, thus displaying an increased zeocin resistance: 32 mutants of ShBlelowA and 96 mutants of ShBlelowB. It is important to note that the assay included a non-irradiated control to corroborate that the background strain did not give rise to colonies able to grow at 100 µg/ml zeocin and thus the colonies that grew were the result of UV irradiation.

Subsequently, the zeocin resistance level of each individual clone was determined to identify those mutants that displayed the strongest resistance to zeocin. To do so, all the mutants obtained (32 mutants of ShBlelowA and 96 mutants of ShBlelowB) were plated on increasing concentrations of zeocin. This step also represented a confirmation for their ability to grow in presence of higher zeocin concentrations compared to it parental strains. This growth performance assay revealed differences in the resistance level between mutants that could be due to different mutations leading to different zeocin resistance levels. All clones that were able to grow at a concentration of at least 100 µg/ml zeocin (resistance to four times higher concentration than its background strain) were selected as candidates: 8 resulting of the mutagenesis of shBlelowA (figure 56-A) and 19 resulting of the mutagenesis of shBlelowB (figure 57-A). Furthermore, we analysed the shBle protein expression level of selected mutants by immunoblot to corroborate the higher expression compared to it parental strains (figure 56-B and figure 57-B).

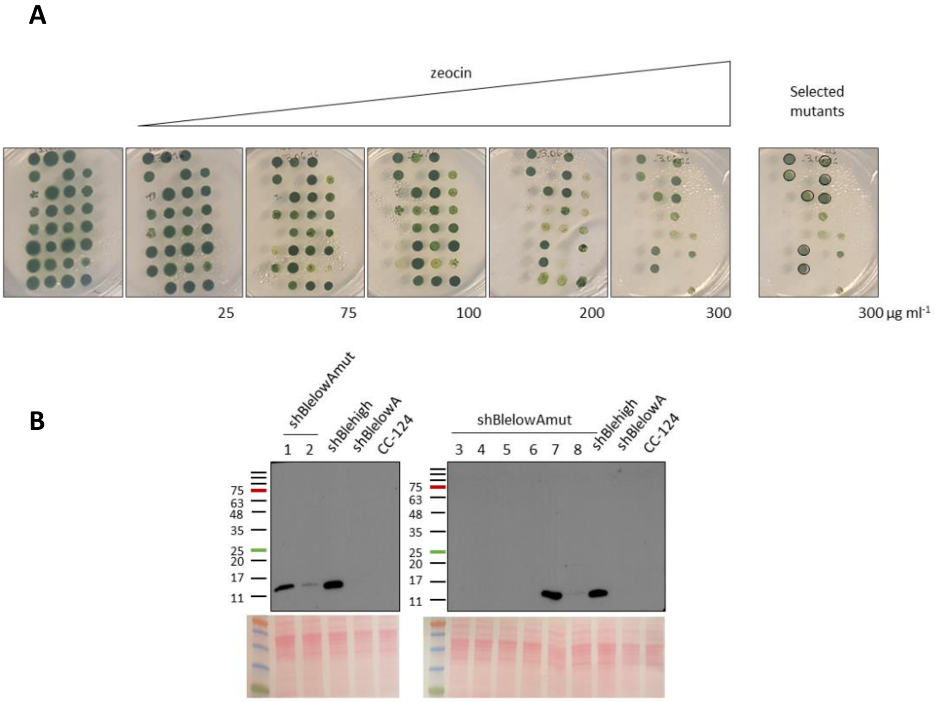

A
zeocin
Selected
mutants
25
75
100
200
300
300 µg ml-1
B
shBlelowAmut
1
2
shBlehigh
shBlelowA
CC-124
shBlelowAmut
3
4
5
6
7
8
shBlehigh
shBlelowA
CC-124
75
63
48
35
25
20
17
11
75
63
48
35
25
20
17
11

**Figure 56. Zeocin sensitivity test and immunoblot analysis of mutants obtained from strain shBlelowA. A)** zeocin sensitivity test of shBlelowA mutants; Mutants were spotted onto TAP agar plates containing increasing concentrations of zeocin. **B)** Immunoblot of shBlelowA mutants; 40 µg TSP loaded in each lane, primary antibody anti shBle. Membrane stained with Ponceau showed below immunoblot exposure as a loading control.

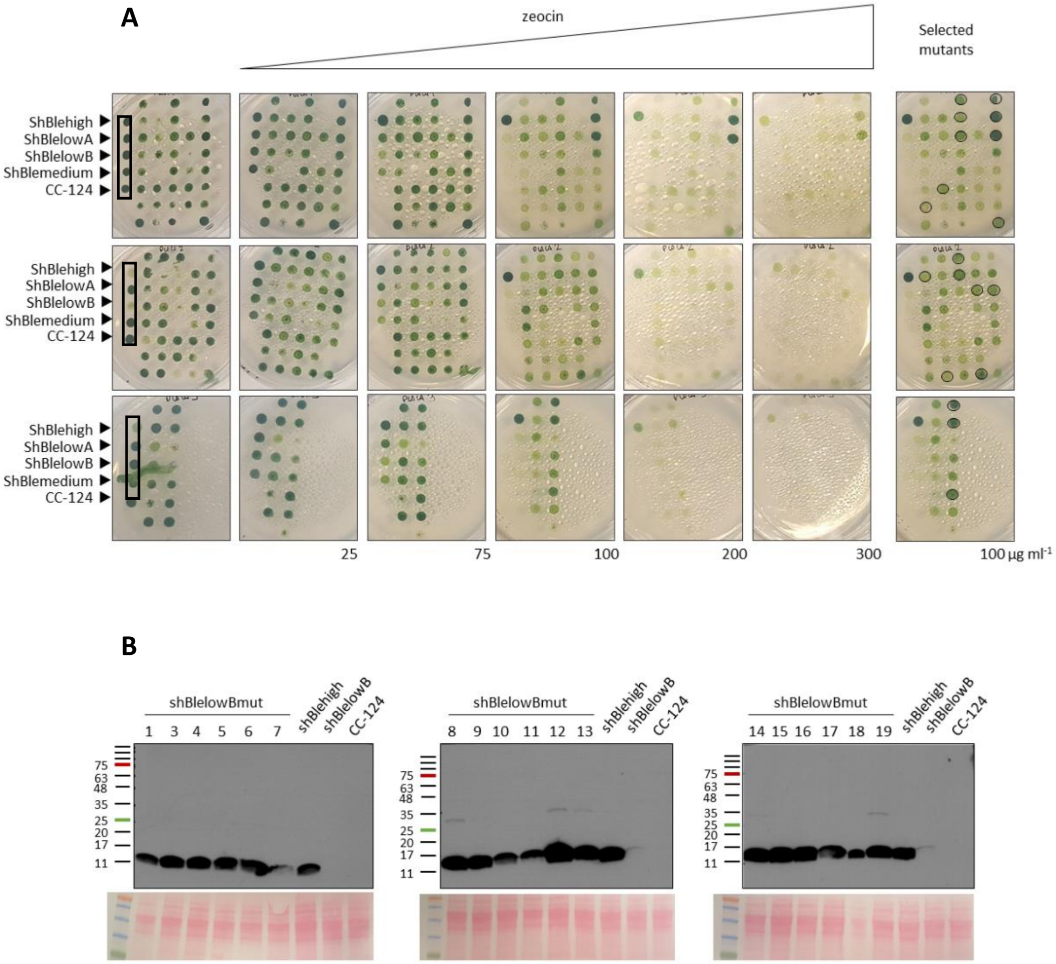

A
zeocin
Selected mutants
ShBlehigh
ShBlelowA
ShBlelowB
ShBlemedium
CC-124
25
75
100
200
300
100 µg ml-1
B
shBlelowBmut
1 3 4 5 6 7
shBlehigh
shBlelowB
CC-124
8 9 10 11 12 13
14 15 16 17 18 19
75
63
48
35
25
20
17
11

**Figure 57. Zeocin sensitivity test and immunoblot analysis of mutants obtained from strain shBlelowB. A)** zeocin sensitivity test of shBlelowB mutants; Mutants were spotted onto TAP agar plates containing increasing concentrations of zeocin. **B)** Immunoblot of ShBlelowB mutants; 40 µg TSP loaded in each lane, primary antibody anti shBle. Membrane stained with Ponceau are shown below immunoblot exposure as a loading control.

From this point on, we continued only with those mutants that showed at least the same shBle protein level by immunoblot than our control shBlehigh: all the clones of shBlelowB except for lowB#7, and only lowA#1 and lowA#7 from the shBlelowA mutants.

Following the reasoning beyond our experimental design: higher zeocin resistance, increased shBle protein and enhanced nuclear transformation efficiencies were the three expected indicators of reduced gene silencing when compared with background strains. At this point we had already checked the first two. However, both can be caused by a mutation that increases only the expression of ShBle: mutations enhancing promoter, mRNA stability or chromatin accessibility as examples. To further confirm or reject the putative high expressing clones these clones were transformed with a reporter gene: gLuc gene was our choice. We transformed the selected candidates with vector pPL-ARSss-glucTEVhEGF-hyg$^R$ (*2.2 Comparison of the efficiency of two different cis regulatory regions to drive transgene expression,* figure 29) which carries two cassettes: i) the gene *AphVII*' conferring resistance to hygromycin B driven by β-tubulin cis regulatory elements (5' and 3') and ii) the cassette ARSss–gLuc-3HA tag-TEV-hEGF-6xhis tag, the whole cassette driven by cis regulatory regions (5' and 3') of Chlamydomonas Ribosomal protein L23 (*RPL23*, Cre04.g211800) (López-Paz et al. 2017). Transformation experiments always included the wild type strain CC-124 (background strain of shBlelowB and shBlelowA) and the background strain ShBlelowB or ShBlelowA. We found variable transformation efficiencies between experiments, even for the control strains (CC-124 and ShBlelowB). Nonetheless, as an increased transformation efficiency could be one indicator for a mutant with reduced gene silencing, transformation efficiency fold increase related to background strain CC-124 is shown in figure 58. Although most mutants did not show increased transformation efficiencies there were promising strains: mutB3, mutB12, mutB13, mutB14, mutB15 and mutB19.

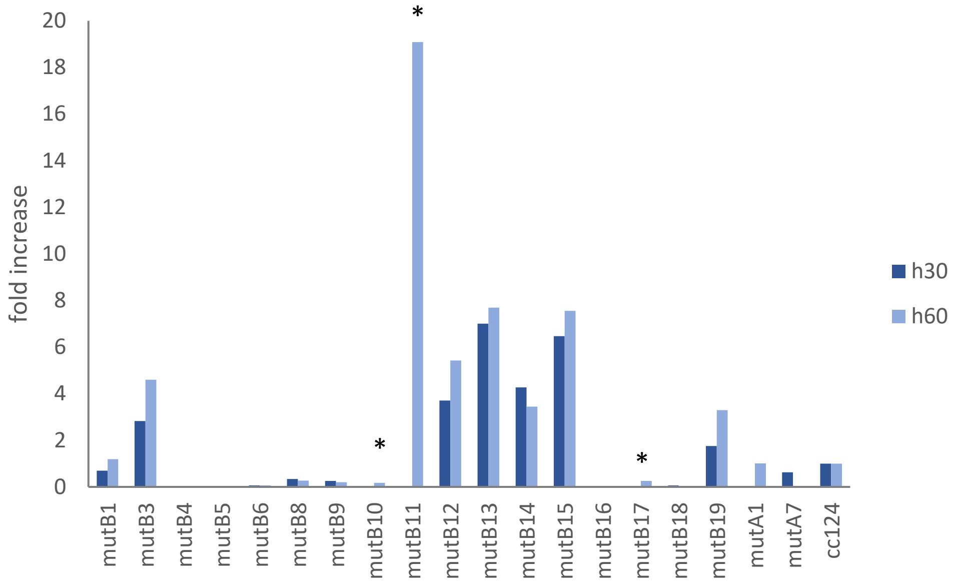


**Figure 58. Transformation efficiency (number of colonies obtained/µg transformed DNA) fold increase of the mutants referred to CC-124.** Transformation efficiency was determined at two different antibiotic concentrations (hygromycin 30 or 60). * indicates mutants able to grow in 30 µg/ml of hygromycin B without transformation and thus there is only measure of transformation efficiency at 60 µg/ml of hygromycin B.

Independently of the transformation efficiency obtained we further analysed the expression levels of the transformants of all our mutants by luciferase activity assay except for mutB16 for which we did not obtain transformants. Luminescence values were normalized by culture density (OD 750 nm) and expressed as relative luciferase units (RLUs). Transformants that showed luciferase activity above background (>1 x $10^3$ RLUs) were considered positives. We used two metrics to assess performance in each mutant. The first was the proportion of positive transformants (luciferase activity above background: >1 x $10^3$ RLUs) that was statistically analysed by Fisher's exact test. The second metric was the distribution of positive RLU values binned by expression level and was statistically analysed by Wilcoxon-Mann-Whitney rank sum test. Both metrics were analysed between each mutant and its parental strains (CC-124 and either shBlelowB or shBlelowA). Differences were considered significant when p < 0.05. Results are shown only for transformants with significant p-values in table 10.

Regarding the mutants of shBlelowA strain (mutA1 and mutA7), the luciferase activity results for both mutants and shBlelowA were lower that the obtained for wild type strain CC-124 and consequently no statistical analysis was pursued with these strains. The fact that shBlelowA gave transformants with generally lower expression levels than CC-124 transformants may indicate a mutation that reduces its recombinant protein expression.

**Table 10. P-values resulting of the statistical analysis to determine candidates to be a high expressing strain.** Only results for transformants with significant p-values are shown. Statistical analysis was performed with two metrics: proportion of positive transformants (Fisher's exact test) and expression level of these positive transformants (Wilcoxon-Mann-Whitney rank sum test). Statistical analysis was performed against CC-124 strain and shBlelowB background strain. Results were considered significant when p < 0.05. Significant p-values are highlighted with an asterisk. Whitney RST; Wilcoxon-Mann-Whitney rank sum test. positives; number of transformants which RLU/DO750nm is superior to 1000.

| | against CC-124 | | against shBlelowB | |
|---|---|---|---|---|
| Mutant | Fisher | Whitney RST | Fisher | Whitney RST |
| MUTB12 | 0.835 | 0.022* | 0.527 | 0.011* |
| MUTB13 | 0.001* | 0.759 | 0.000* | 0.359 |
| MUTB14 | 0.101 | 0.376 | 0.040* | 0.646 |
| shblelowB | 0.831 | 0.313 | | |

As indicated by the results presented in table 10, only mutants mutB12 and mutB13 remain as candidates to be high expressing strains. On one hand, there was a significant increase in expression levels for positive transformants of mutB12 compared with shBlelowB and CC-124 positive transformants determined by a two-tailed Mann-Whitney U test (p < 0.05). On the other hand, although its frequency of positives is significantly higher, MutB13 does not show significantly higher expression levels. However, instability of hEGF may hamper the correct analysis of high expressing transformants by limiting the fusion protein accumulation. To further confirm or reject the high expressing strains we tested the accumulation of the fusion protein gLuc:$(SP)_{20}$:TEV:hEGF (*2.5 Introduction of an intronic sequence within reporter gene to enhance nuclear expression in Chlamydomonas* - figure 41) that has shown an increased stability thereby avoiding possible difficulties in analysing expression levels due to instability. Screening of transformants was performed as exposed and the metrics used to statistically analyse the results were the same stated before. There was a significant increase ($p = 0.0046$) in expression levels for positive transformants of mutB13 compared with CC-124 positive transformants determined by a two-tailed Mann-Whitney U test. In this case, there was not a significant increase in expression levels for positive transformants of mutB12. Nevertheless, the highest expressing transformants from mutant backgrounds showed higher expression level than the highest expressing strains from CC-124 background (figure 59).

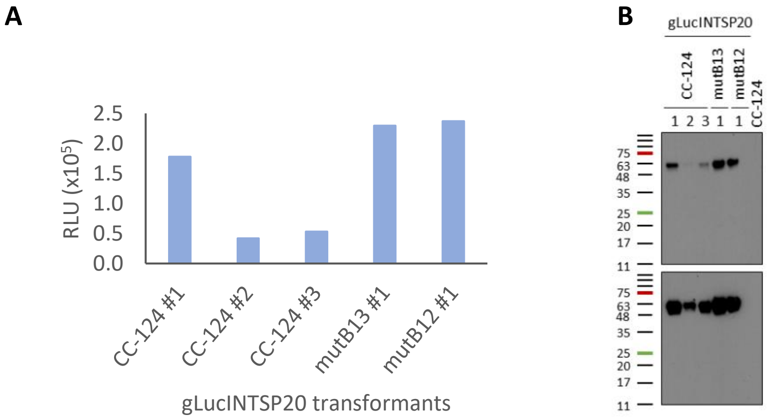


**Figure 59. Comparison of pPL-ARSss-gLucINTRON-$(SP)_{20}$-TEV-hEGF highest expressing transformants from CC-124, mutB12 and mutB13 strains. A) gLuciferase activity of selected transformants.** Assay was performed with 20 µl of total culture. **B) Immunoblot analysis of selected transformants.** 12 µl of 18X concentrated culture supernatant were loaded. Primary antibody against gLuc. Two exposures are shown: the one showed above corresponds to seconds of exposure and the one showed below corresponds to minutes of exposition. gLucINT-$(SP)_{20}$: pPL-ARSss-gLucINTRON-$(SP)_{20}$-TEV-hEGF.

To further confirm that mutB12 and mutB13 strains carry a mutation that helps transgene expression, these strains were mated with CC-125 pPL-ARSss-glucTEVhEGFhygB$^R$ selected transformants (*Strategy 2- 1. Analysis of recombinant protein expression in diploid cells* - figure 47). The expression of both transgenes (gLuc and shBle) was analysed in the progeny. It is expected that progeny containing gLuc or shBle transgene will be divided in two populations: high or low expression (referred to transgene levels). If mutation segregates independently of transgene high expression levels will correlate with the mutation improving transgene expression.

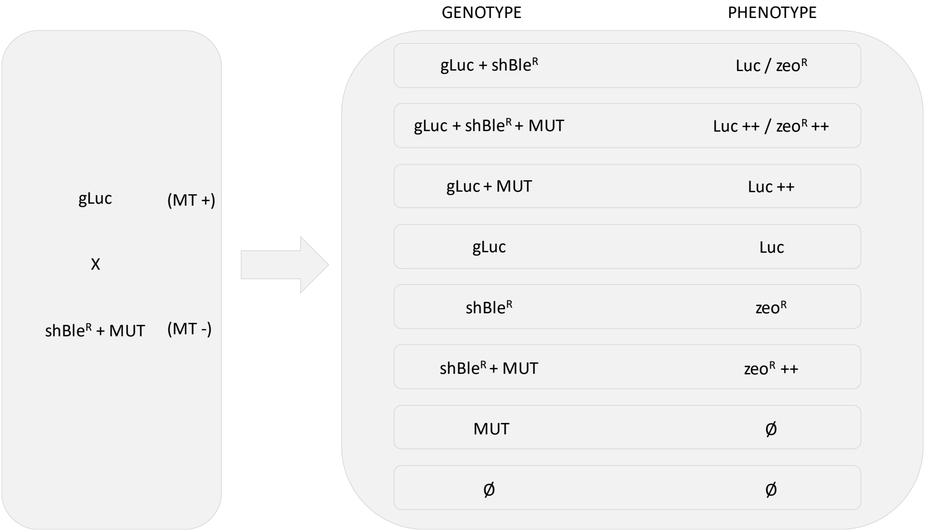


**Figure 60. Resulting progeny from a cross between a high expressing strain and gLuc wild type transformants.** gLuc: gLuciferase gene. shBle$^R$: resistance gene to bleomycin family of antibiotics, MUT: mutation facilitating transgene expression, Luc: luciferase activity, zeo$^R$: resistance to zeocin, ++: refers to higher expression levels, MT: mating type.

A cross between CC-124 and CC-125 pPL-ARSss-glucTEVhEGFhygB$^R$ transformant was included as a control. gLuciferase activity levels greatly varied between progeny of the same crossing even for the control cross. These differences may be due to growth under non-optimal conditions in 96-well plates that cause variability in transgene expression. Also, differences between parental strains CC-125 and CC-124 that are not completely isogenic may add variability. The high number of progeny clones needed made the assay not feasible with cultures grown under optimal conditions. The difficulties in analysing gLuc expression levels in the progeny hampered the confirmation or rejection of the presence of a mutation favouring transgene expression in our high expressing strain candidates.

# Discussion

In the context of recombinant proteins there is a challenge to develop expression systems and methodologies appropriate to address the cost and demand issues for a variety of applications (Fletcher et al. 2007; Ferrer-Miralles et al. 2009). One of the alternative expression hosts proposed is *Chlamydomonas reinhardtii*. Despite all the advantages of this expression host (described in Introduction, page 12), the low expression levels of nuclear transgenes represent a limitation to its use for recombinant protein production. In chapter 1 we test several strategies designed to reduce or overcome this limitation. Through design of the DNA cassette on the base of a secreted fusion protein comprising a gLuciferase and hEGF, expression levels ranging from 1 to 100 µg RP /L of culture have been achieved.

## *DNA cassette design and optimization to increase nuclear RP expression yields*

Our starting point was the fusion strategy of Ble$^{R}$2AGOI that had been previously reported as successful to select high expressing clones of Chlamydomonas (Rasala et al. 2012, 2014; Plucinak et al. 2015). Two versions of 2A were tested. However, the range of products or variability in transgene expression obtained makes it difficult to establish a correlation between protein expression and resistance to zeocin. This may be due to poor processing and/or intermediate fusion proteins that result in poor efficiency of shBle protein by preventing its biological function when compared to monomeric shBle. To summarize, despite having been previously reported efficient in Chlamydomonas (Rasala et al. 2012, 2014; Plucinak et al. 2015) we could not validate the Ble:2A:GOI strategy to select high transgene expressing transformants at least for the proteins tested. The fact that the proteins tested may be considered unstable (may be the case of hEGF) or were not adapted to Chlamydomonas codon usage (as is the case of mCherry) may explain the differences in processing efficiency by 2A peptides. Moreover, other studies have reported partial processing with 2A peptides (Rasala et al. 2012; Kong et al. 2015; López-Paz et al. 2017). Also, strain and/or growth conditions may contribute to the variability in processing between studies. Interestingly, it was reported that when using the fusion consisting of shBle:2A and an enzyme from *Botryococcus braunii*, zeocin resistance level was not correlated with accumulated RP (Kong et al. 2015). They also reported the inability to sort highly expressing transformants of the target protein among the transformants that survived high zeocin concentrations. In the light of our results, their results may be attributable to poor processing and/or intermediate fusion proteins that result in poor efficiency of shBle protein by preventing its biological function when compared to monomeric shBle.

We decided to keep Ble$^R$ as a marker gene because we were able to increase expression of shBle by maintaining selected transformants on antibiotic containing plates. We further studied the approach based on the fusion comprising Ble$^R$ and GOI by fusing it separated by a TEV protease recognition sequence. It was observed that the screening of clones by resistance to high zeocin concentrations leads to the selection of high expressing strains of shBle protein, but not to the selection of high expressing transformants of the complete fusion protein. The analysis of the insertion of the heterologous cassette showed that the selected high expressing clones had not inserted the complete cassette. Different scenarios were found on cassette integration: insertion of the resistance gene only, insertion in a coding region (transcribing for a chimeric fusion protein consisting of our designed protein and a native protein), insertion of the cassette with deletions that may transcribe for a partial fusion protein or not. These results support two possible explanations, not excluding between them: i) fusion protein results in poor efficiency of shBle protein by preventing its biological function when compared to monomeric shBle or ii) instability of the fusion protein (Ble:TEV:hEGF) hampers the selection of high resistant transformants expressing the fusion protein. Only a small fraction of selected transformants had inserted the complete heterologous cassette thus expressing the full-length protein. The study of one transformant expressing the full-length protein allowed the estimation of the RP expression yield in approximately 1 µg/L culture. This expression level is low and does not allow RP production for industrial purposes. Furthermore, expression levels are below the expression level of 0.25% TSP previously reported in Chlamydomonas with the Ble2A strategy, an strategy also based on selection of transformants resistant to high concentrations of zeocin (Rasala et al. 2012). However, in this strategy the GOI was an enzyme and shBle and GOI were successfully expressed as discrete proteins. It is important to remark that differences in processing efficiency affecting the overall RP yield may be related with the nature of the GOI. Indeed, different yields of different RP are reported in the present work (monomeric shBle compared to fused shBle to hEGF or to fused ShBle to a region of a native Chlamydomonas protein).

Regarding the approach based on the fusion comprising Ble$^{R}$ and GOI, one of the transformants isolated expressed RP to high levels (80 µg/L). We studied the expressed RP and it consisted of part of our designed fusion protein and part of a native protein. Thereby, this chimeric protein is the result of a partial insertion of the heterologous cassette in frame in a coding region. This corroborates that the partial insertion of the cassette leads to different transgenes and prevents the selection of high expressing transformants of the full-length fusion protein with the strategy based on shBle:TEV:hEGF. Regarding the chimeric protein, it is possible that the native protein region increases the stability or solubility of the fusion protein, but it remains to be determined. On the other hand, the gene encoding for this protein has six introns, and one is 997 bp length. Given that the average for introns length in

Chlamydomonas is 373 bp (Merchant et al. 2007) this intron represents an interesting regulatory region to test.

Interestingly, it was observed that selected clones expressing fusion proteins comprising shBle duplicated their expression level by maintenance in presence of zeocin for two months. A model to explain the evolution of antibiotic resistance in bacteria has been proposed in which a combination of continued selection, epigenetic inheritance and stochastic variation allows for the evolution of populations with gene expression patterns providing an increasing antibiotic resistance (Adam et al. 2008). The model is based on the idea that within an isogenic population there is random variation in the expression level of genes, creating phenotypic variation. This epigenetic variation shows an element of heritability and some cells with elevated expression of genes conferring antibiotic resistance survive antibiotic selection. The progeny of this cells will in turn show variation in expression levels and if selective process continues it will lead to an increase of the resistance level. As it has been shown that transgenerational epigenetic effects play a role in adaptive evolution in Chlamydomonas (Kronholm et al. 2017), the model proposed for bacteria may explain our results. Maintenance in presence of zeocin will enhance expression of transgenes for which the strategy shBle:GOI is suitable.

The strategy based on the fusion protein comprising Gaussia Luciferase as a reporter gene and hEGF as GOI is successful as it allows the high throughput screening of transformants leading to the highest expressing transformants of the full-length protein. With this expression strategy it could be observed that the fusion protein is accumulated in medium independently of strain type (with or without cell wall) and the maximum of expression through culture stages corresponds to the initial stationary phase of growth except for a fusion protein that accumulates in medium for 48h in stationary phase. This latter fusion protein consists of shBle:$(SP)_{20}$:TEV:hEGF and the fact that accumulates in medium for longer periods induces us to think that the inclusion of stabilizing regions may improve expression yields. We report that the maximum of RP expression through culture stages is protein dependent, being consistent with previously reported results that show accumulation of a fusion protein comprising gLuc and a GOI until stationary phase of culture (Lauersen et al. 2013b) and secreted GFP until 4 days later (Ramos-Martinez et al. 2017). Moreover, the effect of culture conditions on yield at each growth stage has been reported (Lauersen et al. 2015a) and remains to be determined for our fusion proteins.

Interestingly, despite detecting intracellular and extracellular fusion protein by immunoblot against gLuciferase, only intracellular fusion protein was detected by immunoblot against HA tag. It is likely that a PTM prevents the detection of the 3HA tag from secreted proteins in Chlamydomonas. A similar effect has been shown in a secreted RP with a FLAG tag from insect cells. In this case, tyrosine O-sulfation abolished the FLAG-anti-FLAG interaction rendering a well-established tag system unsuitable for secreted proteins (Schmidt et al. 2012). Tyrosine sulfation is a PTM that occurs in secreted and transmembrane proteins of higher eukaryotic species and in all mammalian cell lines tested, but not in prokaryotes or in yeast. The functional importance of tyrosine sulfation in many proteins is not clear although it often functions in protein-protein interactions (Yang et al. 2015). It seems that sulfation does not depends on a defined consensus sequence (Rosenquist and Nicholas 1993) but three to four acidic amino acid residues are normally found within five residues of sulfated tyrosines (Rosenquist and Nicholas 1993). Moreover, it has been proposed that the charge of the residue in the amino-terminal (-1) position of the tyrosine is critical and must be neutral or acidic, whereas a basic residue abolishes sulfation (Lin et al. 1992). In the aminoacidic sequence of 3HA tag there are tyrosines that present the characteristics to be sulphated and according to sulfinator tool results (https://web.expasy.org/sulfinator/) at least one of these tyrosines would be sulfated. To rule out or confirm this hypothesis, several sulfatase reactions may be performed with secreted RP to medium. However, if a PTM prevents the detection of the 3HA tag from secreted proteins in Chlamydomonas remains to be determined.

Comparison between two cis regulatory regions (5’ and 3’) allowed to determine that the expression of RP driven by cis regulatory regions (5’ and 3’) of Chlamydomonas Ribosomal protein L23 (*RPL23*, Cre04.g211800) (López-Paz et al. 2017) was significantly better in terms of expression yield than expression driven by AR promoter (*HSP70-RBCS2* chimeric promoter including *RBCS2* first intron downstream) and 3’UTR *RBCS2*.

Since it has been shown that the use of alternative signal peptides affects the efficiency of protein secretion and the overall yield of secreted proteins (Knappskog et al. 2007; Kober et al. 2013; Molino et al. 2018), different secretion peptides have been studied in terms of expression yield and secretion efficiency: Chlamydomonas arylsulfatase 1 secretion sequence (ARSss), metalloprotease gametolysin secretion sequence (GAMss) and Carbonic Anhydrase 1 secretion sequence (CAHss). Expression level resulted significantly higher with CAHss compared to GAMss and ARSss and resulted significantly higher with the use of GAMss compared to ARSss. Therefore, changing and optimizing secretion sequences represents a tool for improving transgene expression.

A successful strategy to enhance yields and stability of secreted RP in plant cell cultures has been demonstrated based on (Ser-Pro)n glycomodules (Shpak et al. 1999; Xu et al. 2007). In previous studies in Chlamydomonas, this glycomodules conferred an enhanced protein stability and allowed accumulation of RP in culture medium. Due to the enhanced stability, yields of a (Ser-Pro)n-fused fluorescent protein increased up to 12 fold compared to fluorescent protein without glycomodules reaching a maximum yield of 15 mg/L (Ramos-Martinez et al. 2017). In the present work, a significant increase in the expression yield of a secreted fusion protein comprising hEGF was achieved either with $(SP)_{10}$ and $(SP)_{20}$ glycomodules. Furthermore, a fusion protein comprising the $(SP)_{20}$ glycomodule after the gLuciferase gene and before the TEV protease cleavage sequence was designed with the aim to take advantage of the stabilizing effect of the glycomodule region together with the possibility of obtaining discrete hEGF (without glycomodules) after TEV protease cleavage. The highest expression levels reported in the present study were achieved with this fusion protein. This increased expression levels may be due to: i) an intronic sequence placed inside the gLuc gene, ii) the $(SP)_{20}$ glycomodule located between the reporter gene and the GOI and iii) the lack of 3HA tag. While all the other secreted fusion proteins accumulated until stationary phase of growth, this fusion protein accumulated for 48h at stationary phase of culture. It is unlikely that this effect is due to the intronic sequence placed inside the gLuc gene or due to the elimination of the 3HA tag. Contrary to the other fusion proteins, there is no detectable degradation of this fusion protein by immunoblot suggesting that the $(SP)_{20}$ glycomodule placed N-terminally of hEGF confers higher stability. Despite that, the effect of both the intronic region and the lack of the 3HA tag cannot be analysed separately from the effect of $(SP)_{20}$ glycomodule in the present work. Further studies are required to analyse the effect of the three regions separately.

Given that fusion proteins comprising $(SP)_{10}$ and $(SP)_{20}$ glycomodules showed an enhanced stability allowing accumulation in medium of RPs, we investigated the existence of native glycomodules from Chlamydomonas in its more abundant secreted proteins that would confer higher stability to such proteins and thus may be used for enhancing RP yields. For that purpose we studied the presence of putative glycomodules in the most abundant Chlamydomonas proteins present in medium under different conditions (Baba et al. 2011). Possible glycomodules (not necessarily with $(SP)_n$ motif) presenting several combinations and lengths of repetitive sequences were identified. We tested five sequences representing different native glycomodules by cloning them C-terminally to hEGF and none of this native glycomodules surpassed the performance of $(SP)_{10}$ and $(SP)_{20}$ glycomodules. However, PHC21A native glycomodule significantly increased the expression yield in comparison to the fusion protein without glycomodule. It is important to note that when we studied the glycomodules present in the most abundant secreted proteins in Chlamydomonas, we detected that one protein may contain several glycomodule region suggesting that the introduction of glycomodules at more than one site in the fusion protein may increase the stability and overall yield of expressed RP. Also, since the highest expression yield reported in the present study was achieved with the fusion protein gLuc:$(SP)_{20}$:TEV:hEGF, it is interesting to further study PHC21A native glycomodule cloned C-terminally of gLuc to test the maximum expression yield achievable with such glycomodule. Also, it is interesting to test the expression of a fusion protein carrying more than one glycomodule sequence.

It is important to highlight that statistics to compare regulatory regions, secretion peptides or stabilizing regions have been performed on the base of gLuciferase activity and this activity could be dependent on fusion protein. Despite that, the direct relation between gLuciferase activity and fusion protein has always been corroborated by immunoblot. Since gLuc activity may overestimate RP because of degradation, expression yields have been always estimated by densitometry of immunoblots.

The highest expression levels reached in the present study have been achieved using UVM4 strain, a mutant strain that was isolated because it has an elevated expression of transgenes (Neupert et al. 2009). However, this strain showed higher degradation of recombinant proteins than the wild type strain

CC-124. Also, UVM4 culture supernatant presented higher viscosity when concentrated by filtration than that of CC-124 strain (data not shown). These facts represent a limitation for scalability and downstream processing of UVM4 strain. Since Chlamydomonas cell wall is composed of glycoproteins and carbohydrates (Ehrenberg et al. 1971; Roberts et al. 1972) such components are released to medium during growth. Indeed, it has been reported that cell wall less strains release more than twice the amount of protein per cell than the wild type (Zhang and Robinson 1990). As a result, it is likely that the higher viscosity found in UVM4 culture supernatant is due to a higher content of glycoproteins and carbohydrates in its medium.

With the strategy based on a fusion protein comprising a reporter gene and a GOI, secreted RP yields have been increased from 1 to 100 µg/L by testing several regulatory regions, secretion peptides and stabilizing regions. It must be considered that the vector used to achieve the highest expression level reported in the present work do not include all the best strategies tested. Furthermore, merging of the best strategies becomes mandatory as a next step together with further analysis of native SPs cloning more than one and cloning one placed C-terminally of gLuc. It is likely that, with these steps, expression yield will be further increased. To our knowledge, the highest expression levels reported in Chlamydomonas nucleus were achieved with a fusion protein comprising GFP and glycomodules that reached a yield of 12 mg/L (Ramos-Martinez et al. 2017). Despite not reaching such expression yields, our results are not directly comparable because GOI are different.

### *Effect of transgene copy number on nuclear transgene expression levels*

The present study investigates gene dosage as a strategy to improve expression levels in the nucleus of Chlamydomonas from two approaches: 1) comparison of RP expression in diploid strains carrying two copies of transgene compared with its haploid counterparts, and 2) analysis of RP expression in haploid strains containing one or two copies of transgene.

Regarding diploid cells, our initial hypothesis was that an increase of RP may be caused by higher biomass accumulation in diploid strains at the latter culture growth stages (due to nitrogen starvation). However, there is not an increase in gLuc activity at the end of the culture growth in the diploid progeny analysed and thus our hypothesis must be discarded. However, further study of diploid cells for production of RP may still be interesting because they have previously shown an increased resistance to stresses (Kwak et al. 2017).

Preliminary results have been obtained regarding the direct relation between copy number and transgene expression level in haploid cells. Such preliminary results indicate a direct relation between gene copy number and nuclear transgene expression. Further studies are needed to confirm the preeliminary results obtained.

### *Isolation of Chlamydomonas strains with improved expression of nuclear transgenes*

The experimental approach designed to isolate mutagenized strains with increased transgene expression led to the selection of two candidate strains. Both candidate strains have an increased shBle expression level compared to its background strain (shBlelowB) but when tested for expression of other transgenes, the increase in expression levels is not always statistically significant. We intended to study the presence of a favouring mutation by mating the candidate strains and studying its progeny. However, the difficulties in analysing transgene expression in the progeny hampered the confirmation or rejection of the presence of a mutation favouring transgene expression in our strain candidates. These differences may be due to growth under non-optimal conditions in 96-well plates that cause variability in transgene expression added to differences between parental strains CC-125 and CC-124 that are not completely isogenic. When tested for the expression of a reporter protein, the highest expressing transformants from mutB12 and mutB13 strains showed higher expression levels than the highest expressing transformants from CC-124 background. However, there are different expression levels between transformants suggesting that expression of nuclear transgenes in our candidate strains is still subjected to silencing mechanisms similarly as reported with other isolated high expressing strains (Kong et al. 2015). It has been shown that silencing in Chlamydomonas occurs at transcriptional and post-transcriptional levels (Cerutti 1997; Wu-Scharf et al. 2000; Jeong Br et al. 2002). If the favouring mutation affects a transcriptional silencing mechanism the increase in expression would not be significant if the mRNA or protein expressed are unstable or silenced. Due to that it is important to test the candidate strains with several transgenes. Also, if the favouring mutation affects a post-transcriptional silencing mechanism, the transgene will be still subjected to transcriptional silencing and to position effect. Given that multiple mutations can occur during UV mutagenesis, it would be interesting to map the mutation(s) and to study if it can favour transgene expression in an additive manner with other described mutations on silencing mechanisms (Wu-Scharf et al. 2000; Zhang et al. 2002; Casas-Mollano et al. 2008; Kong et al. 2015). Since the isolated strains are wild type they can be mated to accumulate mutations favouring transgene expression obtaining a powerful strain for high transgene expression.

# Methodology

## *Chlamydomonas strains and culture conditions*

Algal cells were cultivated mixotrophically in liquid Tris-acetate-phosphate (TAP) medium at 25ºC under continuous illumination at 150 µE and 200 rpm orbital shaking unless otherwise stated. Strains were routinely maintained on TAP agar plates every three weeks.

The Chlamydomonas strains used throughout the study are: CC-125, CC-124, CC-1690 (21gr), CC-1883, UVM4 and UVM11. Strains CC-1690 (21gr), CC-125 and CC-124 were obtained from the Chlamydomonas Resource Center (http://chlamycollection.org/). Strains UVM4 and UVM11 are two mutants graciously provided by Prof. Dr. Ralph Bock (Neupert et al. 2009).

Optical density at 750nm was determined either with BioTek PowerWave XS plate reader or with a spectrophotometer UVmini1240 (Shimadzu).

## *Construction of expression cassettes*

Two cis regulatory regions were used in this study to drive expression of GOI. On one hand, *HSP70A/RBCS2* chimeric promoter (Schroda et al., 2000) further improved by insertion of the *RBCS2* first intron downstream (Sizova et al., 2001), abbreviated as AR in the present study, was used in combination with 3'UTR *RBCS2*. On the other hand, *RPL23* (Cre04.g211800) cis regulatory elements (5' and 3') were used. PCR amplification from CC-124 Chlamydomonas strain genome allowed us to obtain 5'UTR *RPL23* (F409/R410) and 3'UTR *RPL23* (F411/R412).

TRX h genomic sequence was amplified (F223/R222) from Chlamydomonas strain CC-124 to clone it in our vectors. The sequence of the mature form of hEGF (53aa) was codon-optimized for Chlamydomonas nuclear expression and synthesized by GeneArt (Life Technologies, USA). The mCherry Ds gene was obtained from the plant binary vector CD3-960 (Nelson et al. 2007). A codon-optimized Gaussia princeps luciferase (gLuc) gene was synthesized (Ruecker et al. 2008). Table 12 resumes all the cloning done to complete the present study.

**Table 11. Sequences of oligonucleotides used for vectors construction in this study.**

| Oligonucleotides names | Oligonucleotides sequence (5'-3') |
|---|---|
| **R222- TRX h F** | CCCCTGCAATGCTTTTAC |
| **F223- TRX h R** | ATGGGCGGTTCTGTTATT |
| **F409 – 5' RPL23 F** | GGCCTCTAGAGCACGTCCATAATGAAAGGTC |
| **R410- 5' RPL23 R** | CGGAAGAATTCACTCTGCGCGCAGACAAGAG |
| **F411- 3'UTR RPL23 F** | GGGCATATGTAAAGGTGAGGCGGGGGTTCG |
| **R412- 3'UTR RPL23 R** | CCAGGTACCCCAGGCGCAGCTTCGCC |
| **F485- $(SP)_{10}$ F** | TATGAGCCCCTCCCCCAGCCCGAGCCCTTCGCCTTCGCCCTCGCCGTCGCCATCCCCGAGCCCCCTGCAGTAAGA |
| **R486- $(SP)_{10}$ R** | TATCTTACTGCAGGGGGCTCGGGGATGGCGACGGCGAGGGCGAAGGCGAAGGGCTCGGGCTGGGGGAGGGGCTCA |
| **F487- GAMss F** | TATCGAATTCACCATGTCGCTGG |
| **R488- GAMss R** | GGCAGATCTGGCCCACGCAGGAGCTGTGC |
| **F504- GP1 F** | TATGCCTCCTTCGCCGGCTCCTCCATCGCCGGCTCCTCCATCGCCGGCTCCTCCATCGCCGGCTCCTCCTTCGCCGGCTCCTCCATCGCCGGGCCTGCA |
| **R505- GP1 R** | GGCCCGGCGATGGAGGAGCCGGCGAAGGAGGAGCCGGCGATGGAGGAGCCGGCGATGGAGGAGCCGGCGATGGAGGAGCCGGCGAAGGAGGCA |
| **F506- LCL F** | TATGAGCAACTCCCCTGCCCCCGCCTCGTCGGCCCCTGCCCCGGCCCGCTCCTCGTCGGCCAGCTGGCGTGACGCCCCCGCCTCGAGCTCCAGCTCGAGCCTGCA |
| **R507- LCL R** | GGCTCGAGCTGGAGCTCGAGGCGGGGGCGTCACGCCAGCTGGCCGACGAGGAGCGGGCCGGGGCAGGGGCCGACGAGGCGGGGGCAGGGGAGTTGCTCA |

**Table 12. Clonings performed for the present study.**

| Vector | Select. gene | Content | Cloning |
|---|---|---|---|
| **pARshBle2A** | Ble$^{R}$ | AR:Ble2A:RBCS2 | AR chimeric promoter (XbaI/EcoRI) fused to BleR gene: FMDV 2A (EcoRI/PstI, XhoI) and 3'UTR (NdeI/KpnI) in pGEM7Z(-) |
| **pAR-his-shBle** | Ble$^{R}$ | AR:his:Ble2A:RBCS2 | 6xhis introduced into pARshBle2A by overlap PCR |
| **pARshBle2AhEGF** | Ble$^{R}$ | AR:Ble2A:3HA:EGF:RBCS2 | Synthetic 3HA (PstI/XhoI) and synthetic EGF (XhoI/NdeI) introduced into pARshBle2A. |
| **pARshBle2ATRXH** | Ble$^{R}$ | AR:Ble2A:3HA:crTRX:RBCS2 | Synthetic 3HA (PstI/XhoI) and amplified TRXH (F222/R223, XhoI/NdeI) introduced into pARshBle2A. |
| **pARshBle2AmCherry** | Ble$^{R}$ | AR:Ble2A-mCherry:RBCS2 | mCherry from CD3-960 cloned into pARshBle2A (PstI/XhoI) |
| **pARshBle2AEhEGF** | Ble$^{R}$ | AR:Ble2AE:3XHA:EGF:RBCS2 | Synthetic Ble:FMDV2Aextended replacing Ble:FMDV2A into pARshBle2AhEGF (EcoRI/PstI) |
| **pAR-ARSss-glucTEVhEGF** | - | AR:ARSss:gLuc:3HA:TEV:EGF | Synthetic ARSssgluc3HATEV replacing shBle:2A into pARshBle2ATRXH (EcoRI/XhoI) |
| **pARshBle2AETRXH** | Ble$^{R}$ | AR:Ble2AE:3XHA:TRXH:RBCS2 | TRXH:3'UTR from pARshBle2ATRXH replacing EGF:3'UTR into pARshBle2AEhEGF (BamHI/BamHI) |

| Vector | Select. gene | Content | Cloning |
|---|---|---|---|
| **pAR-ARSss-glucTEVhEGF-HygB$^R$** | *AphVII’* | 5'RPL23:ARSss:gLuc:3HA:TEV:EGF:3'RBCS2 + AphVII’ (hygB) | Aph7' gene from Phyg3 (KpnI/EcoRV) into pAR-ARSss-glucTEVhEGF (KpnI/SmaI) |
| **pPL-gLuc:hEGFHis:RBCS2** | - | 5'RPL23:ARSss:gLuc:3HA:TEV:EGF:3'RPL23 | PRPL23 5'UTR amplified (F409/410) used to replace AR chimeric promoter in pAR-ARSss-glucTEVhEGF (EcoRI/XbaI) |
| **pPL-gLuc:EGF-PL** | - | 5'RPL23:ARSss:TEV:EGF:3'RPL23 | 3' region of RPL23 amplified (F411/412) replacing 3' RBCS2 in pRPL23:gLuc:hEGFHis:RBCS2 (KpnI/NdeI) |
| **pPL-His-gLuchEGF-PL** | - | 5'RPL23:ARSss:HisLUCEGF:3'RPL23 | 6XHis:gLuc amplified (F433/R434) replacing gLuc:TEV in pRPL23:gLuc:hEGF:RPL23 (BglII/XhoI) |
| **pPLshBleTEVhEGF_no3HA** | Ble$^R$ | 5'RPL23:his:shBle:EGF:3'RPL23 | ShBle amplified (F457/R458) from pARshBle2A replacing luc:3HA in pPL-His:gLuchEGF-PL (EcoRI/XhoI) |
| **pPLshBleTEVhEGF** | Ble$^R$ | 5'RPL23:his:Ble:3HA:TEV:EGF:3'RPL23 | Synthetic 3HATEV (codon adapted) introduced in pPLshBleTEVhEGF_no3HA (XhoI) |
| **pPL-ARSss-glucTEVhEGF-hygB$^R$** | *AphVII’* | 5'RPL23:ARSss:his:gLuc:3HA:TEV:EGF:3'RPL23 + AphVII’ (hygB) | *aph VII* gene from phyg3 (KpnI/EcoRV) introduced in pPL-His:gLuchEGF-PL (SmaI/KpnI) |
| **pPL-ARSss-glucTEVhEGF-paro$^R$** | *AphVIII'* | 5'RPL23:ARSss:his:gLuc:3HA:TEV:EGF:3'RPL23 + aph VIII (paro) | *aph VIII* gene from pSI103 introduced in pPL-His:gLuchEGF-PL (KpnI/SmaI) |
| **pPL-ARSss-glucTEVhEGF-(SP)$_{10}$-hygB$^R$** | *AphVII’* | 5'RPL23:ARSss:his:gLuc:3HA:TEV:EGF:(SP)$_{10}$:3'RPL23 + AphVII’ (hygB) | (SP)$_{10}$ (F485/R486-annealing+phosp.) introduced in pPL-ARSss-glucTEVhEGF- hygB$^R$ (NdeI) |

| Vector | Select. gene | Content | Cloning |
|---|---|---|---|
| **pPL-ARSss-glucTEVhEGF-(SP)$_{20}$-hygB$^{R}$** | *AphVII’* | 5'RPL23:ARSss:his:gLuc:3HA:TEV:EGF:(SP)$_{20}$:3'RPL23 + AphVII’ (hygB) | synthetic (SP)$_{20}$ replacing (SP)$_{10}$ in pPL-ARSss-glucTEVhEGF-(SP)$_{10}$- hygB$^{R}$ (NdeI/PstI) |
| **pPL-GAMss-glucTEVhEGF-paro$^{R}$** | *AphVIII'* | 5'RPL23:GAMss:his:gLuc:3HA:TEV:EGF:3'RPL23 + AphVIII | GAMss (amplified F487/R488 from synthesis product) replacing ARSss in pPL-ARSss-glucTEVhEGF-paro$^{R}$ (EcoRI/BglII) |
| **pPL-ARSss-glucTEVhEGF-LCL-hygB$^{R}$** | *AphVII’* | 5'RPL23:ARSss:his:gLuc:3HA:TEV:EGF:LCL:3'RPL23 + AphVII’ (hygB) | LCL (F506/R507-annealing+phosp.) replacing (SP)$_{10}$ in pPL-ARSss-glucTEVhEGF-(SP)$_{10}$- hygB$^{R}$ (NdeI/PstI) |
| **pPL-ARSss-glucTEVhEGF-GP1-hygB$^{R}$** | *AphVII’* | 5'RPL23:ARSss:his:gLuc:3HA:TEV:EGF:GP1:3'RPL23 + AphVII’ (hygB) | GP1 (F504/R505-annealing+phosp.) replacing (SP)$_{10}$ in pPL-ARSss-glucTEVhEGF-(SP)$_{10}$- hygB$^{R}$ (NdeI/PstI) |
| **pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF** | - | pRPL23 5':ARSss:his:gLuc-intron-gLuc:(SP)$_{20}$:TEV:hEGF:3'RPL23 | synthetic ARSss:his:gLuc-intron-gLuc:(SP)$_{20}$:TEV replacing ARSss:his:gluc:TEV into pPL-His-gLuc-hEGF-PL (EcoRI/XhoI) |
| **pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF-paro$^{R}$** | *AphVIII'* | ARSss:his:gLuc-intron-gLuc:(SP)$_{20}$:TEV:EGF+ Aph VIII (paro) | *Aph VIII* gene from psi103 into pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF (KpnI/SmaI) |
| **pPL-ARSss-glucTEVhEGF-GP2-hygB$^{R}$** | *AphVII’* | 5'RPL23:ARSss:his:gLuc:3HA:TEV:EGF:GP2:3'RPL23 + AphVII’ (hygB) | Synthetic GP2 replacing (SP)$_{10}$ in pPL-ARSss-glucTEVhEGF-(SP)$_{10}$-hygB$^{R}$ (NdeI/PstI) |
| **pPL-ARSss-glucTEVhEGF-PHC21A-hygB$^{R}$** | *AphVII’* | 5'RPL23:ARSss:his:gLuc:3HA:TEV:EGF:PHC21A:3'RPL23 + AphVII’ (hygB) | Synthetic PHC21A replacing (SP)$_{10}$ in pPL-ARSss-glucTEVhEGF-(SP)$_{10}$- hygB$^{R}$ (NdeI/PstI) |
| **pPL-ARSss-glucTEVhEGF-PHC21B-hygB$^{R}$** | *AphVII’* | 5'RPL23:ARSss:his:gLuc:3HA:TEV:EGF:PHC21B:3'RPL23 + AphVII’ (hygB) | Synthetic PHC21B replacing (SP)$_{10}$ in pPL-ARSss-glucTEVhEGF-(SP)$_{10}$- hygB$^{R}$ (NdeI/PstI) |

| Vector | Select. gene | Content | Cloning |
|---|---|---|---|
| **pPL-CAHss-glucTEVhEGF-paroR** | *AphVIII'* | 5'RPL23:CAHss:his:gLuc:3HA:TEV:EGF:3'RPL23 + aph VIII (paro) | CAHss (F533/R534-annealing+phosp.) replacing GAMss in pPL-GAMss-glucTEVhEGF-paroR (EcoRI/BglII) |

In several cases DNA sequences were synthesized as two complementary oligos to be annealed. That was the case of a synthetic sequence $(SP)_{10}$ obtained from two codon optimized and complementary oligonucleotides (F485/R486). Extra bases at the 5’ and 3’ end of the double-stranded fragment were added to generate overhangs compatible with NdeI digested vector. Moreover, a Pst I site was also included at the 3’ end (immediately 5’ of Nde I site) of the double stranded fragment to easily substitute this fragment (Nde I / Pst I) considering that the Nde I site at the 3’ end was not restored with the cloning into the vector. The oligonucleotides were resuspended in mQ water (final concentration of 10 µM) and phosphorylated 37ºC 20 min (10 units PNK from NEB, reaction buffer 250 mM Tris pH 7.6, 50 mM $MgCl_2$, 5 mM ATP, 5 mM DTT, 25% (w/v) polyethylene glycol-8000 – 2.5 µM of each oligo) and afterwards, PNK was inactivated at 75ºC 10 min. Then, to obtain the double stranded DNA fragment, the oligos were mixed in equal molar amounts in annealing buffer (10 mM Tris pH 8.0, 1 mM EDTA, 50mM NaCl), incubated at 94°C for 5 minutes and allowed to cool down at room temperature.

Sequences LCL and GP1 (LCL-F506/R507, GP1-F504/R505) were also synthesised as oligos, phosphorylated and annealed as indicated to replace $(SP)_{10}$ (Nde I / Pst I) in pPL-ARSss-glucTEVhEGF-$(SP)_{10}$-hygB$^{R}$.

### *Chlamydomonas matings*

Chlamydomonas mating type plus (MT+) and mating type minus (MT-) gametes were generated by incubating cells (previously grown for 4-7 days in TAP agar plates) for at least 4 hours in nitrogen-free liquid HSM. Afterwards, MT+ and MT- gametes were mixed for 2-4 hours and mating reactions were followed by microscopy until zygotes began to form. Then the cells were plated on HSM agar plates, incubated under 150 µE for 16 hours and then placed in the dark for at least 5 days to allow maturation of zygotes. Unmated cells were killed through exposure to chloroform vapours for 45 seconds and plates were placed under continuous light until colonies were visible (between five to seven days). Once grown, part of the biomass was resuspended in TAP-Tween 20 0.005% and a series of dilutions were plated in TAP agar plates. Once single colonies were visible, those were picked to a 96 well plate each well containing 200 µl of TAP medium. Generally, colonies were grown for five days to one week prior to assay or replicate plating. Progeny was further analysed appropriately in each case.

To obtain diploids, zygotes were not allowed to mature and were plated directly on the proper antibiotics that would allow to select only fused cells carrying both parental antibiotic resistances. Diploid confirmation was done by genotype with specific MT+ and MT- genes: minus-specific *MID* gene (minus dominance) or the plus specific *FUS1* gene (fusion) (Werner and Mergenhagen 1998).

### *Chlamydomonas nuclear transformation by glass beads*

Cell wall less strains (cc1883, UVM4 and UVM11) were transformed by agitation in the presence of glass beads. Between 1.5-3 µg of linearized vector DNA (Sca I, Bioron) were used in each transformation performed as previously described (Kindle 1990). Transformations were plated on TAP agar with the appropriate antibiotic (25 µg/L paromomycin, 30 and 60 µg/L hygromycin B or 10 µg/L zeocin). Plates were placed under continuous light until antibiotic resistant colonies were clearly visible (between five to seven days). Transformants were picked from plate and transferred to a 96 well plate each well containing 200 µl of TAP medium. Generally, colonies were grown for three to five days prior to assay or replicate plating. Replicate plate generation was performed with Replica plater (R2383, Sigma-Aldrich). The transformed DNA were either the linearized vectors (single restriction enzyme that cut only once).

### *Chlamydomonas nuclear transformation by electroporation*

Cell walled strains (CC-124, CC-125 and 21gr) were transformed by electroporation. The protocol followed in this study is based on the one described (Brown et al. 1991) including some changes. Cells were grown to 1–2 $10^6$ cells/ml in TAP medium 25 ºC under constant illumination of 150 µE on a rotary shaker (200 rpm). Cells were harvested in the presence of 0.005% tween 20 by centrifugation (3000 g 5 min) and washed with 10 ml of ice-cold MAX Efficiency® Transformation Reagent for Algae (A24229, Invitrogen) for each 500 ml of culture. Then, cells were resuspended in ice-cold MAX Efficiency® Transformation Reagent for Algae at a final concentration of $2.5x10^8$ cells/ml. 250 µl of cells were incubated with 0.8-2.5 µg of DNA (linearized vectors) for 5 minutes on ice in a 4 mm cuvette (BioRad). An exponential electric pulse of 1000 V/cm was applied to the sample using a MicroPulser™ Electroporator (BioRad). Cells were recovered for approximately 20 hours in 10 ml of TAP 40 mM sucrose without direct light and gently shaking. After the recovery, cells were plated to two TAP-agar plates supplemented with the appropriate antibiotic (25 µg/L paromomycin, 30 and 60 µg/L hygromycin B or 10 µg/L zeocin). Resulting colonies were treated as described above.

### *Zeocin sensitivity tests*

In order to establish the antibiotic resistance level of different clones, liquid cultures grown in 96 well plates were tested for growth on TAP agar plates with increasing concentrations of zeocin (indicated in each figure). These zeocin sensitivity tests were performed by replicating 96 well plates (comparable grown colonies) with Replica plater (R2383, Sigma-Aldrich).

### *UV mutagenesis*

First, two strains with low expression of selectable marker were obtained: CC-124 transfomants with low expression of shBle were selected by its ability to grow at 25 µg/mL of Zeocin but not on plates with 100 µg/mL of Zeocin. These low-Zeocin-resistant transformants selected (shBlelowA and shBlelowB) were used as background strains for UV mutagenesis. Both strains were grown in liquid TAP until exponential growth phase (1-2 × $10^6$ cells/mL). Twenty milliliters of culture were transferred onto a 14 cm diameter Petri dish (168381, Thermo Fisher Scientific) and exposed to UV light using a Stratalinker® UV Crosslinker 1800 (Stratagene). After exposing cells to the desired quantity of irradiation, cells were immediately covered to be protected from light. Irradiated cells and non-irradiated cells were incubated overnight in the dark to minimize the amount of light-activated repair. Cells were then harvested (3000 gs 5 min) and resuspended in 750 µl of TAP. Resuspended cells were spread on a TAP agar plate supplemented with 100 µg/mL of Zeocin to select mutants with increased levels of shBle expression.

### *Genotype analysis*

DNA isolation was made following Scott Newman's method (Newman et al. 1990) subjected to some changes.

Cells were grown on TAP agar plates for 3-5 days. Cells were scraped off plate into 1.5ml Eppendorf tube containing 50 µl TEN buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA, 150 mM NaCl). After resuspending the cells by vigorous vortexing, these were incubated 95ºC 5min. 100 µl of water and one microliter of RNase (R6148 – 20 mg/ml, Sigma-Aldrich) were added and allowed to react for 10 min at RT. Subsequently, 300 µl of SDS-EB buffer (2% SDS, 400 mM NaCl, 40 mM EDTA, 100 mM Tris-HCl, pH 8.0) were added and vortexed to mix. DNA was extracted once with 350 µl of phenol/CIA (1:1) for one minute by vortexing and phases were separated by centrifugation (14000 gs 5 min):

aqueous phase was transferred to a new tube. The extraction was repeated with 300 µl CIA (24:1) and the aqueous phase transferred again to a new tube. Two volumes of absolute ethanol were added, and DNA was allowed to precipitate ON at -80ºC. Next morning the tubes were centrifuged 14000 g 20 min at 4ºC and ethanol was eliminated carefully. A wash of the pellet was made again with two volumes of original aqueous phase of 70% ethanol: centrifugation 14000 gs 5 min 4ºC and ethanol eliminated very carefully. The genomic DNA pellet was allowed to dry for 10 minutes and resuspended in 50 µl of milli-Q water by pipetting and through a final incubation of 65ºC 20 min.

The integration of the pPLshBleTEVhEGF heterologous cassette was determined by PCR analysis of this genomic DNA. PCR was developed under standard conditions for 35 cycles using RedTaq polymerase (Sigma). The oligonucleotides used are listed in table 13.

**Table 13. Sequences of oligonucleotides used for PCR amplification in this study.**

| Oligonucleotides names | Oligonucleotides sequence (5’-3’) |
|---|---|
| **F475** - BleF1 | ACATGGCCAAGCTGACCAG |
| **R476** - BleR2 | CCCGGAAGTTCGTGGACAC |
| **F477** - BleF2 | ACGACGTGACCCTGTTCATC |
| **R478** - TEVR | TGGAAGTACAGGTTCTCGCG |
| **F479** - BleF3 | GCAACTGCGTGCACTTCG |
| **R480** - EGFRev | CACTTCAGGTCGCGGTACTG |
| **R448** - UTR-1rev | CATGTCAGAGGCCCGAAAAC |
| **R481** - UTR-2rev | TACTCCTTTCGTGCTGGCAG |
| **R428** - UTR-3rev | CTCCCGGTTCCTCGCTATTC |
| **R482** - HA2 | TACGGGTAGCCAGCGTAGT |

| Oligonucleotides names | Oligonucleotides sequence (5’-3’) |
|---|---|
| **R483** - HA1 | GCACGTCGTAGGGGTAGC |
| **R484** - endEGF | ACATATGGTAGCGCAGCTCC |
| **R497** – Adenosylhomocysteinase | AAGATGAAGCGGTCCACCTG |
| **R490** – predicted | TCCAGGGCCTGCTTGATTTG |

Mating type determination was made by genotyping specific MT+ and MT- genes: minus-specific *MID* gene (minus dominance) or the plus specific *FUS1* gene (fusion) (Werner and Mergenhagen 1998).

**Table 14. Sequences of oligonucleotides used for mating type determination by PCR amplification in this study.**

| Oligonucleotides names | Oligonucleotides sequence (5’-3’) |
|---|---|
| **F338 –** *FUS1* gene (MT plus) | GAGAACCTGTACTTCCAGAGC |
| **R339 –** *FUS1* gene (MT plus) | GAGAACCTGTACTTCCAGAGC |
| **F341 –** *MID* gene (MT minus) | CGACGACTTGGCATCGACAGGTGG |
| **R340 –** *MID* gene (MT minus) | CTCGGCCAGAACCTTTCATAGGGTG |

### *Protein isolation and Immunoblot analyses*

Unless otherwise stated, 50 ml cultures were harvested at mid-log phase of grwoth by adding Tween 20 to 0.005% and centrifuging 3000 g 5 min. Pelleted cells were then resuspended in 200-300 µl of PBS or PBS triton 1% (indicated in each figure) in presence of 1X Protease inhibitors (P9599, Sigma) and 10 µM MG132. Immediately, resuspended cells were frozen in liquid nitrogen and maintained at -80ºC until analysed. To proceed to analyse the samples, those were rapidly thawed (through shaking with hands and vortexing) and centrifuged 12000 g 10 min. The resulting supernatant is considered the soluble fraction throughout the current study. If culture supernatant was the fraction to be analysed, cells were separated as indicated and supernatant concentrated by centrifugation at 12000 gs through a membrane of 10 kDa (Vivaspin 500, 10 kDa MWCO PES, VS0101, Sartorius).

Samples were quantified by Bradford method (B6916, Sigma) and denatured at 95ºC for 5 min in presence of 1x laemmli sample buffer (60mM tris ph 6.8, 2% SDS, 10% glycerol, 200mM DTT, bromophenol blue).

Samples were loaded into denaturing SDS-PAGE (Any kD™ Mini-PROTEAN® TGX™ Precast Protein Gels, BioRad) and ran at 200 V for 30 min. Transference to nitrocellulose membranes (1704159-Trans-Blot® Turbo™ Midi Nitrocellulose Transfer Packs, BioRad) was made with Trans-Blot® Turbo™ Transfer System (BioRad) at 1.3 A, 25 V, 7 min. Blots were blocked in PBS containing 5% non-fat dry milk for 1h at room temperature (RT) and incubated with primary antibodies diluted in PBX (1x PBS pH 7.4, 0.1% Tween-20) with 3% dry milk at 4°C overnight. Antibody dilutions were as follows: rat-anti-HA (1:2000) (3F10, Roche, Switzerland), polyclonal rabbit serum anti shBle protein (1:1000) (cat # ant-sh, Invivogen), polyclonal rabbit anti-gLuciferase antibody (1:1000) (NEB) or a rabbit polyclonal anti-hEGF (1:500) (ab9695, Abcam). After washing in PBX for three times 10 min, the blot was incubated with horseradish peroxidase (HRP) conjugated goat-anti-rat-IgG (1:2500) (Thermo Fisher Scientific) or goat-anti-rabbit-IgG (1:40000) (Thermo Fisher Scientific) for 1h at RT, washed as described above, and processed for chemi-luminescence detection (WBLUF0100-Luminata Forte, Merck) using autoradiographic film (Super RX, Fujifilm). The exception is figure 45 that was detected using Odyssey Fc (LiCor) instead of autoradiographic film.

### *Densitometry*

Densitometry was performed using the Fiji Software excepting the case of figure 46/table 9 that was estimated with Image Studio Software Lite

Software. A dilution series of shBle:3HA or gLuc:3HA was loaded as the samples. After separation of known quantities of ShBle:3HA protein, a calibration curve was generated, and the quantity equivalents of samples were determined. Quantification is corrected by comparison of the sizes between standard and target RP. We have assumed that transfer onto the membrane was uniform and that the affinity of the primary antibody against HA tag was equal for samples and standards since they share the same epitope.

### *gLuciferase activity assay*

Bioluminescence assays were performed at RT with a 96-well microplate luminometer (Berthold LB 96V MicroLumat Plus Luminometer, Berthold technologies) with automated substrate injection. Unless otherwise stated, 100 µl of algal culture were mixed with 100 µl of a coelenterazine buffer (100 mM Tris, 500 mM NaCl, 10 mM EDTA, 10 µM coelenterazine) directly before measurement. Coelenterazine was always added to the buffer 30 min before the assay from a 2mM stock (in methanol). Light emission was recorded for 1.5 s with a delay of 0.2s.

Screenings were performed with culture from colonies grown in 96-well plates. Kinetic assays (figure 27, 36, 44, 50 and 51) were performed with culture extract or culture supernatant from cultures grown in Erlenmeyer flasks.

### *Immunoprecipitation - Pierce Crosslink IP kit, anti HA-Roche clone F10*

1L cultures were grown to 1-3 x $10^6$ cells/ml and harvested by adding Tween 20 to 0.005% and centrifuging 4000 g 10 min. Supernatant was discarded and pelleted cells were resuspended in 2 ml of Extraction Buffer A (25 mM sodium phosphate, 300 mM sodium chloride, 10 % glycerol, 1% Triton, ph 7.4, 1X plant protease inhibitors (P9599, Sigma), 10 µM MG132, 1 mM phenantroline).

Immediately, resuspended cells were frozen in liquid nitrogen and maintained at -80ºC until analysed. To proceed to analyse the samples, those were rapidly thawed at room temperature and centrifuged 12000 g 10 min: the resulting supernatant is considered SN1. 1ml of extraction buffer A was added to the resulting pellet and the process was repeated: pellet was resuspended, frozen in liquid nitrogen, thawed at room temperature and centrifuged 12000 g 10 min to collect the supernatant considered SN2. SN1 and SN2 were mixed and considered from now on as “input”. This “input” was diluted with extraction buffer without triton x-100 1:1 to proceed with immunoprecipitation (final 0.5% triton x-100).

The binding of antibody to protein A/G plus agarose (Cat No. 26147 Pierce Crosslink Immunoprecipitation kit - Thermo Fisher Scientific) was made following the manufacturer instructions: 40 µl of the resin slurry (gently mixed to obtain an even suspension) were added to a Pierce spin column and centrifuged 1000 g 1 min. Flow through was discarded. The resin was washed twice with 400 µl of 1X coupling buffer by adding the buffer, centrifuging and discarding the flow through. Then, 6 µg of antibody were prepared for coupling (30 µl Roche HA 3AF10 for each reaction) by adjusting the volume to 200 µl with mQ$H_2O$ and 20X coupling buffer and added to the resin. The coupling reaction was incubated on a rotator at room temperature for 2 h. After the two hours, the column was placed into a collection tube and centrifuged (flow-through was saved to verify antibody coupling). The resin was washed twice: once with 200 µl 1X coupling buffer and once with 600 µl 1X coupling buffer, centrifuged and flow-through discarded.

A preclearing step was made using control agarose resin. For each sample, 80 µl of the control agarose resin slurry were added into a Spin column. The column was centrifuged to remove storage buffer and then was washed with 100 µl of 1X coupling buffer, centrifuged and flow-through discarded. Then the 600 µl of "input" were added and incubated 1h at 4ºC on a rotator. Then, the column was centrifuged 1000 g 1 min: the column containing the resin was discarded and the flow-through (pre-cleared input) was saved.

To proceed with the immunoprecipitation, the precleared lysate (600 µl) was added to the antibody-resin in the column and incubated in a rotator at 4ºC for 16 hours. Then the column was centrifuged and flow-through saved until confirming the result of the IP. The resin was washed three times for 10 minutes with 600 µl of Extraction Buffer A with 0.5% triton X-100 at RT. The elution was made by adding 100 µl of HA peptide (Prod. 26184, Thermo Fisher Scientific, 1 mg/ml in filtered PBS 1X), incubated with shaking-rocking 30 min at RT, centrifuged and collected as first eluate (E1). The elution was repeated three times to obtain E2, E3 and E4.

Finally, E1, E2, E3 and E4 were concentrated with Vivaspin 500 (10 kDa, Z614025, Sartorius) up to 25-30 µl. The result of IP was analysed by loading fractions of each step in a Coomassie stained SDS-PAGE and by immunoblot.

### *Protein sequencing by LC-MSMS*

This work was performed at the Proteomics Platform of Barcelona Science Park, University of Barcelona; a member of ProteoRed-ISCIII network.

The sample was in-gel digested as follows: The gel band was washed with ammonium bicarbonate (25 mM $NH_4HCO_3$) and acetonitrile (ACN). The sample was reduced (DTT 20 mM; 60 min, 60ºC) and alkylated (iodoacetamide 55 mM; 30ºC, 30 min, in the dark). Afterwards, the sample was digested for 16 hours at 37ºC with trypsin (sequence grade modified Trypsin, Promega). Finally, the resulting peptide mixture was extracted from the gel matrix with 10% formic acid (FA) and ACN and dried down in a SpeedVac vacuum system.

The dried-down peptide mixture was analysed in a nanoAcquity liquid chromatographer (Waters) coupled to a LTQ-Orbitrap Velos (Thermo Fisher Scientific) mass spectrometer. The tryptic digest was resuspended in 1% FA solution and an aliquot was injected for chromatographic separation. Peptides were trapped on a Symmetry C18TM trap column (5µm 180µm x 20mm; Waters) and were separated using a C18 reverse phase capillary column (ACQUITY UPLC M-Class Peptide BEH column; 130Å, 1.7µm, 75 µm x250 mm, Waters). The gradient used for the elution of the peptides was 1 to 40 % B in 60 minutes, followed by gradient from 40% to 60% in 10 minutes. (A: 0.1% FA; B: 100% ACN, 0.1% FA), with a 250 nL/min flow rate.

Eluted peptides were subjected to electrospray ionization in an emitter needle (PicoTipTM, New Objective) with an applied voltage of 2000 V. Peptide masses (m/z 300-1700) were analysed in data dependent mode where a full Scan MS was acquired in the Orbitrap with a resolution of 60,000 FWHM at 400 m/z. Up to the 15th most abundant peptides (minimum intensity of 500 counts) were selected from each MS scan and then fragmented in the linear ion trap using CID (38% normalized collision energy) with helium as the collision gas. The scan time settings were: Full MS: 250 ms (1 microscan) and MSn: 120 ms. Generated .raw data files were collected with Thermo Xcalibur (v.2.2).

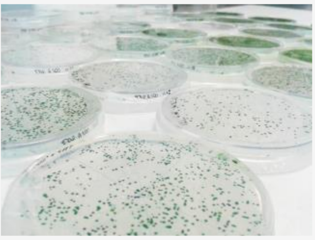
A database was created by merging all protein entries for *Chlamydomonas reinhardtii* present in the public database Uniprot (v. 2/5/17) with a database containing common laboratory contaminant proteins; the following entry was also added to the database:

```
 >GAT_quimera
MAMHMAKLTSAVPVLTARDVAGAVEFWTDRLGFSRDFVEDDFAGVVRDDVTLFISA
VQDQQDQVVPDNTLAWVWVRGLDELYAEWSEVVSTNFRDASGPAMTEIGEQPWG
REFALRDPAGNCVHFVAEEQDSRGYPYDVPDYAGYPYDVPDYAGSYPYDVPDYAGRE
```

NLYFQSRSRNSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELRYHM

The .raw file obtained in the mass spectrometry analysis was used to search against the database described above. The software used was Thermo Proteome Discoverer (v.1.4.1.14) with Sequest HT as the search engine. Both a target and a decoy database were searched to obtain a false discovery rate (FDR), and thus estimate the number of incorrect peptide-spectrum matches that exceed a given threshold. To improve the sensitivity of the database search, Percolator (semi-supervised learning machine) was used in order to discriminate correct from incorrect peptide spectrum matches. The following search parameters were applied: Database/Taxonomy: uniprot_Creinhardtii_170502_cont.fasta Enzyme: Trypsin Max. miscleavage sites 2 Fixed modifications : Carbamidomethyl (C) Variable modifications: Oxidation (M), Peptide tolerance: 10 ppm and 0.6Da (respectively for MS and MS/MS spectra) Percolator Target FDR (Strict): 0.01; Validation based on: q-Value

The search results were visualized in Proteome Discoverer (v.1.4.1.14) and exported to Excel as a list of identified proteins. The results have been filtered so only proteins identified with at least 2 high confidence peptides (FDR> 1%) are included in the list.



# Chapter 2. Purification of human Epidermal Growth Factor produced in Chlamydomonas

# Introduction

Growth factors (GF) represent a large family of proteins that modulate cell responses such as cell adhesion, cell survival, proliferation, chemotaxis and differentiation. GF are used in several applications including regenerative medicine, tissue engineering or cosmetics (Hajimiri et al. 2015; Martino et al. 2015).

Tissue engineering and regenerative medicine combine biomaterials, GF, and stem cells to repair failing organs. The field of tissue engineering is expanding with several products in clinical use and numerous other in clinical studies (Guilak et al. 2014). Since GF are often required to promote tissue regeneration (Tabata 2003), as regenerative medicine and cell therapy projects advance to the clinic, access to a reliable source of stem cells, biomaterial and GF will be needed (Gimble et al. 2007). Thus, difficulties in the supply of these materials may prevent progression of the industry (Karnieli et al. 2017). A related field that also uses GF is the *in vitro* fecundation (IVF) (Chronopoulou and Harper 2015).

*In vitro* cell culture of mammalian lines needs the addition of GF to maintain the desired characteristics and growth. Apart from the addition of specific GF depending on cell lines, traditionally, GF and other supplements are added to *in vitro* cultures of mammalian cells in the form of fetal bovine serum (FBS). However, the use of FBS is problematic because there exists variability in composition between batches, there is an inherent risk of contamination by animal pathogens (Van Der Valk et al. 2004; Schiff 2005) and the collection of serum causes animal suffer (Van Der Valk et al. 2004). Replacing FBS is highly desired as demonstrated by several attempts to achieve it (Lindroos et al. 2009; Tan et al. 2015; Karnieli et al. 2017; Lee et al. 2018) and to replace FBS, recombinant GF will be needed. Although defined media have been successfully used for culturing numerous cell lines their cost is considered to be high (Babcock et al. 2012).

The cosmetic industry is continuously searching new anti-aging products. GFs are synthesized by skin cells and because of its role in the formation of the extracellular matrix (ECM) and control of other GF, they have become important targets of cosmetic applications. Nowadays, several anti-aging products contain recombinant GFs, mostly expressed in *E.coli* (http://www.caregen.com/) although they are also produced in plants (http://plantaderma.es/, https://nbm21.en.ec21.com/).

Current RP expression hosts include bacteria, yeasts, insect cell lines, plants, mammalian cell lines and transgenic animals, each of these production platforms having its limitations (Demain and Vaishnav 2009; Ferrer-Miralles et al. 2009; Corchero et al. 2013). See the general introduction of this thesis for more extensive information about expression platforms (*Introduction, page 1*). The most suitable expression host has to be selected depending on each product (Schmidt 2004). Since GF differ greatly in size and characteristics (PTMs needed, oligomerization, etc.) there are GF that may be expressed in *E.coli* such as hEGF but others require complex PTMs to be fully active and thus are best suited for expression in eukaryotic systems, for example Human Platelet Derived Growth Factor (PDGF) (Deepa et al. 2013). However, in the context of the increasing demand caused by cell therapy projects reaching clinical approval, alternatives to the costly and limited production in mammalian cells are needed. It is important to note that animal-free, endotoxin-free GF are highly desired for stem cell cultures, serum-free applications, chemically-defined media and other *in vitro* pre-clinical work. Transgenic plants as GF production hosts have an advantageous low cost of cultivation, high mass production, flexible scale-up, lack of human pathogens and addition of eukaryotic PTMs (Ferrer-Miralles et al. 2009). There are examples of marketed plant expressed GF used as cell culture media components (www.invitria.com) or in cosmetics (http://plantaderma.es/, www.nbm21.en.ec21.com). Main challenges for plant-derived RPs include low yields, the long process required to generate transgenic plant lines, the associated scale-up costs, PTM patterns differing from mammalian ones, the possibility of gene flow and the negative public perception of transgenic plants (Ferrer-Miralles et al. 2009; Hernández et al. 2014).

Compared to plants, microalgae have some extra benefits: the quick production of RP (only requires weeks from initial transformants to production volumes), the growth in containment, no risk of displacement of cultures, no risk of gene flow and an easier downstream processing in case of secreted RP (Rasala and Mayfield 2015; Rosales-Mendoza 2016). If RP is secreted to medium there is no need of breaking cells, avoiding the release of intracellular content that would hamper the integrity of RPs (due to intracellular proteases) and the downstream processing (due to cell debris including sugars, proteins, DNA, etc.). On the contrary, in the case of plants, cells must be broken to recover RP involving a more complex downstream processing. In the context of the demand for a scalable production of efficient endotoxin-free and animal-free GF, microalgae represent a promising host.

Human Epidermal Growth Factor (hEGF), a common cellular activator, is our model for the GF family. Our objective is to validate Chlamydomonas as a candidate host for the production of GF. The challenge we face in the present chapter is to recover RP from culture medium starting from low concentrations (µg/L), and to purify and process the fusion protein to obtain functional hEGF. The purification method mainly depends on the characteristics of the target protein, the quantities needed and the desired purity grade. For example, high purity levels may be not necessary for cosmetic applications but are mandatory for therapeutic uses. Affinity chromatography is a frequent first purification step when it is possible to tag the target protein as it is our case. If required, additional purification steps are included to remove impurities. For instance, hEGF and other GFs expressed in plants (*Nicotiana benthamiana*) are purified up to 97% of purity through an initial step of affinity chromatography followed by anionic exchange (http://plantaderma.es/). In our case, since the starting RP concentrations are low, it is likely that we will need several purification steps including different methods to achieve purified hEGF. This chapter aims at developing a scalable optimized purification method to recover and purify RP from Chlamydomonas medium, therefore validating it as an industrial host.

# Results

The objective of the present chapter is to validate Chlamydomonas as an expression platform for a functional small and simple protein and to assess the suitability of Chlamydomonas as an industrial RP host. This general goal will be achieved by pursuing specific objectives: to purify hEGF from culture medium and to test its functionality.

## 1. *Purification of different hEGF isoforms*

### 1.1. Development of a purification procedure at lab scale

Fusion proteins comprising several hEGF isoforms have been successfully expressed and secreted to medium with yields from 1 to 100 µg RP/L culture (Chapter 1 - *Strategy 1 - 2.6 Characterization and quantification of transgene expression during cell culture growth* - table 9). The concentration of expressed RP in culture medium hampers its direct purification and a previous concentration step is required. At lab scale this concentration step was solved by lyophilization. After concentration, the purification process designed has several steps as shown in figure 61.

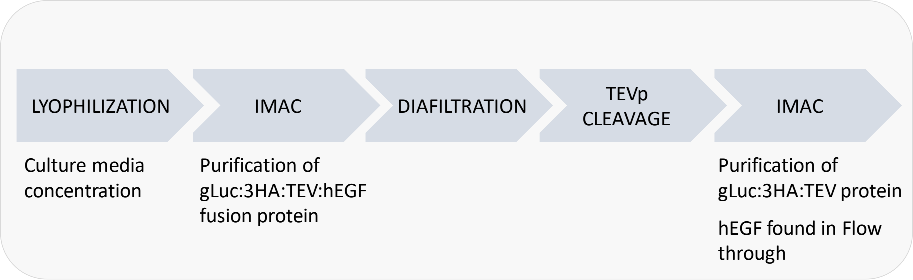

LYOPHILIZATION
IMAC
DIAFILTRATION
TEVp CLEAVAGE
IMAC
Culture media concentration
Purification of gLuc:3HA:TEV:hEGF fusion protein
Purification of gLuc:3HA:TEV protein
hEGF found in Flow through

**Figure 61. Schematic representation of initial purification proposal.**

To optimize IMAC as the second step of the purification proposal three resins were tested as they may show differences on yield of RP purification or purity level achieved: $Ni^{2+}$, $Co^{2+}$ and $Ni^{2+}$-EXCEL. $Ni^{2+}$-EXCEL is a resin optimized for purification of proteins secreted to medium. Since culture supernatant containing RP is concentrated by lyophilization, the concentration of salts may interfere with the purification and due to that IMAC was tested with and without a previous step of dialysis. Transformant CC-124#3 expressing the fusion protein gLuc:3HA:TEV:hEGF (transformed with pPL-GAMss-gLucTEVhEGF vector, see *Strategy 1 - 2.3 Comparison of the effect of several secretion peptides on expression yields* – figure 32) with an expression level previously quantified at 8 µg RP/L (Chapter 1 - *Strategy 1 - 2.6 Characterization and quantification of transgene expression during cell culture growth* - table 9) was selected for the development of the IMAC step. Supernatant from a culture at stationary phase of growth was concentrated 80X by lyophilization and once resuspended was (both directly and after dialysis to the purification buffer) purified either by $Ni^{2+}$, $Co^{2+}$ or $Ni^{2+}$EXCEL resins (for each 500 µl of culture approximately 1 µl of resin slurry was added). Purification efficiency was assessed by immunoblot (figure 62).

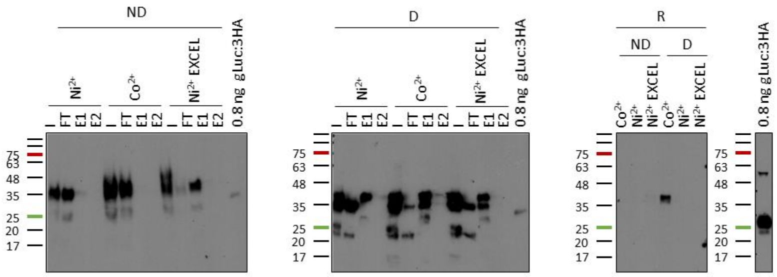


**Figure 62. Immunoblot assessment of IMAC purification with different resins and sample conditions.** 12 µl were loaded in each lane. Primary antibody against gLuciferase. The immunoblot showed at right was exposed during minutes while the other two immunoblots were exposed during seconds. D: dialyzed, ND: non-dialyzed, I: input, FT: flow through, E1: first elution, E2: second elution, R: third elution together with resin.

As it can be seen in the immunoblot, the $Ni^{2+}$EXCEL resin allows an efficient purification avoiding the dialysis step (there is not RP found in flow through) while there is no binding of RP to the other two resins (RP is only found in flow through). Besides, the three resins efficiently purify RP from the dialysed sample (only a small proportion of RP is found in flow through). Also, most purified RP is eluted in the first elution (E1). Furthermore, regarding the purification of dialysed samples, it can be observed that two bands appear at both input and first elution: one at the expected molecular weight (34.2 kDa) and one slightly lower. It can also be observed that only the lower band is detected in the flow through (FT) indicating that only the fusion protein of the expected molecular weight is purified. A hypothesis is that there is degradation or cleavage of the 6xhis tag that causes the loss of RP during purification. Given these results, the best option is to directly purify with $Ni^{2+}$EXCEL resin avoiding the step of dialysis. It will be necessary to determine if the purification of the other fusion proteins show the same performance.

Purified RP by IMAC from transformant UVM4#4 expressing the fusion protein gLuc:3HA:TEV:hEGF:$(SP)_{20}$ (transformed with pPL-ARSss-gLuc-TEV-hEGF-$(SP)_{20}$ vector, see Chapter 1 - *Strategy 1 - 2.4 Use of glycomodules to enhance expression and secretion of nuclear transgenes of Chlamydomonas*– figure 38) with an expression level previously estimated at 29 µg/L (*Strategy 1 - Chapter 1 - Strategy 1 - 2.6 Characterization and quantification of transgene expression during cell culture growth* - table 9) was selected because its higher yields would allow to study minimal TEV protease concentrations needed for cleavage. Several TEV protease and DTT concentrations were tested at different temperatures (results not shown) and the optimal reaction conditions were determined: 3h at room temperature (RT), 0.2 units of TEV protease per 20 µl of reaction without addition of DTT.

### 1.2 Purification and processing of different hEGF isoforms

Selected transformants were grown until stationary phase of growth (OD750nm 2.5) because it was previously seen that RP is accumulated in culture medium until this stage of growth (Chapter 1 - *Strategy 1 - 2.6 Characterization and quantification of transgene expression during cell culture growth* - figure 44). Transformant UVM4#35 was the exception since it was grown for 48 h after reaching stationary phase of growth because it had been shown that gLuc:$(SP)_{20}$:TEV:hEGF accumulates in culture medium without detectable degradation until this culture stage (Chapter 1 - Strategy 1 - *2.6 Characterization and quantification of transgene expression during cell culture growth - figure 44*). 400 ml of culture supernatant from each transformant were purified following the purification proposal presented in figure 61. Immunoblots were performed to assess the results of these purification processes. The same purification was performed for CC-124 and UVM4 strains and no signal was detected by immunoblot (results not shown). As it is shown in figure 63, all the different fusion proteins behave similarly to the isoform previously tested: all were recovered and eluted in first fraction. Also, the different fusion proteins were successfully digested by TEV protease but gLuc:$(SP)_{20}$:TEV:hEGF (expressed by transformant UVM4#35). This fusion protein was successfully recovered from medium but was not cleaved by TEV protease presumably because the TEV protease cleavage site is just next the $(SP)_{20}$ glycomodule and its physical conformation may hamper the cleavage. A silver stained SDS page was used to assess the purification level achieved in fractions submitted to evaluation of hEGF proliferation activity (*2.2 In vitro activity assessment of partially purified hEGF isoforms*) and it showed (figure 63-C) that purified hEGF is a minor part of the protein fraction.

Several TEV protease and DTT concentrations were tested at different temperatures to see if the fusion protein gLuc:$(SP)_{20}$:TEV:hEGF could be cleaved by TEV protease but no cleavage was detected (results not shown).

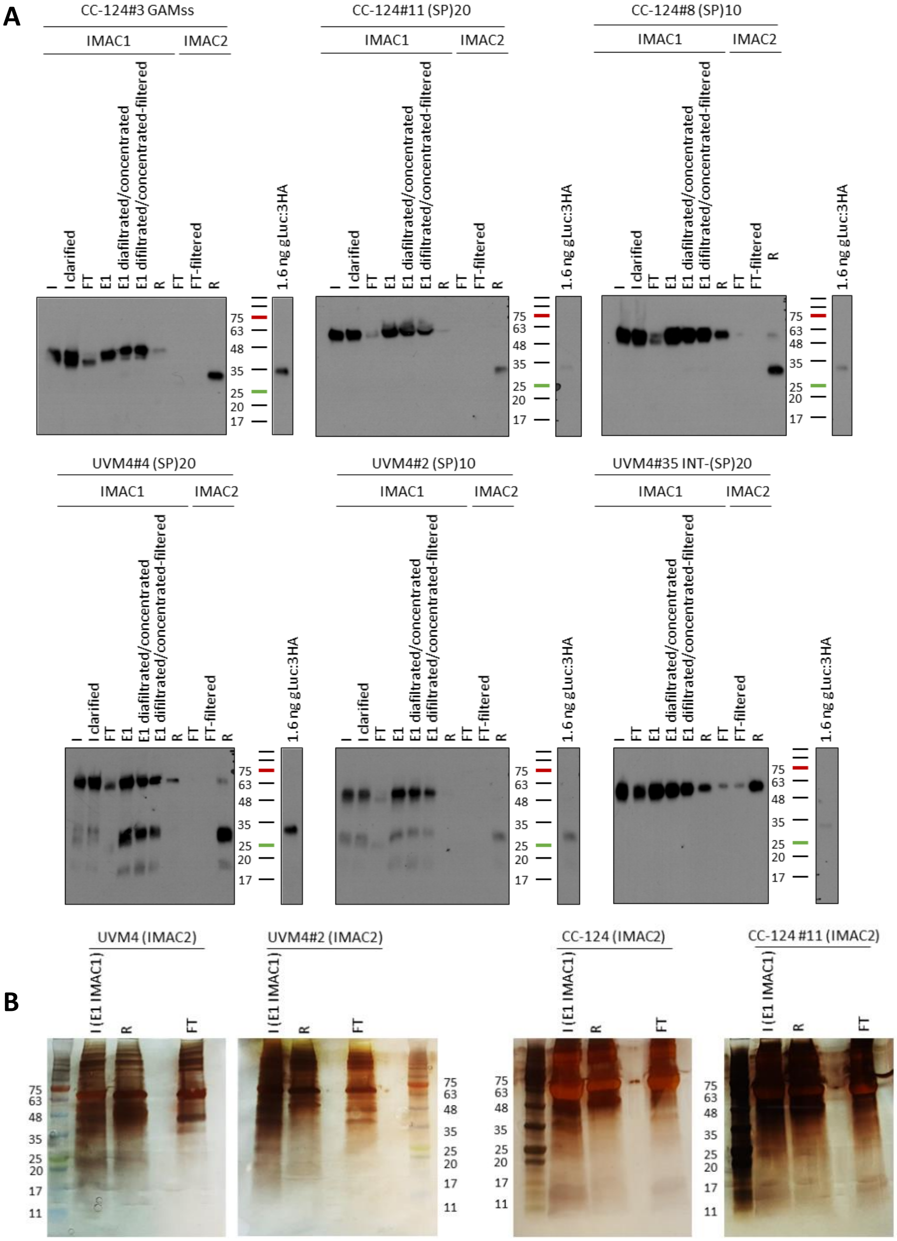

A
CC-124#3 GAMss
IMAC1
IMAC2
I
I clarified
FT
E1
E1 diafiltrated/concentrated
E1 difiltrated/concentrated-filtered
R
FT
FT-filtered
R
1.6 ng gLuc:3HA
75
63
48
35
25
20
17
CC-124#11 (SP)20
CC-124#8 (SP)10
UVM4#4 (SP)20
UVM4#2 (SP)10
UVM4#35 INT-(SP)20
B
UVM4 (IMAC2)
UVM4#2 (IMAC2)
CC-124 (IMAC2)
CC-124 #11 (IMAC2)
I (E1 IMAC1)
R
FT
11

**Figure 63. Analysis of RP purification from selected transformants. A) Immunoblot assessment of RP purification from selected transformants.** Equivalent volumes were loaded in each lane. Primary antibody against gLuciferase. I: input – lyophilized culture supernatant resuspended in ultrapure water, I clarified: input after clarification by centrifugation, FT: flow through, E: elution, E diafiltrated/concentrated: elution after a change of buffer and a concentration of 2.5X. GAMss: pPL-GAMss-glucTEVhEGF-hygB$^R$, (SP)10: pPL-ARSss-glucTEVhEGF-(SP)$_{10}$-hygB$^R$, SP(20): pPL-ARSss-glucTEVhEGF-(SP)$_{20}$-hygB$^R$, INT-(SP)20: pPL-ARSss-gLucINTRON-(SP)$_{20}$-TEV-hEGF. Immunoblot exposures shown were selected to maximize information from each fusion protein purification process and are not directly comparable. **B) Silver stained SDS page of fractions of second IMAC** (after TEV protease cleavage of elution from first IMAC). 20 µl were loaded at each lane.

### 1.3 IEX purification of processed hEGF

As explained, following the proposal presented in figure 61 hEGF represented a minor proportion of the total protein content in the obtained fractions. Thus, the inclusion of a subsequent purification step is needed. To achieve higher purity levels ion exchange chromatography (IEX) was proposed as schematized in figure 64.

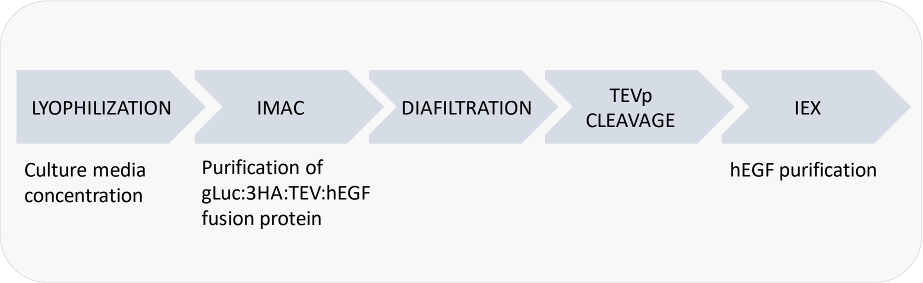

LYOPHILIZATION
IMAC
DIAFILTRATION
TEVp CLEAVAGE
IEX
Culture media concentration
Purification of gLuc:3HA:TEV:hEGF fusion protein
hEGF purification

**Figure 64. Schematic representation of improved purification proposal.**

Transformant CC-124#3 expressing the fusion protein gLuc:3HA:TEV:hEGF was selected to further develop the downstream processing to obtain purified hEGF. This transformant was selected because, with an expression level previously estimated at 8 µg RP/L (Chapter 1 - *Strategy 1 - 2.6 Characterization and quantification of transgene expression during cell culture growth* - table 9), is the highest expressing transformant of a fusion protein which cleavage releases discrete hEGF. RP from 5 L of culture supernatant (from a culture at stationary phase of growth with an OD750nm of 2.25) was purified following the proposal presented in figure 64 (details in methodology). IMAC (figure 65) and IEX (figure 66) results were assessed by immunoblot against gLuciferase and against hEGF.

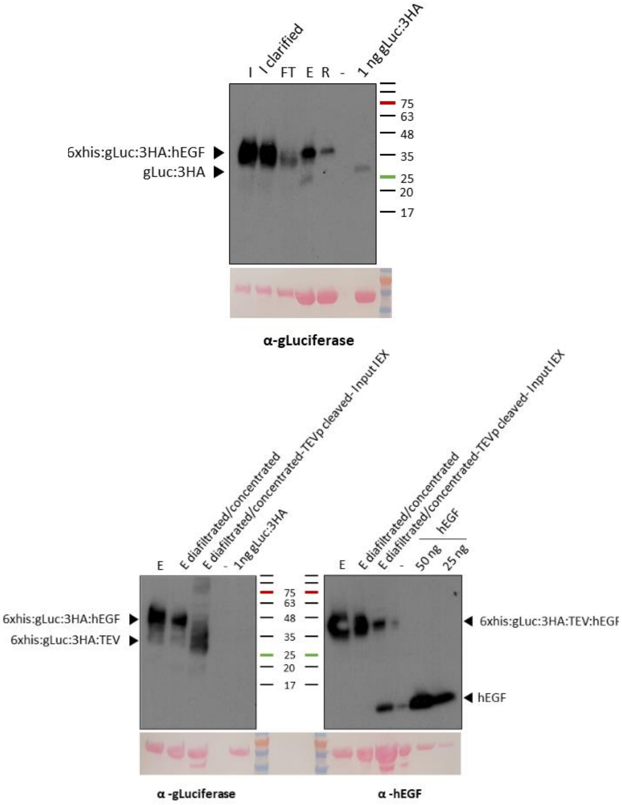
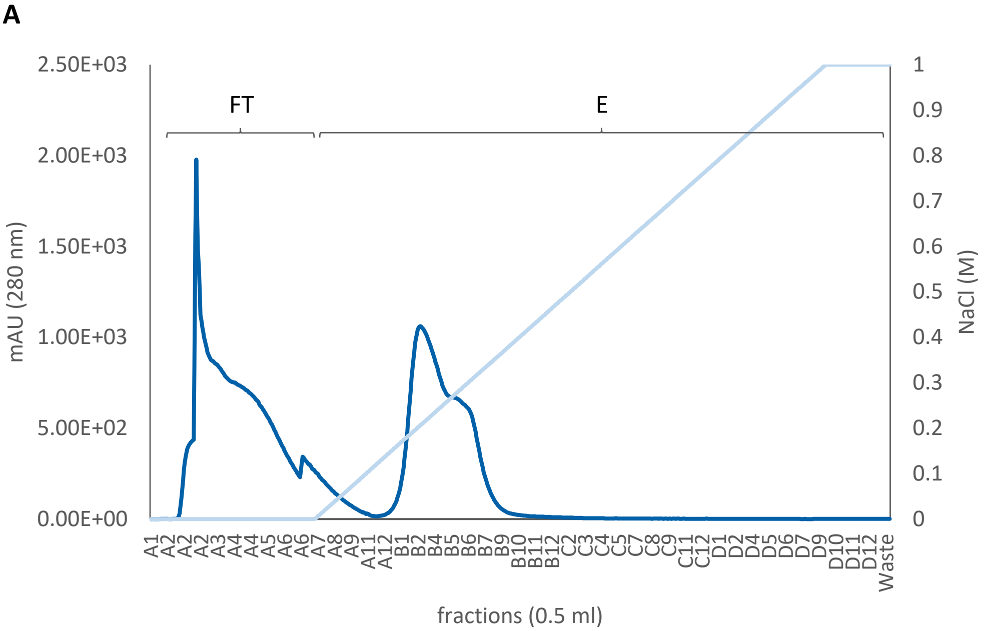
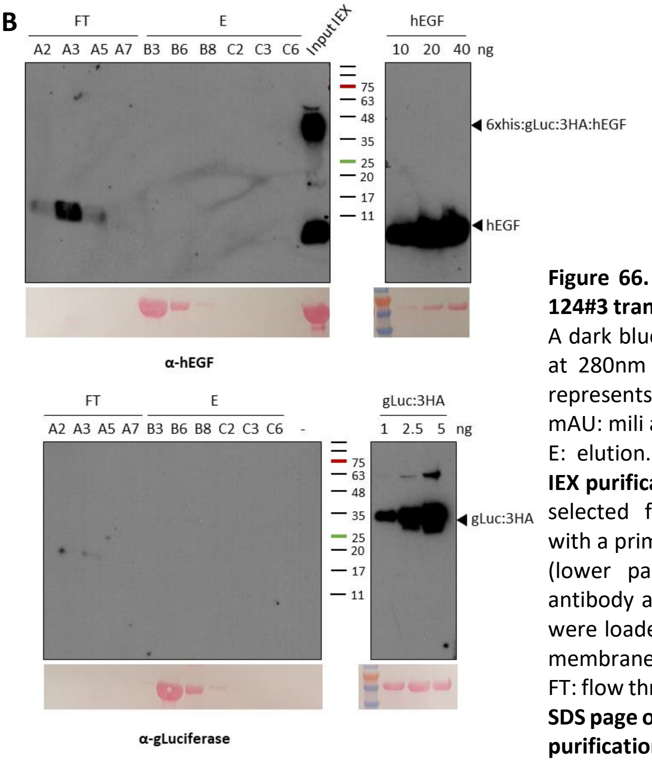

I
I clarified
FT
E
R
-
1 ng gLuc:3HA
75
63
48
35
25
20
17
6xhis:gLuc:3HA:hEGF
gLuc:3HA
α-gLuciferase
E
E diafiltrated/concentrated
E diafiltrated/concentrated-TEVp cleaved- Input IEX
-
1ng gLuc:3HA
6xhis:gLuc:3HA:hEGF
6xhis:gLuc:3HA:TEV
α -gLuciferase
E
E diafiltrated/concentrated
E diafiltrated/concentrated-TEVp cleaved- Input IEX
-
hEGF
50 ng
25 ng
6xhis:gLuc:3HA:TEV:hEGF
hEGF
α -hEGF

**Figure 65. Immunoblot assessment of RP IMAC purification from CC-124#3 transformant.** Equivalent volumes were loaded in each lane. Primary antibody used is indicated below immunoblots. Ponceau stained membranes are shown below immunoblots. I: input – lyophilized culture supernatant resuspended in ultrapure water, I clarified: input after clarification by centrifugation, FT: flow through, E: elution, E diafiltrated/concentrated: elution after a change of buffer and a concentration of 2.7X, E diafiltrated/concentrated-TEVp cleaved: previous fraction subjected to TEV protease digestion.

Once fusion protein had been purified by IMAC and digested by TEV protease, the next step was to purify processed hEGF by IEX. Since the hEGF expressed in the present work has an isoelectrical point (pI) of 5.8 (https://web.expasy.org/compute_pi/) a standard anionic exchanger column purification was tested. The purification with an anionic exchanger column was performed at pH 8, expecting that hEGF would bind to the column and elute at some point of the linear gradient. However, as it is shown in figure 66, hEGF did not bind to the column and is found in the FT fractions. Moreover, both the chromatogram and a silver stained SDS page (figure 67-C) showed that the flow through fraction contains a variety of proteins where hEGF represents a minority of the total protein content. Since pI may be far from the theoretical value, to lower the ionic strength of the buffer or to increase the pH of buffer are proposed to optimize the purification. Alternatively, purification with a cationic exchanger column may be tested.




B
FT
E
A2 A3 A5 A7 B3 B6 B8 C2 C3 C6
Input IEX
hEGF
10 20 40 ng
75
63
48
35
25
20
17
11
6xhis:gLuc:3HA:hEGF
hEGF
α-hEGF
FT
E
A2 A3 A5 A7 B3 B6 B8 C2 C3 C6 -
gLuc:3HA
1 2.5 5 ng
75
63
48
35
25
20
17
11
gLuc:3HA
α-gLuciferase

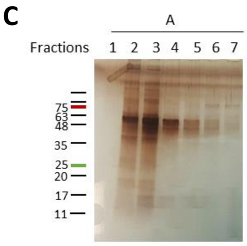

C
A
Fractions
1 2 3 4 5 6 7
75
63
48
35
25
20
17
11

**Figure 66. IEX purification of RP from CC-124#3 transformant. A) IEX chromatogram.** A dark blue line represents the absorbance at 280nm (left axis) and a light blue line represents the molarity of NaCl (right axis). mAU: mili absortion units, FT: flow through, E: elution. **B) Immunoblot assessment of IEX purification of RP.** Two immunoblots of selected fractions were performed: one with a primary antibody against gLuciferase (lower panel) and one with a primary antibody against hEGF (upper panel). 12 µl were loaded at each lane. Ponceau stained membranes are shown below immunoblots. FT: flow through, E: elution. **C) Silver stained SDS page of flow through fractions from IEX purification.** 12 µl were loaded in each lane.

# 2. *hEGF activity assessment of isoforms*

## 2.1 Development of *In vitro* activity assay conditions

To assess the activity of our hEGF isoforms a robust *in vitro* assay was needed. In the assay named “MTT test” the yellow tetrazolium MTT (3-(4, 5-dimethylthiazolyl-2)-2, 5-diphenyltetrazolium bromide) is reduced by metabolically active cells and the resulting intracellular purple formazan can be solubilized and quantified by spectrophotometric means. As MTT test measures the number of cells metabolically active (alive) it does not measure proliferation but viability. However, it is generally accepted as a measure of proliferation (Mosmann 1983).

In order to develop an assay to assess the activity of recombinant hEGF, the induction of proliferation by a commercial hEGF was tested in NIH 3T3 cells in complete medium (DMEM+10%FCS) and in basal medium (DMEM) by MTT test. As can be seen in figure 67, there is not an increase of proliferation induced by hEGF in complete medium. When tested in basal medium a tendency can be observed in NIH3T3 cells: with increasing concentrations of hEGF there is an increase of the number of cells at the end of the assay. However, it is necessary to highlight that NIH3T3 cells mostly died in basal medium and thereby the tendency is more related with an increase of survival than with an increase in proliferation.

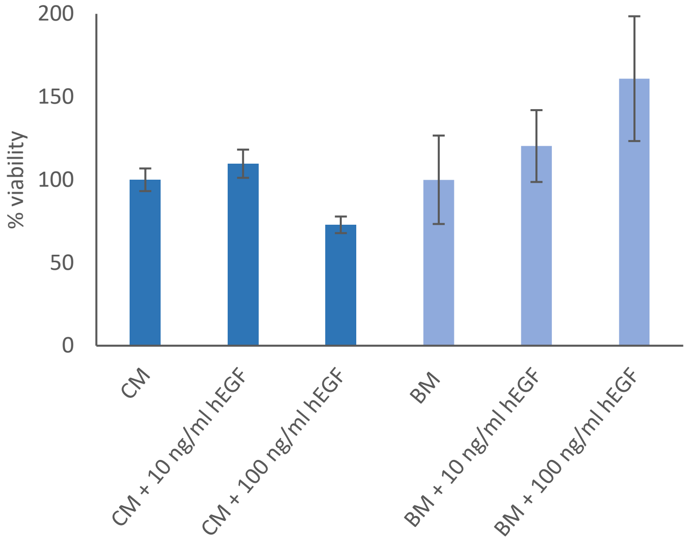


**Figure 67. hEGF proliferation induction in NIH 3T3 cells in complete medium (CM) and basal medium (BM) by MTT test.** Error bars show the standard deviation of three biological replicates. Initial concentration of cells was 3x10$^3$ cells per well.

A subsequent MTT test was performed with NIH 3T3 cells under different medium conditions: basal medium + 0.5% fetal calf serum (FCS), basal medium + 1% FCS, basal medium + 2% FCS.

As can be seen in figure 68, there is an increase in proliferation induced by hEGF under the three conditions tested (0.5%, 1% and 2% FCS). In all three conditions there is a maximum of induction of proliferation at 10 ng/ml. However, it seems that the induction of proliferation has more impact as lower is the FCS percentage and thus the best condition to test proliferation induced by hEGF is in basal medium + 0.5% FCS. Although the tendency of the induction of proliferation can be appreciated, the variability between replicates prevents the assay from being robust.

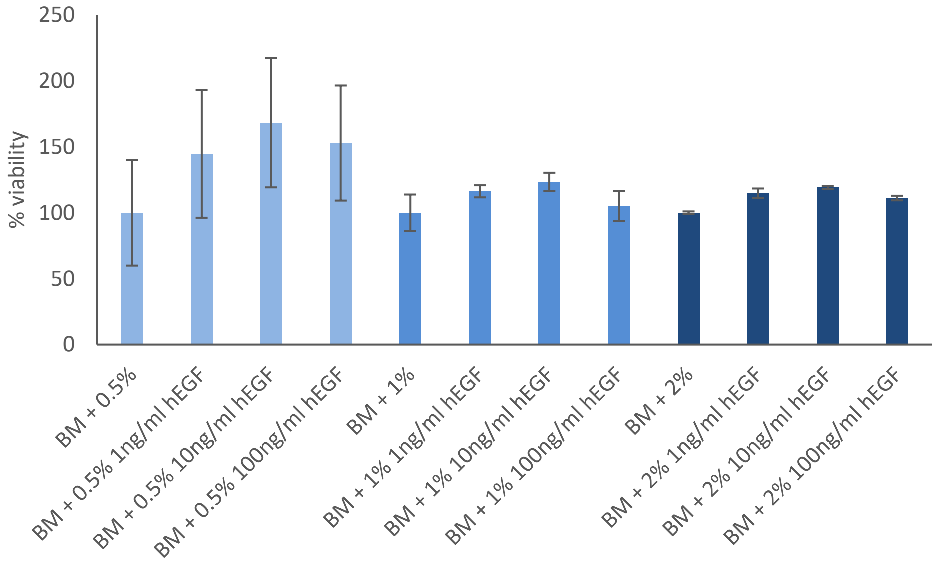


**Figure 68. hEGF proliferation induction in NIH 3T3 cells under different medium by MTT test.** Error bars show the standard deviation of three biological replicates. Initial concentration of cells was 6x10$^3$ cells per well.

With the aim to improve the assay, a subsequent MTT was performed starting with a higher number of cells: higher optical density at the end of the assay was expected to reduce variability between replicates. This was confirmed as seen in figure 69.

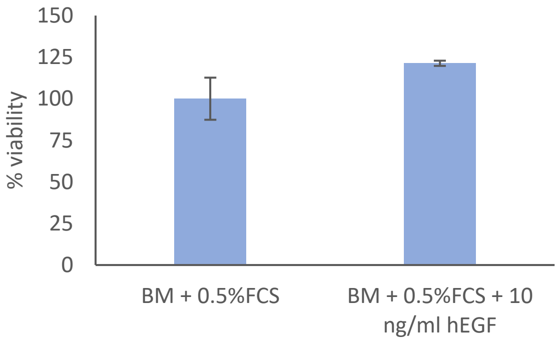


**Figure 69. hEGF proliferation induction assessment in NIH3T3 cells in basal medium (BM) + 0.5% FCS.** Error bars show the standard deviation of six biological replicates. Initial concentration of cells was 1x10$^4$ cells per well.

Summarizing, we established the optimal conditions for assessment of hEGF functionality by induction of proliferation in NIH3T3 cells by MTT assay: basal medium + 0.5% FCS, hEGF between 1 and 100 ng/ml and a starting number of $1x10^4$ cells per well.

### 2.2 *In vitro* activity assessment of partially purified hEGF isoforms

As it can be seen in figure 63 (*1.2 Purification and processing of different hEGF isoforms*), isoforms of hEGF from the selected transformants were successfully recovered from culture supernatant by IMAC and successfully cleaved by TEV protease. The exception was the fusion protein gLuc:$(SP)_{20}$:TEV:hEGF (expressed by transformant UVM4#35) which was not cleaved by TEV protease. Proliferation induction of different processed hEGF isoforms (FT IMAC2) obtained from UVM4 transformants was tested: hEGF:$(SP)_{10}$ (obtained from transformant UVM4#2) and hEGF$(SP)_{20}$ (obtained from transformant UVM4#4). Instead of processed hEGF, the fusion protein gLuc:$(SP)_{20}$:TEV:hEGF purified by IMAC (E1 IMAC1) from transformant UVM4#35 was tested. Both fractions (E1 from IMAC and FT from second IMAC) obtained from UVM4 were tested as a negative control. Results are shown in figure 70.

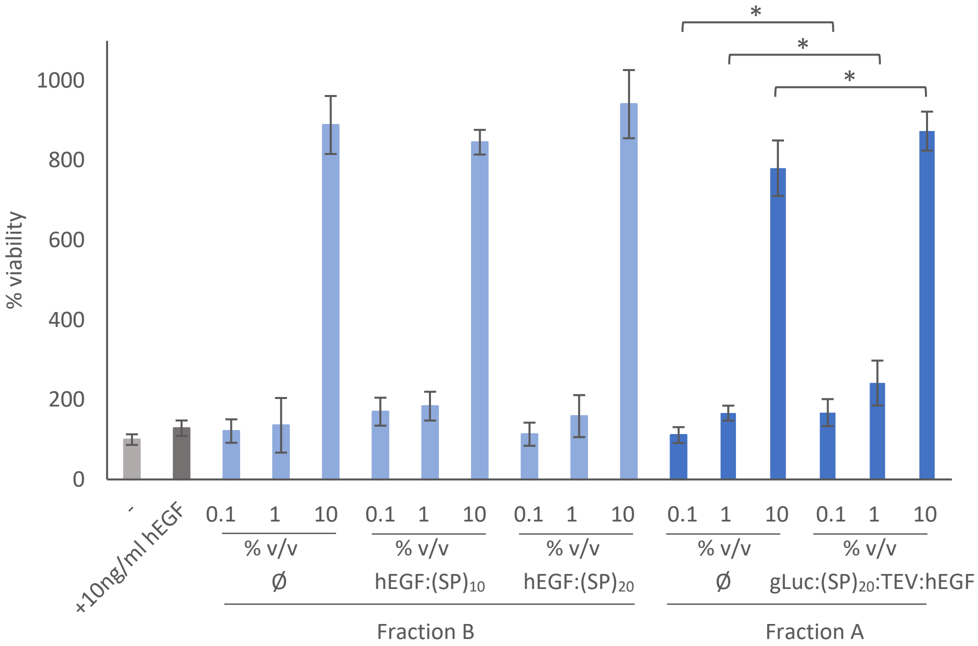
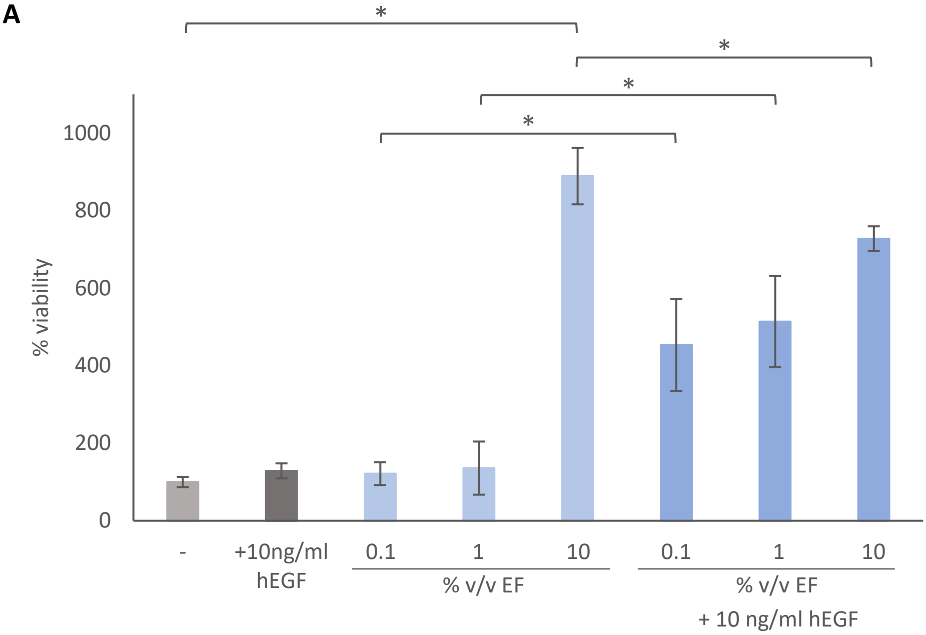


**Figure 70. MTT proliferation induction assessment of hEGF isoforms expressed in Chlamydomonas.** Fraction A – corresponds to E1 of first IMAC (figure 63) and thus contains IMAC enriched full length fusion protein. Fraction B – corresponds to FT of second IMAC (figure 63) and thus contains processed hEGF. Both fractions contain RP 200X relative to initial culture volume. Assay was performed in basal medium containing 0.5 % FCS in NIH3T3 cells. Error bars show the standard deviation of six biological replicates. Initial concentration of cells was 1x$10^4$ cells per well. - ; basal medium containing 0.5 % FCS, * indicates significant increase in viability determined by a two-independent sample t-test ($p < 0.05$). hEGF:$(SP)_{10}$ was obtained from transformant UVM4#2, hEGF:$(SP)_{20}$ was obtained from transformant UVM4#4 and gLuc:$(SP)_{20}$:hEGF was obtained from transformant UVM4#35, Ø makes reference to fractions obtained from UVM4 strain.

Unexpectedly, fractions from non-transformed UVM4 strain induced proliferation. That hampers the analysis of hEGF proliferation activity of partially purified RPs and corroborates the need to further purify RP to test its activity. Besides, the fusion protein gLuc:$(SP)_{20}$:TEV:hEGF shows an statistically significant higher viability when compared to its UVM4 control fraction at the three concentrations tested. That may indicate hEGF activity but that cannot be determined by the present assay because while UVM4 culture was harvested when it arrived at stationary phase, transformant #35 was let to grow for 48h after reaching stationary phase of the culture. Thereby, it is likely that the components responsible of proliferation activity of the wild type fractions are enriched during this extra period.

In a subsequent assay, lyophilized culture supernatant from CC-124 strain (dialysed to PBS) and a UVM4 fraction (FT second IMAC, figure 63) were tested in presence and absence of a commercial hEGF to evaluate if the induction of proliferation had an additive behaviour or if it was masked. As it can be seen in figure 71, both fractions induce proliferation. Whereas the culture supernatant from CC-124 is concentrated 53X relative to initial culture, the partially purified fraction of UVM4 is 200X concentrated relative to initial culture. However, when compared to the negative control (0.5% FCS), the concentrated supernatant showed significant increase in proliferation for the three concentrations tested while the partially purified fraction from UVM4 showed significant increase in proliferation only in the less diluted concentration tested. Thus, concentrated culture supernatant from CC-124 is more effective than the IMAC purified fraction from UVM4. When compared between themselves in presence or absence of hEGF, it is important to highlight that the effect of the fractions and the commercial hEGF seem to be synergistic rather than additive. The reason why the induction of proliferation when hEGF is added to the wild type fraction from UVM4 is more pronounced than when added to the concentrated medium may be an overdosing effect similar to the tendency observed between 0.5%, 1% and 2% of FCS.



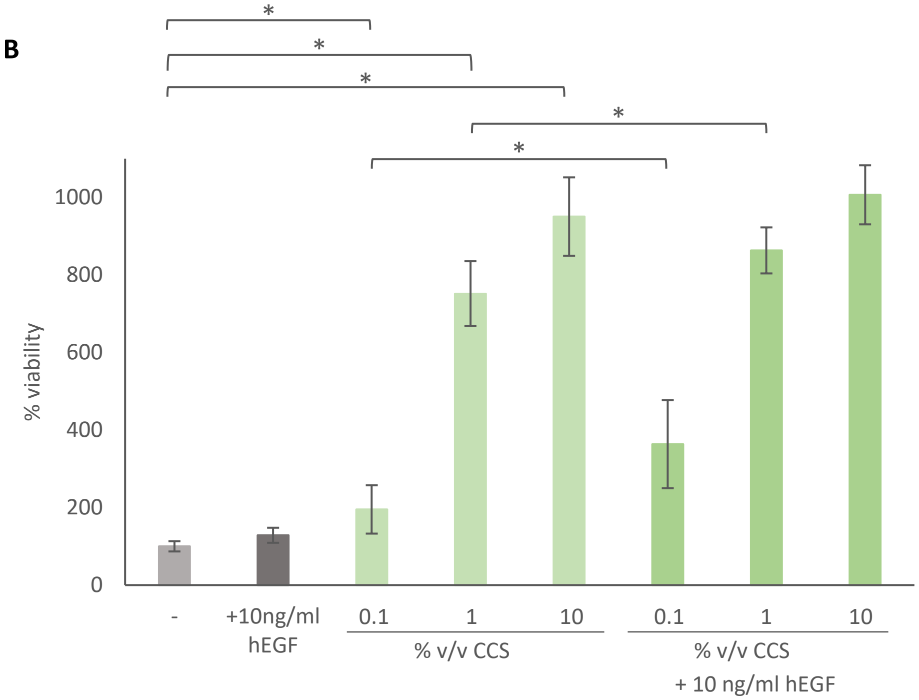


**Figure 71. Proliferation induction assessment of Chlamydomonas extracellular fractions in presence or absence of hEGF by MTT assay. A) Proliferation induction assessment of IMAC enriched fraction (EF). B) Proliferation induction assessment of concentrated culture supernatant (CCS).** All the conditions tested were assayed in basal medium containing 0.5 % FCS in NIH3T3 cells. Error bars show the standard deviation of six biological replicates. Initial concentration of cells was $1x10^4$ cells per well. EF: IMAC enriched fraction 200X relative to initial culture volume, CCS; concentrated culture supernatant 53X relative to initial culture volume, - ; basal medium containing 0.5 % FCS, * indicates a significant increase in viability determined by a two-independent sample t-test ($p < 0.05$).

# Discussion

As reported in Chapter 1, hEGF has been successfully expressed as a fusion protein to gLuciferase and a combination of stabilizing regions. The challenge we face in the present chapter is the purification and processing of the different fusion proteins and testing the activity of varied Chlamydomonas expressed hEGF isoforms. A scalable protocol based on culture supernatant concentration, IMAC purification and TEV protease processing has been developed and validated. Unexpectedly, over the course of these experiments it was found that Chlamydomonas secreted components induce proliferation of murine fibroblasts. The reported proliferation induced by Chlamydomonas extracellular fractions as well as the synergistic effect seen when tested together with hEGF, unveils the potential of secreted components of Chlamydomonas. However, the reported activity found in Chlamydomonas extracellular fractions hindered the functional testing of expressed hEGF.

To purify the fusion proteins expressed in Chapter 1, a purification was proposed based on affinity chromatography (specifically IMAC) to recover RP from concentrated culture supernatant. The step of culture supernatant concentration was solved at laboratory scale by lyophilization. Nevertheless, it is worth to note that the scalability of this step has been proven since we have successfully concentrated Chlamydomonas culture supernatant containing expressed hEGF up to 200-fold by tangential flow filtration (TFF) with a membrane of 10 kDa (results not shown). Regarding the lab scale procedure performed, RP was enriched and partially purified from concentrated culture supernatant (lyophilized) by IMAC and subsequently the fusion proteins were processed by TEV protease releasing discrete hEGF. However, hEGF represented a minimal proportion of total protein in the final fraction and, thus, further purification steps are needed. Our purification strategy was based on affinity chromatography that is a frequent first purification step when the target protein can tagged as it is in our case. The addition of a different extra tag would allow the performance of two affinity chromatography purifications in tandem, likely allowing higher purity levels. Another option is to add a purification step based on another type of chromatography. Ion exchange chromatography (IEX) separates proteins with differences in surface charge. In the present work an anionic exchanger column was tested for the purification of processed hEGF and the target protein did not bind to the column. Since pI may be far from the theoretical value, to lower the ionic strength of the buffer or to increase the pH of buffer are proposed to optimize the purification. Alternatively, an option is to test the purification with a cationic exchanger column.

It is important to note that the expression levels hamper the efficient purification of RP. In order to maximize RP yields, apart from optimizing RP expression itself, Chlamydomonas culture should be optimized. Chlamydomonas is usually cultivated in batches of TAP medium (containing acetate as a carbon source) and despite a fast growth, the culture enters the stationary phase of growth without reaching high densities, resulting in 1-2 g/L of dry weight biomass (Fields et al. 2018). For example *E.coli* grows to more than 100 g/L of dry weight biomass (Korz et al. 1995). High density cultures of other algae have been reported (Mandalam and Palsson 1998; Xiong et al. 2008; Kim et al. 2011), but only a few studies have explored the optimization of Chlamydomonas culture. One of these studies reported biomass densities of 9 g/L dry weight by using a combination of fed-batch and hollow-fiber cell-recycle systems (Chen and Johns 1995). Recently, a mixotrophic fed-batch cultivation strategy has been shown to significantly increase biomass density, productivity, and yield of RP (Fields et al. 2018). In this study culture densities were increased from 0.45 ± 0.03 to 23.69 ± 0.5 g/L dry weight and total RP

(GFP) was increased up to 2.5-fold (Fields et al. 2018). Despite these advances, a comprehensive study to extend the growth of Chlamydomonas is needed: exhausted nutrients, secondary metabolites inhibiting growth and medium optimization have to be analysed. Optimization of Chlamydomonas growth will enhance the host by maximizing its productivities and allowing more efficient purification processes.

Once recovered from concentrated culture supernatant, the different fusion proteins were successfully digested by TEV protease, releasing discrete hEGF except for the fusion protein gLuc:$(SP)_{20}$:TEV:hEGF (expressed by transformant UVM4#35). This fusion protein was successfully purified but was not cleaved by TEV protease, presumably because cleavage site is next to the $(SP)_{20}$ glycomodule and its physical conformation may hamper the cleavage. It has been shown in Chapter 1 that, while the other fusion proteins reach their maximum expression level when culture arrives at stationary phase, gLuc:$(SP)_{20}$:TEV:hEGF accumulates without detectable degradation for 48 h once the culture reaches stationary phase. The same factor that hinders the processing by TEV protease may be protecting the RP from proteases found in medium and thus may be responsible of the accumulation of this fusion protein to higher levels and for longer culture times than the other isoforms. This fusion protein accumulated in medium surpassing the 100 µg/L that is the highest yield reported in the present work. Thus, further work is essential to obtain processed hEGF at this expression yields. As a next step we propose the inclusion of a flexible linker between the TEV protease recognition sequence and the $(SP)_{20}$ glycomodule to analyse if the fusion protein is accumulated to the same levels (compared to the fusion protein without flexible linker) and if it can be efficiently cleaved by TEV protease. Flexible linkers are included to obtain a certain degree of movement or interaction (Chen et al. 2013).

To assess the activity of our hEGF isoforms a robust *in vitro* assay was needed. However, performance of the assay in complete medium did not allow to see induction of proliferation by hEGF and thus the assay had to be optimized. The optimal conditions for assessment of hEGF activity by induction of proliferation in NIH 3T3 by MTT assay were established: basal medium + 0.5% FCS, hEGF between 1 and 100 ng/ml and a starting number of $1x10^4$ cells per well.

Several fractions containing partially purified fusion protein or processed hEGF were tested *in vitro* on murine fibroblasts. Remarkably, all tested fractions including fractions from non-transformed UVM4 strain induced proliferation. The proliferation induction of Chlamydomonas secreted components is so high that hampers the analysis of recombinant hEGF activity.

Two fractions from non-transformed strains were tested in presence and absence of a commercial hEGF to evaluate if the induction of proliferation had an additive behaviour or if it was masked. The two extracellular fractions from non-transformed strains tested were concentrated culture supernatant from CC-124 strain and a UVM4 IMAC enriched fraction. When compared to the control condition (FCS 0.5%), the concentrated supernatant from CC-124 showed a significant increase in proliferation induction for the three concentrations tested while the partially purified fraction from UVM4 showed a significant increase in proliferation induction only in the less diluted concentration tested. When compared between themselves in presence or absence of hEGF, both show a synergistic effect on proliferation induction. Even though both resulted statistically significant, the induction of proliferation when hEGF was added to the fraction from UVM4 was more pronounced than when added to the concentrated culture supernatant from CC-124. This may be an overdosing effect similar to the observed at increasing concentrations of FCS.

The induction of proliferation by extracellular components described is interesting as it may have several applications. Extracellular polymeric substances (EPS) have recently been shown to have high value in diverse applications: anti-adhesive materials, antitumoral agents, anti-inflammatory, antivirals and skin care products (Xiao and Zheng 2016). The most abundant extracellular metabolites are carbohydrates, commonly named exopolysaccharides, but proteins, nucleic acids and lipids among others are also present in the medium. Polysaccharides from microalgae of the species *Parachlorella kessleri* or *Parachlorella beijerinckii* (Patent US 8,927,522 B2), and exopolysaccharides from *Pseudoalteromonas s.p* (Patent EP2646115 B1) have been demonstrated beneficial for skin care. Similarly, a glycoprotein from *Pseudoalteromonas antartica* (Patent EP1402898 B1) has been reported to improve the adhesion of human dermal fibroblasts and promote the cellular growth of human epidermal keratinocytes. Also, a purified fungal glycoprotein (Patent EP2875805 B1) has shown a stimulatory effect on keratinocyte cell regeneration. Given all the exposed, it is expected that the extracellular content of Chlamydomonas holds an unexplored potential.

Skin care is a potential application for the proliferation induction by extracellular components described since the cosmetic industry is continuously searching natural products to delay skin aging. Skin aging is manifested as a reduction in the extracellular matrix components (ECM) and polysaccharides and glycoproteins play a key structural role by forming a matrix to support the cells in the dermis and also regulate the activity of cells that surround them (Kleinman et al. 1981). Recombinant GF are already being used in cosmetics since GF are one of the components of ECM controlling cell growth, proliferation, and differentiation. It is important to note that the animal origin of some cosmetic actives used nowadays (as collagen) involve safety risks and are not well perceived from market. The synergistic effect described between Chlamydomonas secreted components and hEGF suggests that it may become a natural and effective cosmetic treatment.

Nowadays, the biopharmaceutical industry needs to adopt animal-free cell-culture systems to eliminate the variability between batches and the risks of contamination. Besides, the challenge is to adopt this culture components while reducing costs. Currently, plant hydrolysates (produced by the enzymatic or acidic digestion of a given raw material such as cotton or wheat) are being used to culture a variety of cell lines such as CHO (Merten et al. 1999; Ballez et al. 2004; Babcock et al. 2012). Thus, a potential application for the proliferation induced by Chlamydomonas secreted components described may be related to cell culture.

Given the potential industrial applications of the proliferation induced by Chlamydomonas secreted components, we have filed a patent application titled "*IMAC-enriched microalgal culture supernatant and uses thereof"*.

# Methodology

### *Chlamydomonas strains and culture conditions*

Algal cells were cultivated under standard conditions as previously stated (see *Chapter 1, Methodology*, page 129).

### *Protein isolation and Immunoblot analyses*

Immunoblots of concentrated culture supernatant were performed as previously stated (see *Chapter 1, Methodology*, page 129).

### *Lyophilization and dialysis of wild type culture supernatant*

To obtain a cell free concentrated culture supernatant, culture was centrifuged 4000 g 10 min to obtain a cell-free supernatant that was frozen at -80ºC and concentrated at least 50-fold by lyophilization followed by resuspension in $mQH_2O$. Concentrated supernatant was dialyzed against phosphate buffered saline 3 times 1/150 the volume of the sample. A second centrifugation (3000 g 5 min) was performed to remove aggregates and cell debris. Concentrated and dialysed culture supernatant was tested in proliferation assays.

### *IMAC purification*

Once culture reached late phase of growth, 400 ml of culture supernatant were separated from cells by the addition of 0.005% Tween20 followed by centrifugation at 4000 g 10 min, concentration 40X by lyophilization and resuspension in $mQH_2O$. Concentrated culture supernatant was centrifuged at 3000 g 5 min to remove aggregates and cell debris. Clarified concentrated supernatant was applied to 0.5 mL of Nickel Sepharose Excel resin (17371201, GE) previously equilibrated in Buffer A (50 mM potassium phosphate pH 8, 0.3

M NaCl, 0.1% Triton X-100), and incubated under slow rotation at 4ºC for 4 h in the presence of 1X Protease inhibitors (P9599, Sigma). Resin bound proteins were washed 5 times with 20 resin volumes of Buffer A. Elution was performed by incubating resin with 5 mL Buffer B (50mM potassium phosphate, 0.3 M NaCl, 0.1% Triton X-100, 0.5M imidazole). Concentration and diafiltration of eluted fraction was performed (Vivaspin® 20 10 kDa MWCO, 28-9323-60, GE Healthcare) to obtain a final 2 ml eluate in Buffer C (50 mM potassium phosphate pH 8, 50 mM NaCl). BSA was added to the elution at a final concentration of 0.1% (w/v).

Three resins were previously tested: Nickel Sepharose Excel resin (17371201, GE Lifesciences), Nickel HisPur™ Ni-NTA Resin (88221, Thermo Fisher Scientific) and HisPur™ Cobalt (89964, Thermo Fisher Scientific). These three resins were tested following the IMAC procedure stated before but with some changes: medium was concentrated 80X by lyophilization, 75 µl of resin were added to purify 500 µl of concentrated culture supernatant and elution was performed in 500 µl (E1 and E2). Resin was loaded into an SDS page once eluted by direct boiling in 500 µl of laemmli sample buffer. Each resin was tested directly with concentrated culture supernatant and with concentrated culture supernatant dialysed to Buffer A (50 mM potassium phosphate pH 8, 0.3 M NaCl, 0.1% Triton X-100).

### *TEV protease cleavage and second IMAC*

Elution from IMAC purification (200X relative to initial culture volume) was cleaved by TEV protease (ProTEV Plus, Promega) by addition of 1 unit of TEV protease per 100 µl of elution. Reaction was incubated under slow rotation during 2h at RT.

After the TEV protease cleavage reaction of elution from IMAC purification, 75 µl of Nickel Sepharose Excel resin (17371201, GE Lifesciences) previously equilibrated in PBS was added to the reaction and incubated for 1 h 45 min at RT under slow rotation. If TEV protease cleavage has been successful, hEGF should be in the flow through while TEV protease, gLuciferase and unespecifically co-purified proteins should be in the resin fraction.

### *IEX purification*

5L of supernatant from a culture that had reached OD750nm 2.25 were purified by IMAC as stated previously with the changes listed above:

- Culture supernatant was concentrated 60X by lyophilization.
- Clarified concentrated culture supernatant was applied to 1 mL of Nickel Sepharose Excel resin.
- Resin bound proteins were washed 4 times with 20 resin volumes of Buffer A.
- Concentration and diafiltration of the eluted fraction was performed (Vivaspin® 20 10 kDa MWCO, 28-9323-60, GE Healthcare) to obtain a final volume of 1.6 ml in Buffer C (50 mM potassium phosphate pH 8, 50 mM NaCl).

Elution from IMAC purification was cleaved by TEV protease by addition of 1 unit of TEV protease per 100 µl of elution. Reaction was incubated under slow rotation overnight at 4ºC.

Cleaved elution was applied to an anionic exchanger column (HiTrap® Q High Performance, GE) previously equilibrated in buffer C (50 mM potassium phosphate pH 8, 50 mM NaCl). Bound protein was washed with 5 column volumes of buffer C and elution was performed with a linear gradient of 20 column volumes: from 0% to 100% of buffer D (50 mM potassium phosphate pH 8, 1 M NaCl).

### *Proliferation assays - Mitogenic Activity Assessment*

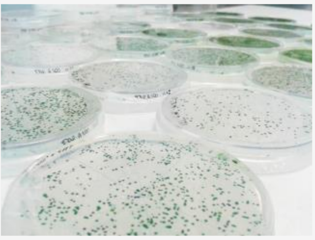
This work was performed by the Biomed division of Leitat.

A MTT (3-(4,5-dimethylthazol-2-yl)-2,5-diphenyltetrazolium bromide) test, with an NIH 3T3 cell line (mouse embryonic fibroblast cells), was used to assess functionality of hEGF produced in Chamydomonas. This assay is dependent on the reduction of the tetrazolium salt MTT by the mitochondrial dehydrogenase of viable cells to form a purple insoluble formazan product and thus measures viable cells at the end of the assay.

The sub-culturing process was carried out in 90% DMEM with 10% FCS at 37°C, 5% $CO_2$ and 90% humidity. Viability was checked every new sub-culture by trypan blue staining. MTT assays were performed with cells between an 80 and 90% of confluency and a viability superior to 90%.

10000 cells (unless otherwise stated) were plated in 100 µl of complete medium (90% DMEM and 10% FCS) with a 1% gelatine coating in 96 well plates. 24 h hours later a wash with DMEM (PBS with $Ca^{2+}$ and $Mg^{2+}$ in the optimization assays) was performed and the products (filter sterilized with a 0.22 µm nylon filter) dissolved in DMEM + 0.5% FCS (100 µl) were added. A commercial hEGF was added at 10 ng/ml (unless otherwise stated) as a positive control. Control experiments were carried out using the complete culture medium. The culture plates were incubated at 37°C in a $CO_2$ (5%) incubator for 72 hours. After the incubation, the culture medium was eliminated and new medium containing 10 µl of MTT (5 mg/mL in PBS, pH 7.4) was added and incubated for 3h at 37ºC. After this incubation, the medium was eliminated, extraction buffer was added (50% formamide and 15% SDS in $H_2O$) and let to react for 1h at 37ºC. After 4 hours, 150 µL of dimethylsulfoxide (DMSO) was added to all wells to dissolve the formazan crystals, and optical density (OD) was measured at a wave length of 570 nm (Multiskan Ascent, Thermo Fisher Scientific). The value obtained at 0% viability (cells treated with

2.5% triton X-100) was considered background signal. For normalization, the cell viability in each condition was expressed as a percentage relative to viability in DMEM 0.5% FCS that was set at 100%.



# Chapter 3. Expression of a monoclonal antibody from Chlamydomonas nuclear genome

# Introduction

Over the past decades, the value of monoclonal antibodies (mAbs) has been demonstrated by their wide and diverse applications as research tools, in industrial processes, in diagnostics, and as therapeutics. Due to their wide range of applications, there is a high demand for mAbs. Global sales revenue for therapeutic mAbs was nearly $75 billion in 2013, representing approximately 50% of the total sales of all biopharmaceutical products, and has been estimated to grow to nearly $125 billion by 2020 (Ecker et al. 2015). This market is continually growing with dozens of new mAb therapeutics in different clinical phases. With 10 new approved mAb products in US in 2017 (6 also in EU), a new record for the annual number of novel antibody therapeutics approved has been set (Kaplon and Reichert 2018). Approved mAbs are mainly produced in Chinese hamster ovary (CHO) cells and some are produced in murine myeloma cell lines NS0 and SP2/0 (Liu 2015). Although this mammalian systems produce effective and safe recombinant mAbs, they have a limited manufacturing capacity and high costs. An alternative to meet market demands is to produce mAbs in heterologous expression systems such as bacteria, yeast, insect cells, or plants. Each of these expression hosts has its own limitations in product efficacy, safety, response time, scalability and cost. See the general introduction of this thesis for more extensive information about expression platforms (*Introduction, page 1*). Briefly, the problem when using bacteria as an expression host is the inability for correctly folding complex proteins. In the case of insect cells, yeast and plants, glycosylation patterns differ from those of humans and this may introduce not only differences in the activity of the RP but also the risk of immunogenicity.

Glycoengineering, generally applied to mAb production in non-mammalian hosts such as yeast, insect cells and plants, aims to humanize the mAb glycosylation profile reducing the immunogenicity risk or improving mAb biological function (Li et al. 2006). Most FDA approved mAbs are human or humanized, being humanization intended to reduce the immunogenicity of rodent-derived mAbs in humans and improve their biological functions (Ahmadzadeh et al. 2014). The effects of glycosylation on pharmacodynamics of antibodies has been well reported: glycosylation is critical on IgG receptor-mediated effector functions (Nose and Wigzell 1983; Tao and Morrison 1989; Liu 2015). Besides, it is also known that glycosylation may positively or negatively impact on the Fc effector functions (Shields et al. 2002). Glycosylation in the variable region can also influence antibody function: it has been shown that glycosylation in the variable fragment region can enhance or reduce the antigen-binding activities (Wallick et al. 1988; Wright et al. 1991).

New mAb modalities are also being developed to deliver cytotoxic drugs (antibody-drug conjugates, immunotoxins and immunoliposomes) (Pirie et al. 2011; Sapra and Shor 2013) or immunomodulatory cytokines (immunocytokines) to tumor cells (Carnemolla et al. 2013; Danielli et al. 2015). Also, in addition to full-size mAbs, smaller antibody fragments capable of antigen binding are also studied and employed: a variety of single-chain variable fragments (scFvs) formats  consisting of variable domains of heavy chains and light chains (Ahmad et al. 2012), and camelid VHH domains (nanobodies). Camelid VHH domains are the smallest antibody-derived biologically active molecules with a high antigen-binding capacity while being water soluble, heat stable, and pepsin resistant (Van Der LindenY et al. 1999).

Plants are promising as an expression host for mAbs and other RP because are cost effective and scalable as well as safe (no mammalian pathogens).

Furthermore, plants can perform PTMs of target proteins, including glycosylation. However, glycosylation is a concern as the plant glycosylation pattern may induce immunogenicity in humans. Plant expressed mAb started with the expression and assembly of a full-length antibody in transgenic Nicotiana tabacum plants almost 30 years ago (Hiatt et al. 1989). Later, a hybrid secretory IgA and IgG mAb that recognized surface antigen I/II of Streptococcus mutants *in vitro* (Ma et al. 1995) was expressed in *N. tabacum* plants and was further evaluated for the treatment of dental caries in humans (Ma et al. 1998). Since then, different transgenic plant hosts have been used to produce a variety of mAbs (with varied yield and quality) (De Muynck et al. 2010) and engineered antibody fragments of potential importance for a diversity of applications (Benvenuto et al. 1991; Owen et al. 1992; de Neve et al. 1993).

In 2012, the first plant biopharmaceutical was approved for use in humans meaning a key step towards market acceptance and commercial viability for plant biopharmaceuticals (Mor 2015). In addition, the recent success with the experimental treatment of Ebola patients with a mixture of plant-produced mAbs (ZMapp) during the 2014–2015 Ebola crisis in West Africa is a prove of the potential of this approach. Also, a monoclonal antibody for the purification of the hepatitis B vaccine has been expressed in plants (Valdés et al. 2003).

Despite the benefits as an expression host, plants have also some drawbacks: risk of crops displacement, possibility of gene flow and a costly downstream processing. Microalgae have attracted attention as an alternative recombinant protein host because they offer the benefits of plants coupled with high productivities associated with microbial growth. Thus, microalgae are promising as expression hosts due to: low production cost, possibility of growth in containment, no gene flow by means of pollen or other vehicles, high growth rates, better safety (no human pathogens) and, in the case of eukaryotic microalgae, the ability to accomplish PTMs (Rosales-Mendoza 2016). Also, some algae species hold a GRAS (generally recognized as safe) status and thus can be orally administered to humans as oral vehicle of vaccines, therapeutic RP or food supplement both for veterinary or human purposes (Specht and Mayfield 2014; Beltrán-López et al. 2016). Another advantage is the quick generation of RPs, requiring only weeks between the production of transformants and their scale up to production volumes. See the general introduction of this thesis for more information about RP hosts (*Introduction, page 1*).

The diatom *Phaedodactylum tricornutum* has been studied as a feasible alternative host for mAb production. Several fully-assembled mAbs have been expressed in *P. tricornutum* retained in the endoplasmic reticulum (Hempel et al. 2011) or secreted to medium (Hempel and Maier 2012; Hempel et al. 2017) with a maximum expression level reported of 2500 ng/ml (Hempel and Maier 2012). One of this expressed mAbs in *Phaedodactylum tricornutum* showed a reduced affinity to its receptors compared to mammalian mAb, and it has been suggested to be due to differences in glycosylation of the Fc region (Vanier et al. 2017).

Despite the fact that RP have been successfully expressed in other microalgae, current work is performed with *Chlamydomonas reinhardtii* (Chlamydomonas hereafter). Since Chlamydomonas has been a model organism for decades (Harris 2001) it is the best characterized microalgal species and a variety of proteins have been successfully expressed in Chlamydomonas. mAb expression history in Chlamydomonas is briefly presented below.

The first algae-expressed mAb was expressed in Chlamydomonas chloroplast and consisted in a human antibody against a glycoprotein (GP-D) of Herpes Simplex Virus Infection (HSV). The antibody consisted of a unique large single chain (HSV8-lsc) containing the entire IgA heavy chain fused to the variable region of the light chain by a flexible linker (Mayfield et al. 2003). The antibody was expressed in a soluble form and was properly dimerized by disulphide bonds.

Also, a full-length neutralizing human IgG1 mAb, 83K7C, directed against PA83 of *Bacillus anthracis*, was expressed in Chlamydomonas chloroplast (Tran et al. 2009). Both chains accumulated as soluble proteins and were fully-assembled. *In vitro* studies showed that algae-based 83K7C binds its antigen similarly to its mammalian counterpart (Tran et al. 2009).

Immunotoxins (chimeric mAbs fused to eukaryotic toxins) cannot be expressed in eukaryotic cells because of their toxicity and cannot be expressed in bacteria due to the bacterial inability to properly fold complex proteins. Algal chloroplasts allow proper folding and assembly of complex eukaryotic proteins thanks to the presence of chaperones and isomerases while remaining resistant to eukaryotic toxins (Tran et al. 2013b). This advantage allowed the expression of an immunotoxin designed for the treatment of B cell lymphoma in Chlamydomonas chloroplast. This immunotoxin contained an scFv (single-chain variable fragment) against the B cell surface antigen CD22 fused to the translocation and enzymatic domains (domains II and III) of exotoxin A from *Pseudomonas aeruginosa* (PE40) (Tran et al. 2013b). Expression and accumulation (to 0.2–0.4% of TSP) of monomeric and dimeric immunotoxin proteins was shown. Such Immunotoxins possessed binding and cytotoxic activity against B cell lymphoma lines *in vitro*. Moreover, treatment with this immunotoxins significantly prolonged the survival of immunodeficient mice with implanted human B cell xenograft tumors (Tran et al. 2013b). In a different study, monomeric and dimeric immunotoxins αCD22Gel and αCDCH23Gel were also expressed in Chlamydomonas chloroplast by fusing anti-CD22 scFv (single chain variable fragment) to gelonin, a toxin from *Gelonium multiflorum* which fusions to scFv antibodies were previously shown to reduce the viability of targeted tumor cells (Pirie et al. 2011). This immunotoxins accumulated in Chlamydomonas chloroplasts up to 0.1–0.3% of TSP (Tran et al. 2013a) and both were able to bind to and reduce the viability of B cell lymphoma lines but not T cells that do not express CD22 (Tran et al. 2013a).

Finally, three different camelid VHH domain–based antitoxins were expressed in Chlamydomonas chloroplast for the treatment of botulism: two monomers (C2 and H7) and one heterodimer (H7-fs-B5) containing two different VHH domains separated by a flexible spacer (Barrera et al. 2015). The antitoxins accumulated as soluble proteins up to ∼5% of TSP and were able to bind to botulinum neurotoxin serotype A similarly to VHH antibodies produced in *E.coli*. It was also shown that these antitoxins can be orally delivered intact to the stomach and small intestine of mice.

Despite the advantages that chloroplast offers, there are applications for which glycosylation is crucial (antibody glycosylation has a critical role in antibody effector function (Tao and Morrison 1989; Liu 2015)) and for this applications mAbs could not be expressed in Chlamydomonas chloroplast. Consequently, expression of mAbs in Chlamydomonas nucleus is the next step to further study the suitability of Chlamydomonas as a versatile recombinant protein expression host.

The challenge we face in the present Chapter is the nuclear expression of a mAb in Chlamydomonas. To do that, we design a vector to express both chains to high levels and we develop a screening that aims to identify transformants expressing the fully assembled mAb. Analysis of the mAb expression and mAb characteristics will be conducted.

The work presented has been financed by the RETOS-Colaboración 2015 program of Ministerio de Economia, Industria y Competitividad of Spain as a partnership with Leitat Biomed.

# Results

## 1. Cassette design

The selected monoclonal antibody (mAb) to be expressed was m5C3. This monoclonal antibody belongs to murine IgG1 isotype and recognises S100A4 soluble protein. S100A4 is a member of the S100 calcium-binding protein family secreted by tumor and stromal cells and has been shown to be crucial in promoting metastasis (Schmidt-Hansen et al. 2004). Moreover, S100A4 has been shown to be a good biomarker for tumor diagnostic and therapies using antibodies against S100A4 have been considered promising strategies to treat cancer (Hernández et al. 2013).

Our expression strategy is based on the design of a single vector (figure 72) containing three cassettes: one for the expression of an antibiotic resistance for initial selection of transformants (Ble$^R$), one for the expression of the heavy chain (HC) and one for the expression of the light chain (LC). Since repetitive sequences may be silenced in Chlamydomonas (Ceruttti et al. 1997; Schubert 2004), we included three different cis regulatory regions (5’ and 3’): AR promoter and 3’UTR *RBCS2* to drive expression of bleomycin resistance, ferredoxin 1 (*FDX1*, also called *PETF*, Cre14.g626700) cis regulatory regions (5’ and 3’) to drive expression of the heavy chain (HC) and Chlamydomonas Ribosomal protein L23 (*RPL23*, Cre04.g211800) regulatory regions to drive light chain (LC) expression. The *FDX1* and *RPL23* highly expressed regulatory regions have been recently described (López-Paz et al. 2017) and we previously corroborated the increase of RP expression levels with *RPL23* cis regulatory regions (*Chapter 1, Strategy 1, 2.2 Comparison of the efficiency of two different cis regulatory regions to drive transgene expression*). It is important to highlight that *RPL23* gene has naturally occurring introns in its 5’ and 3’ untranslated regions (López-Paz et al. 2017). The 5’cis regulatory region (1004 bp) contains the predicted promoter and 5’ UTR that is predicted to contain an intron. The 3’UTR is also predicted to contain an intron and adjacent DNA is used (total of 725 bp).

The sequences coding for m5C3 were provided by Leitat Biomed (Patent WO2011157724 A1). Sequences coding for m5C3 were codon adapted with IDT web tool (https://eu.idtdna.com/CodonOpt) and murine secretion sequence was maintained.

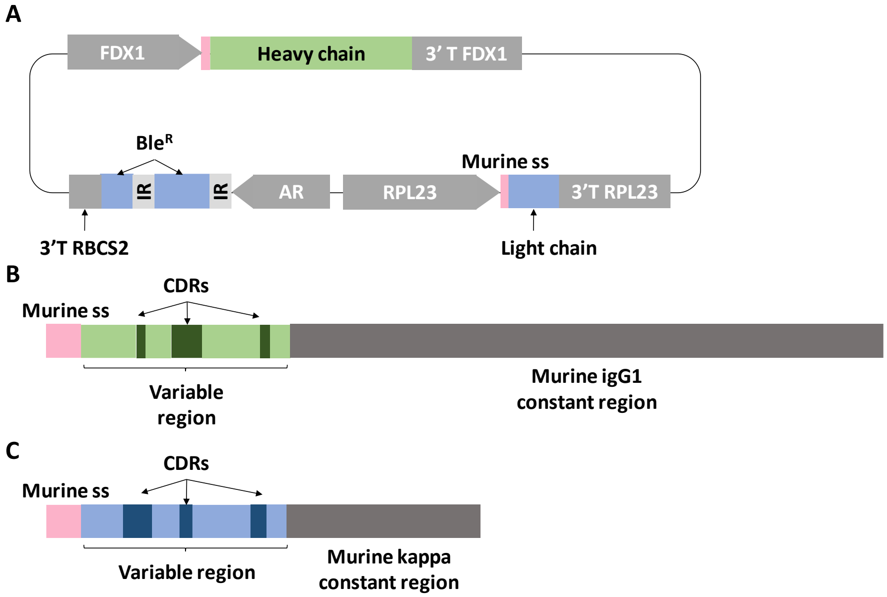

A
FDX1
Heavy chain
3’ T FDX1
Ble^R
IR
IR
AR
RPL23
Murine ss
3’T RPL23
3’T RBCS2
Light chain
B
CDRs
Murine ss
Variable region
Murine igG1 constant region
C
CDRs
Murine ss
Variable region
Murine kappa constant region

**Figure 72. Schematic representation of the vector used for the expression of m5C3 antibody. A) FM-zeo$^R$ vector map. B) Schematics of Heavy chain. C) Schematics of Light chain.** AR: *HSP70-RBCS2* chimeric promoter, IR: intron 1 *RBCS2*, 3’T RBCS2: 3’UTR and terminator of *RBCS2*, RPL23: 5’ *RPL23* cis regulatory elements including 5’UTR and promoter, 3’T RPL23: 3’UTR and terminator of *RPL23*, FDX1: 5’ *FDX1* cis regulatory elements including 5’UTR and promoter, 3’T FDX1: 3’UTR and terminator of *FDX1*, CDRs: Complementarity-determining regions.

## *2. Screening of transformants*

### 2.1 Initial screening of transformants

As mentioned earlier, the high variability in expression levels among Chlamydomonas transformants makes necessary the development of high throughput screenings to select the highest expressing transformants. Since our objective is to select a transformant expressing a fully assembled mAb, the screening should lead to transformants expressing both chains simultaneously in a high throughput manner. Our approach was based on an initial selection of transformants resistant to zeocin, thus indicating the insertion of heterologous DNA. Subsequently, the technique of choice was sandwich ELISA (Enzyme-Linked Immunosorbent Assay). For the present study the plate was coated with an Anti-Mouse IgG, Fc-γ Fragment Specific antibody, then samples were incubated (fully-assembled mAb would be captured) and finally a peroxidase conjugated anti-mouse IgG, F(ab')2 fragment specific (that would bind if a complete mAb has been captured previously) was used to develop the assay (see figure 73).

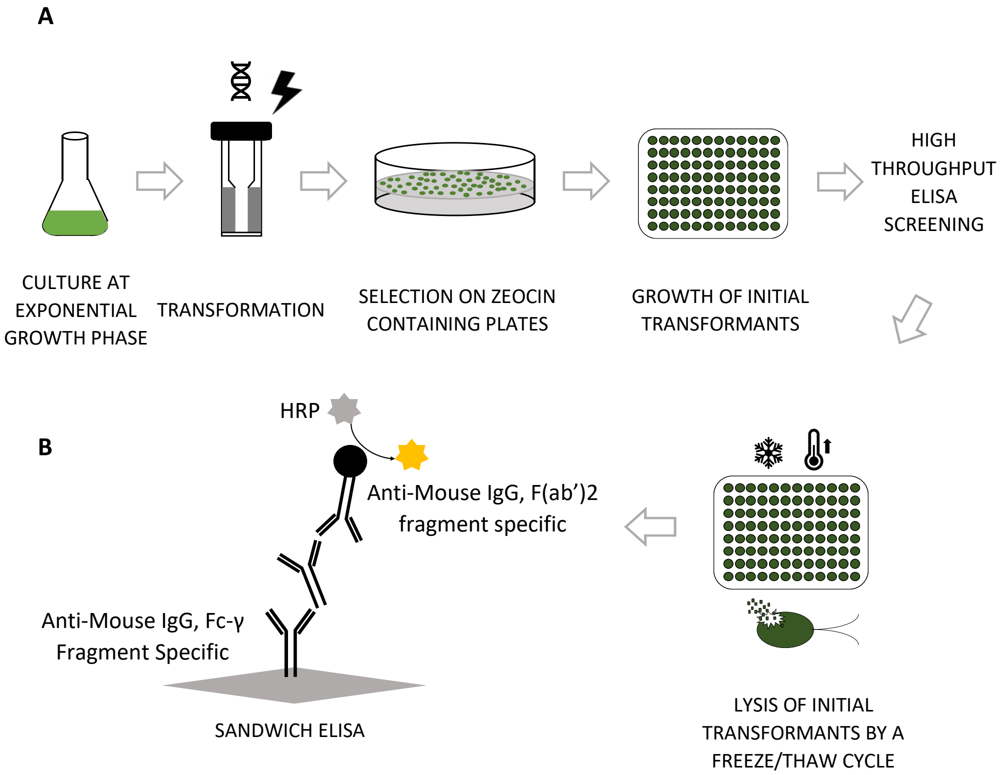
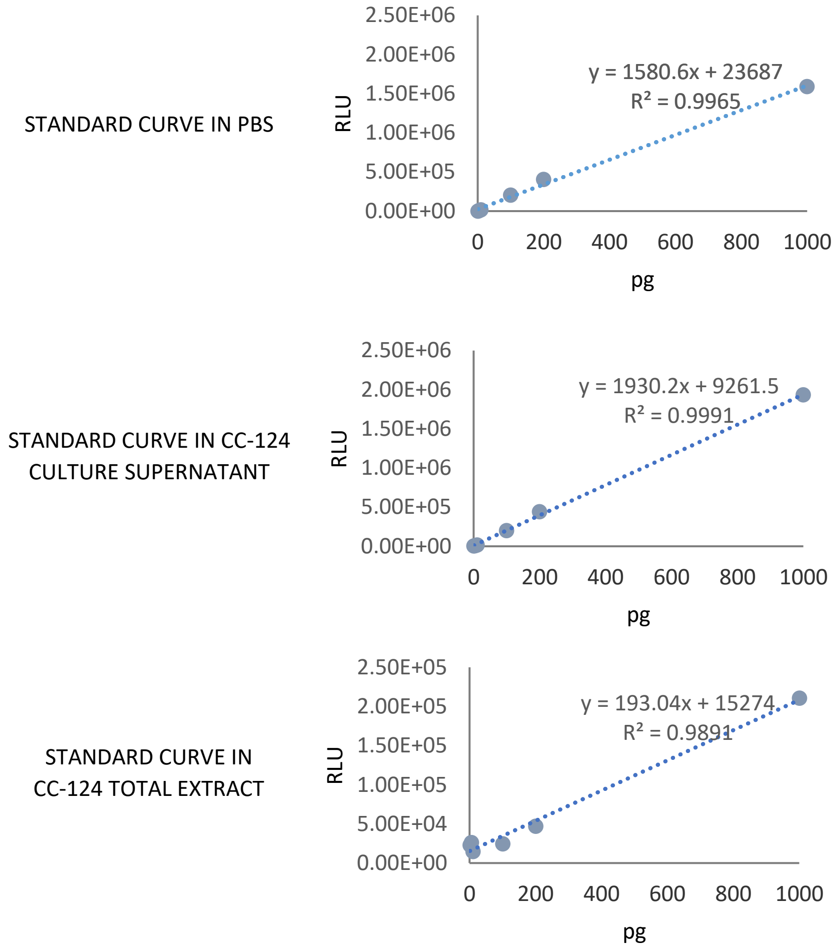

A
CULTURE AT EXPONENTIAL GROWTH PHASE
TRANSFORMATION
SELECTION ON ZEOCIN CONTAINING PLATES
GROWTH OF INITIAL TRANSFORMANTS
HIGH THROUGHPUT ELISA SCREENING
B
HRP
Anti-Mouse IgG, F(ab')2 fragment specific
Anti-Mouse IgG, Fc-γ Fragment Specific
SANDWICH ELISA
LYSIS OF INITIAL TRANSFORMANTS BY A FREEZE/THAW CYCLE

**Figure 73. High throughput screening designed to select Chlamydomonas transformants expressing a fully-assembled mAb. A)** High throughput screening proposed. **B)** Sandwich ELISA detail.

It is important to highlight that selected transformants were always analysed by immunoblot since the secondary antibody used to perform the sandwich ELISA is an antibody against mouse IgG F(ab')2 region (present in both chains) and, although it recognizes preferentially the LC, it is possible that a positive transformant by sandwich ELISA only expresses one chain.

The inclusion in the assay of standard curves together with wild type cell extract or culture supernatant corroborated that neither the crude cell extract nor culture supernatant prevent the mAb from being recognised by the assay (figure 74).


STANDARD CURVE IN PBS

STANDARD CURVE IN CC-124 CULTURE SUPERNATANT

STANDARD CURVE IN CC-124 TOTAL EXTRACT

**Figure 74. Sandwich ELISA standard curves in PBS, in culture supernatant from CC-124 and in total cell extract from CC-124.**

The plasmid FM-zeo$^R$ (see figure 72) was transformed into the cell walled strain CC-124 and 184 initial transformants (resistant to 15 µg/ml zeocin) were screened by sandwich ELISA. Since we had no information about if the used murine secretion sequence would be functional in Chlamydomonas, our screening was performed with 100 µl of total culture (including culture medium and intracellular content from lysed cells that were previously disrupted by a cycle of freeze/thaw). Four out of 184 transformants showed, at least, a 5-fold increase of signal relative to background and were selected for further analysis: 18B1, 19A1, 19A2 and 19A3. Two negative transformants were also selected as negative control strains for further studies: 18B2 and 19B1.

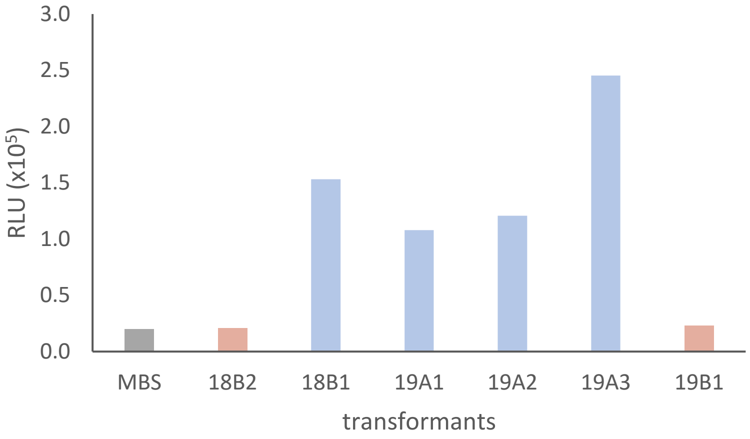


**Figure 75. Sandwich ELISA screening of CC-124 transformants.** The assay was performed with total cell extract. Results are presented as RLU (relative luminescence units). Results are shown only for selected transformants for further analysis (four positives marked in blue and two negatives marked in light brown). MBS; Mean background signal.

The expression level of the six selected transformants was further confirmed in a second assay where medium was separated from cells to examine signal present in each fraction. These results confirmed 19A3 as the highest expressing transformant with 99.5% of the mAb found in culture medium (figure 76).

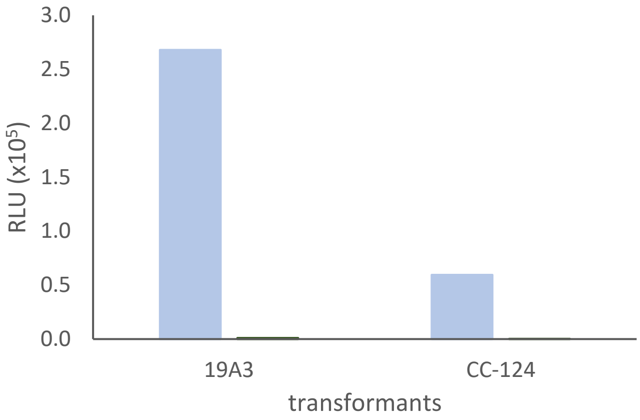


**Figure 76. Sandwich ELISA analysis of intracellular and extracellular mAb expression in 19A3 transformant.** Cells were grown to late log phase of growth and assay was performed with culture supernatant (results presented as blue bars) or cell extracts (results presented as green bars). RLU: relative luminescence units.

In order to study the expression of mAb by 19A3 transformant, mAb was determined by sandwich ELISA in samples taken at different culture stages throughout the culture growth. As it can be seen in figure 77, whereas maximum intracellular expression occurred during mid exponential phase, mAb is accumulated in medium reaching its maximum expression at late log/stationary phase of growth.

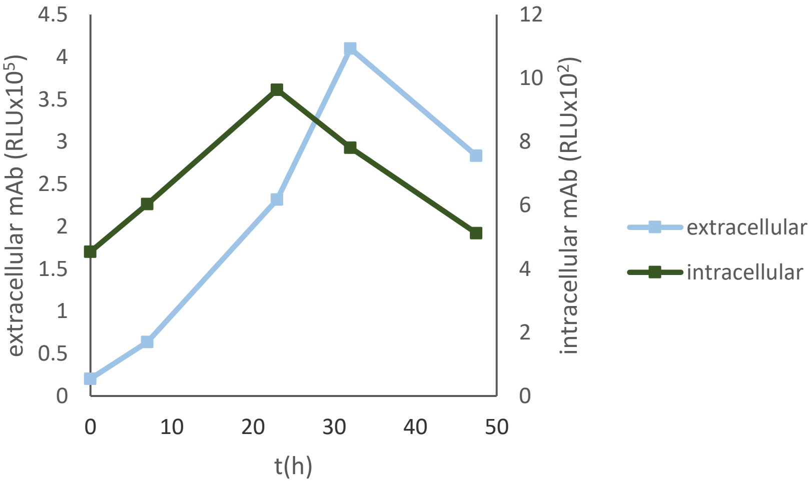


**Figure 77. Kinetics of mAb expression in 19A3 transformant analysed by sandwich ELISA.** Signal corresponding to intracellular mAb is presented in green and referred to right axis. Signal corresponding to extracellular mAb is presented in blue and referred to left axis. Intracellular and extracellular signal correspond to 100 µl of culture.

To be able to evaluate mAb by immunoblot, culture medium from 19A3 transformant at the late exponential phase of growth (named P1) was concentrated by tangential flow filtration (30 kDa) to 45.33 X (named P2) and then it was concentrated by centrifugation-based filtration (10 kDa) to 1700 X (named P3). mAb was analysed by sandwich ELISA and by immunoblot at each concentration step (figure 78). The sandwich ELISA results obtained indicated that mAb concentration increase was not proportional to volume reduction. A subsequent sandwich ELISA analysis performed with higher dilutions of the samples indicated that there is an inhibitor of the assay in culture medium that concentrates together with mAb and hampers the detection of mAb in concentrated samples (results not shown). The immunoblot analysis was performed with an antibody against mouse IgG F(ab')2 region (detects preferentially the LC) and an antibody against mouse IgG (detects preferentially the HC) and showed that, while there was detectable LC, there was undetectable HC. Thereby, we decided to transform 19A3 transformant with an additional copy of HC to increase the expression of the fully-assembled mAb (*2.2 Retransformation and screening of transformants*).

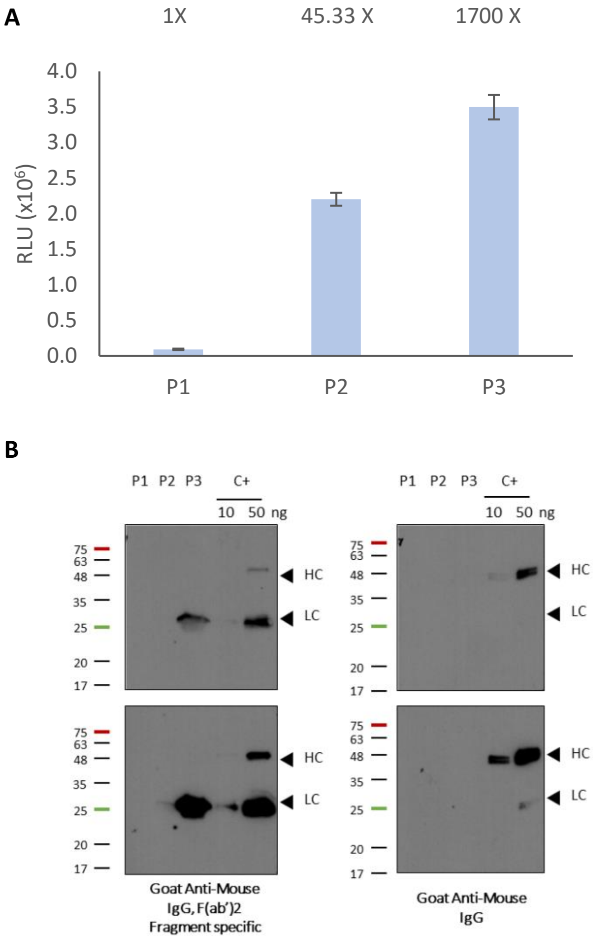


**Figure 78. Analysis of mAb expressed in 19A3 clone after concentration by filtration. A)** Sandwich ELISA results in RLU (relative luminescence units). Error bars show the standard deviation of two technical replicates. Above bars there is indicated the concentration relative to initial culture (volume) at each concentration step. **B)** Immunoblot analysis of concentrated mAb. Primary antibody anti-mouse IgG, F(ab')2 fragment specific (detects preferentially the LC) or anti-mouse IgG (detects preferentially the HC). C+: mouse monoclonal anti-Thymine Dimer antibody, HC: heavy chain, LC: light chain.

As an alternative, the cell wall-less mutant strain UVM4 was also transformed with the plasmid FM-zeo$^R$ (see figure 72). 96 initial transformants (resistant to 15 µg/ml zeocin) were screened by sandwich ELISA. Growth and assay conditions were identical to the ones performed for the screening of CC-124 strain transformants. Nine out of 96 transformants showed, at least, a 50-fold increase of signal relative to background. Four of these nine transformants showed, at least, a 100-fold increase of signal relative to background and thus selected for further analysis.

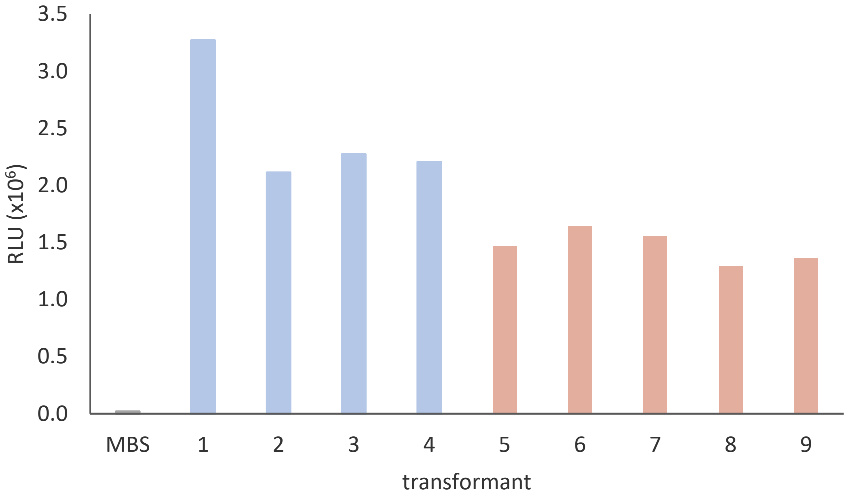


**Figure 79. Sandwich ELISA screening of UVM4 transformants.** The assay was performed with total cell extract. Results are presented as RLU (relative luminescence units). Results are shown only for positive transformants: selected transformants for further analysis are represented as blue bars, positive transformants not selected for further analysis are represented as brown bars. MBS; Mean background signal.

To evaluate mAb expression in UVM4 selected transformants (1, 2, 3 and 4) an immunoblot analysis was performed (figure 80). The immunoblot was performed using an anti-mouse IgG and an anti-mouse F(ab')2 region. Importantly, no band or signal was detected in UVM4 negative control with any of the antibodies used.

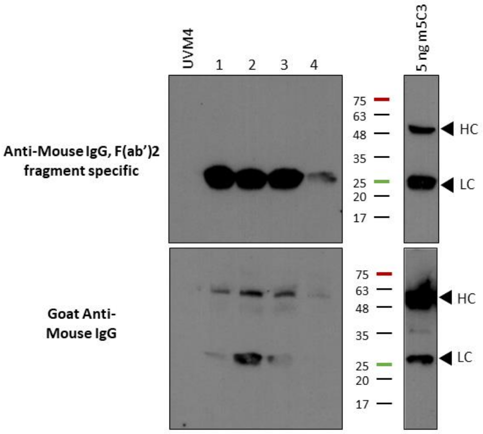

UVM4
1
2
3
4
5 ng m5C3
Anti-Mouse IgG, F(ab')2 fragment specific
Goat Anti-Mouse IgG
75
63
48
35
25
20
17
HC
LC

**Figure 80. Immunoblot evaluation of mAb expressed in UVM4 selected transformants**. Primary antibody anti-mouse IgG, F(ab')2 fragment specific or anti-mouse IgG. 28 µl of 100X concentrated culture supernatant were loaded. m5C3: purified monoclonal 5C3 antibody expressed in mammalian cells, HC: heavy chain, LC: light chain.

Since it expresses both HC and LC, transformant UVM4-2 was selected for further analysis. An immunoblot was performed under reducing and non-reducing conditions with the aim of further characterizing expressed mAb (figure 81). As it can be seen in figure 81, immunoblot under reducing conditions confirmed the expression of both chains and specific signal at a size that corresponds to fully assembled antibody was observed under non-reducing conditions. However, under both conditions (reducing and non-reducing) stoichiometry differs from m5C3: LC is overexpressed relative to HC, presumably due to poor expression or higher instability of this chain compared to LC.

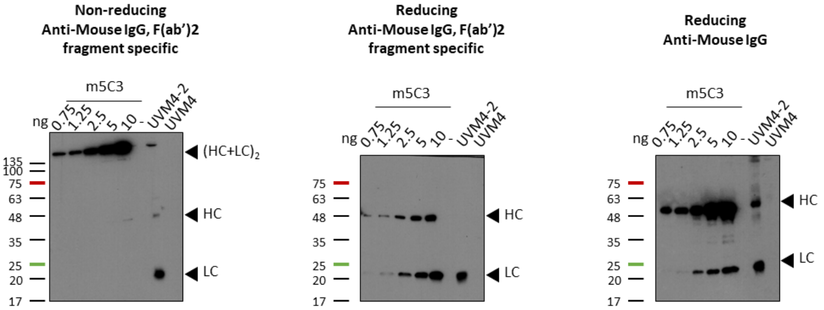

Non-reducing
Anti-Mouse IgG, F(ab')2 fragment specific
Reducing
Anti-Mouse IgG, F(ab')2 fragment specific
Reducing
Anti-Mouse IgG
m5C3
ng
0.75
1.25
2.5
5
10
-
UVM4-2
UVM4
135
100
75
63
48
35
25
20
17
(HC+LC)2
HC
LC

**Figure 81. Immunoblot evaluation of mAb expressed by UVM4-2 transformant**. Primary antibody anti-mouse IgG, F(ab')2 fragment specific or anti-mouse IgG. 12 µl of 80X concentrated culture supernatant were loaded. m5C3: purified monoclonal 5C3 antibody expressed in mammalian cells, $(HC+LC)_2$: complete mAb, HC: heavy chain, LC: light chain.

## 2.2 Retransformation and screening of transformants

As it has been demonstrated for both transformants analysed (CC-124 19A3 and UVM4-2) the expression of chains is differential: there is an excess of LC relative to HC. In the case of 19A3 the expression of HC does not achieve detectable levels. Due to that, we decided to retransform 19A3 transformant with an additional copy of HC to increase the expression of the fully-assembled mAb. To that end, we designed the HC-paro$^R$ vector (figure 82).

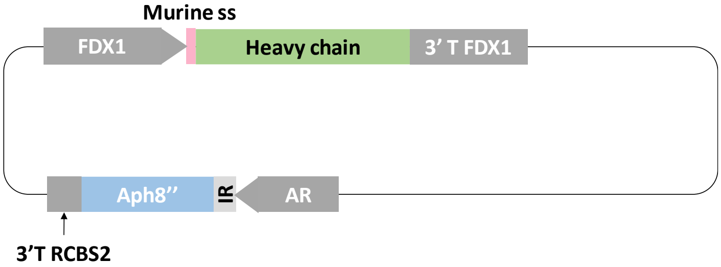


**Figure 82. HC-paro$^R$ vector map.** AR: *HSP70-RBCS2* chimeric promoter, 3'T RBCS2: 3'UTR and terminator of *RBCS2*, IR: intron 1 *RBCS2*. FDX1: 5' *FDX1* cis regulatory elements including 5'UTR and promoter, 3'T FDX1: 3'UTR and terminator of *FDX1*.

HC-paro$^R$ vector was transformed into 19A3 transformant and 376 initial transformants (resistant to 25 µg/ml paromomycin) were analysed by sandwich ELISA as reported (*2.1 Initial screening of transformants*). When performing the sandwich ELISA, the median RLU value of all the initial transformants was taken as background since it was similar to 19A3 signal. Therefore, clones that showed RLU 3-fold above background signal were considered positives and were selected for further analysis (figure 83).

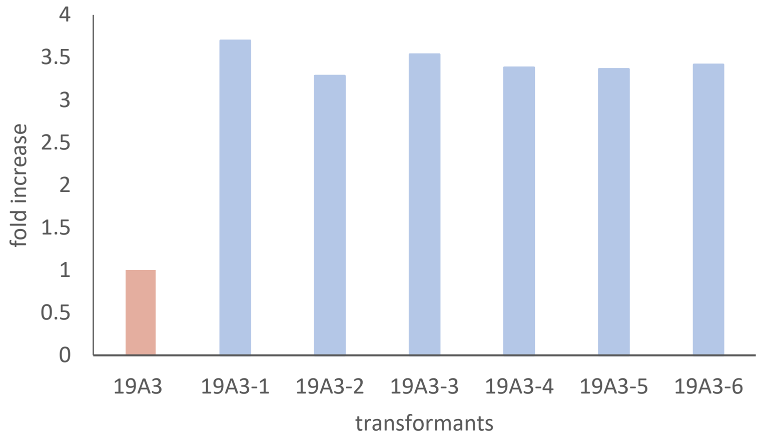


**Figure 83. Sandwich ELISA screening of 19A3 transformants.** The assay was performed with total cell extract. Results for selected transformants are presented as fold-increase relative to 19A3 expression level.

Because in the initial screening some values were close to the saturation range of the assay, a second sandwich ELISA was performed to accurately determine expression levels of the six selected transformants (figure 84).

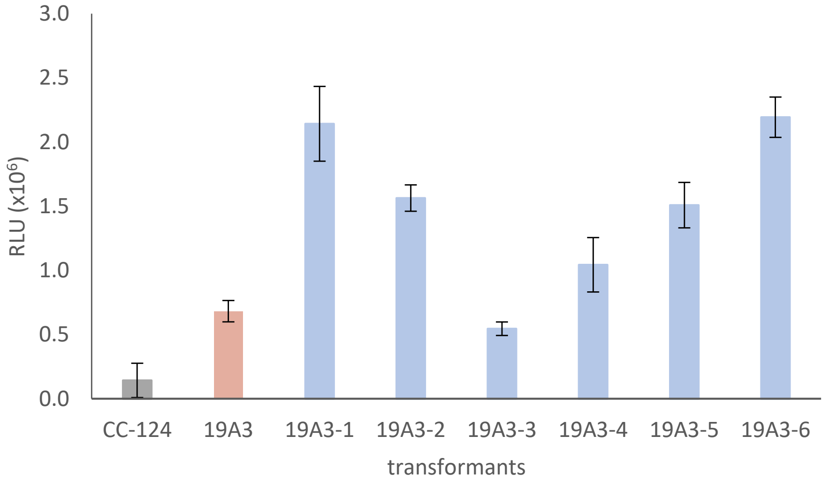


**Figure 84. Sandwich ELISA confirmation of 19A3 selected transformants.** Results are presented as RLU (relative luminescence units). Cells were grown to late log phase and assay was performed only with culture supernatant. Standard deviations of five technical replicates are presented as error bars.

Transformants 19A3-1 and 19A3-6 were the highest expressing transformants and its mAb expression level was estimated by sandwich ELISA in a subsequent analysis (results shown in figure 85). Corroborating the results obtained with 19A3 (figure 76, page 197) in the three analysed transformants the expression of mAb was mainly found in culture medium: approximately 1% of mAb was found in cell extracts.

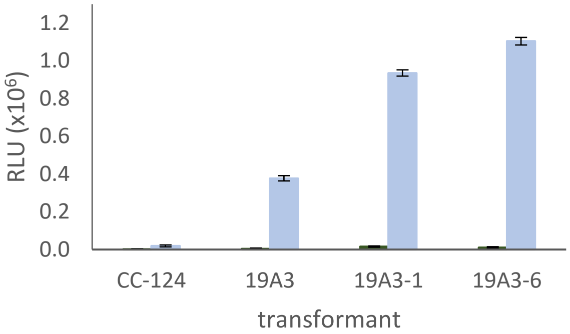


**Table 15. Sandwich ELISA quantification of extracellular ( E ) and intracellular ( I ) mAb expression in selected transformants.**

| Transf. | E (µg/L) | I (ng/L) |
|---|---|---|
| **19A3** | 0.07 | 0.82 |
| **19A3-1** | 0.26 | 4.66 |
| **19A3-6** | 0.32 | 3.17 |

**Figure 85. Sandwich ELISA quantification of intracellular and extracellular mAb expression in CC-124 selected clones.** Cells were grown to late log phase of growth and assay was performed with culture supernatant or cell extracts. m5C3 standard curve was included to estimate expression in transformants. Sandwich ELISA results are presented in RLU (relative luminescence units), green bars represent signal corresponding to intracellular mAb and blue bars represent signal corresponding to extracellular mAb. Error bars show the standard deviations of three technical replicates.

To evaluate mAb expression in the selected transformants 19A3-1, 19A3-6 and 19A3 an immunoblot analysis was performed. Immunoblots were performed using an anti-mouse IgG Fc region and an anti-mouse IgG F(ab')2 region under reducing conditions (figure 86). Also, an immunoblot under non-reducing conditions was performed with an antibody against mouse IgG (figure 86). As it can be seen under reducing conditions there is an increase in the expression of HC in the transformants 19A3-1 and 19A3-6 relative to its 19A3 background strain. Moreover, it can be seen under non-reducing conditions that there is an increase in the expression of fully assembled mAb in transformants 19A3-

1 and 19A3-6 relative to 19A3 that correlates with the increase of expression of HC.

Expression of mAb in selected clones was estimated on the base of these immunoblots. Based on the non-reducing immunoblot approximately 0.4 µg/L of complete mAb was estimated in clones 19A3-1 and 19A3-6 and around a 0.1 µg/L was estimated in 19A3 transformant. Based on the reducing immunoblot using anti-F(ab')2 region approximately 1.5 µg/L of the LC were estimated in 19A3-1, 19A3-6 and 19A3 transformants. In the case of the HC, based on the reducing immunoblot anti- mouse IgG (detects preferentially the HC) around a 0.2 µg/L were estimated for clones 19A3-1 and 19A3-6 whereas the HC was too low to be quantifiable in 19A3 transformant. That estimation of expression corroborates that the increase in HC copy number resulted in an increase of the expression of complete mAb.

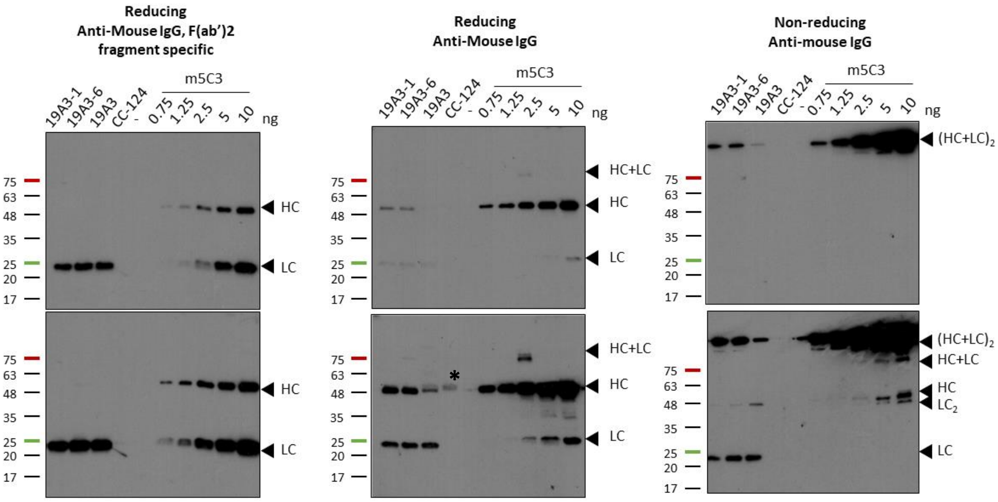


**Figure 86. Immunoblot evaluation of mAb expressed in 19A3-1 and 19A3-6 transformants and comparison to its background strain 19A3**. Primary antibody anti-F(ab')2 fragment specific or anti-mouse IgG. 12 µl of 80X concentrated culture supernatant were loaded. Two exposures are shown for each immunoblot: the one showed above corresponds to seconds of exposure and the one showed below corresponds to minutes of exposition. m5C3: purified monoclonal 5C3 antibody expressed in mammalian cells, (HC+LC)$_2$: complete mAb, HC: heavy chain, LC: light chain. * Indicates a non-specific recognition in CC-124 of a protein slightly higher than the HC.

### 2.3 mAb expression characterization in selected transformants

In order to analyse the expression of mAb in selected transformants (19A3-1 and 19A3-6 of CC-124 strain and UVM4-2) mAb was quantified by sandwich ELISA in samples taken at different culture stages throughout the culture growth (figure 88). Growth rate was monitored by OD (750nm) and no significant differences in growth were observed between transformed and wild type strains (Figure 87).

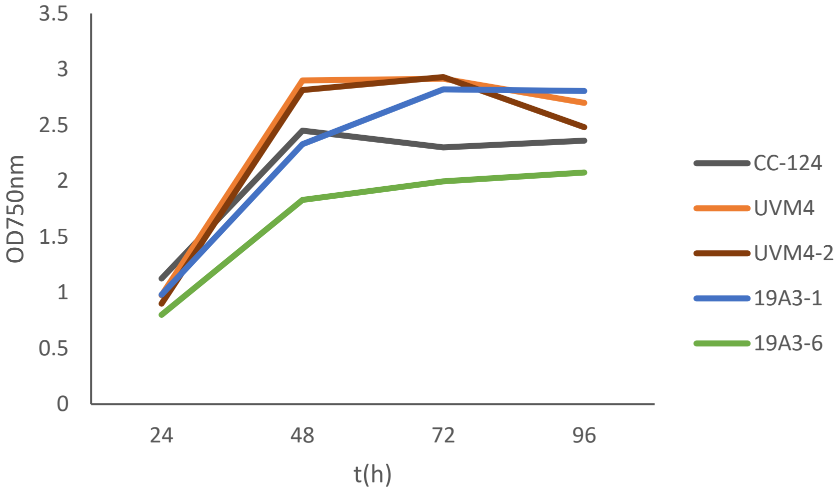
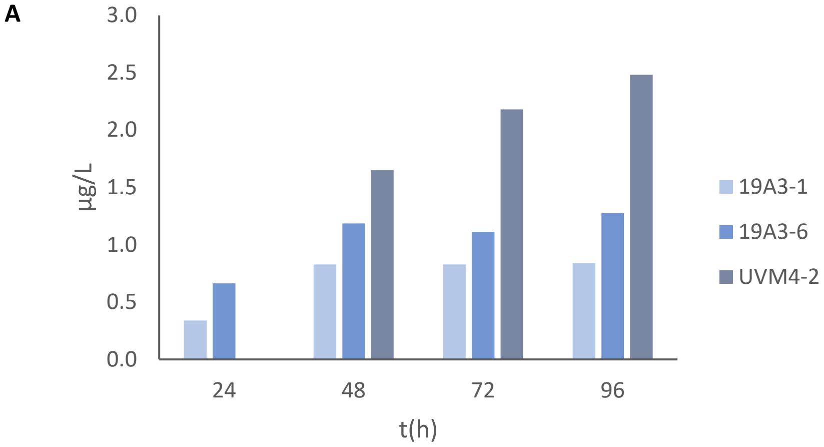


**Figure 87. Kinetics of growth in transformants expressing m5C3.** Growth was monitored by optical density (OD) at 750nm.



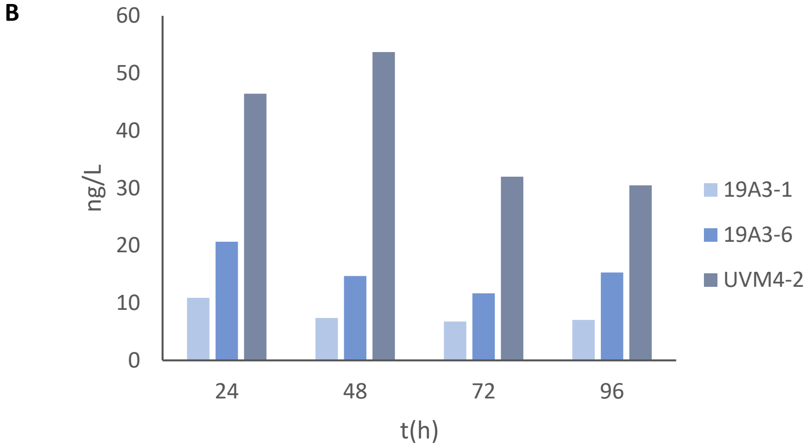


**Figure 88. Kinetics of mAb expression analysed by sandwich ELISA in selected transformants. A)** Quantification of extracellular mAb **B)** Quantification of intracellular mAb.

Corroborating the results obtained with 19A3 transformant (figure 77, page 197), whereas there is accumulation of extracellular mAb at latter stages of culture for all the selected transformants, the highest levels of intracellular mAb are reached at early stages of growth. Values of quantification of extracellular and intracellular mAb at different stage of cell culture for each transformant are presented in table 16.

**Table 16. Sandwich ELISA quantification of mAb expression in selected transformants at different culture stages.** Results in µg/L. I: intracellular mAb, E: extracellular mAb, ND; not detected.

| | 24h | 48h | 72h | 96h |
|---|---|---|---|---|
| **19A3-1 E** | 0.339 | 0.829 | 0.827 | 0.839 |
| **19A3-6 E** | 0.665 | 1.185 | 1.113 | 1.276 |
| **UVM4-2 E** | ND | 1.649 | 2.181 | 2.482 |
| **19A3-1 I** | 0.011 | 0.007 | 0.007 | 0.007 |
| **19A3-6 I** | 0.021 | 0.015 | 0.012 | 0.015 |
| **UVM4-2 I** | 0.046 | 0.054 | 0.032 | 0.030 |

An immunoblot was also performed to evaluate mAb expression at each growth stage. This immunoblot was performed under reducing and non-reducing conditions with an anti-mouse F(ab')2 region. Fully assembled mAb is detected at the different stages of the culture without variations in its expression. Moreover, there is accumulation of mAb in medium at the late exponential phase of growth that corresponds exclusively to the fully-assembled mAb without detectable degradation bands. Moreover, as it can be observed in the immunoblot under non-reducing conditions, there is more fully-assembled mAb in transformant 19A3-6 and 19A3-1 than in transformant UVM4-2. That is not consistent with ELISA quantification results which indicated that transformant UVM4-2 was the highest expressing transformant. However, since transformant UVM4-2 expresses an excess of LC, this excess of LC may be unspecifically recognised by sandwich ELISA. Furthermore, fully assembled mAb was estimated based on immunoblots by densitometry and higher expression of transformant 19A3-6 relative to transformant 19A3-1 was corroborated with an expression level of approximately 1 µg/L consistent with sandwich ELISA results. Considering both sandwich ELISA and immunoblot results transformant 19A3-6 was selected for further mAb characterization.

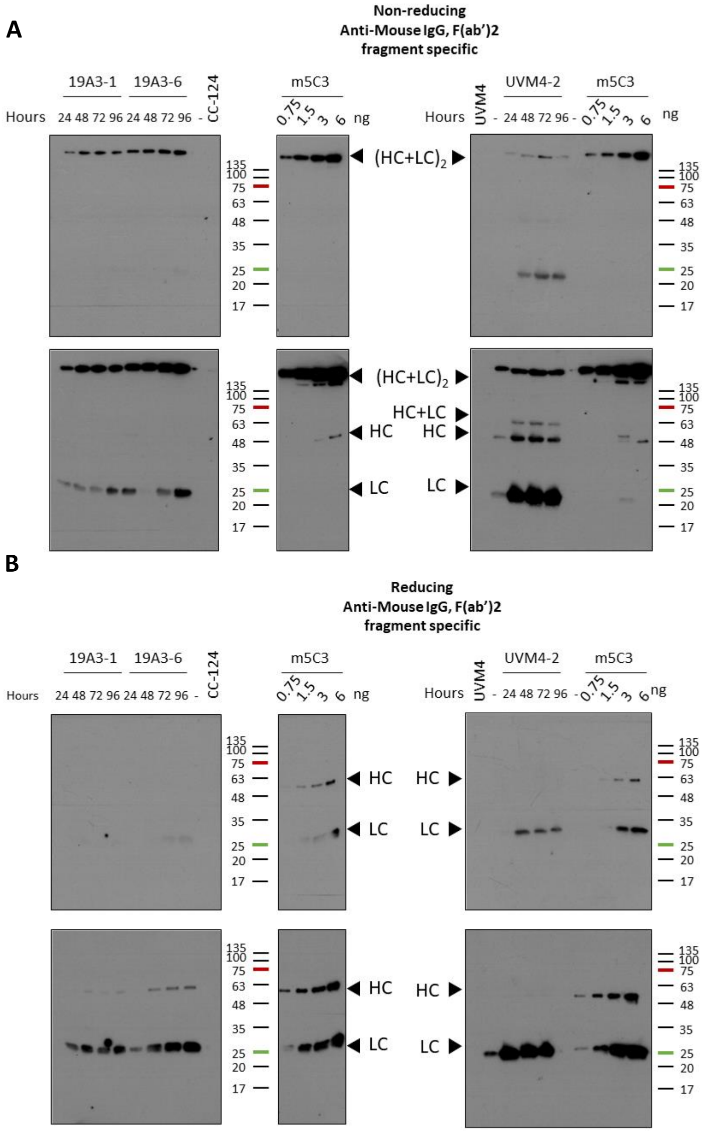

A
Non-reducing
Anti-Mouse IgG, F(ab')2
fragment specific
19A3-1
19A3-6
CC-124
Hours
24 48 72 96 24 48 72 96 -
m5C3
0.75 1.5 3 6 ng
UVM4
UVM4-2
m5C3
Hours
- 24 48 72 96 - 0.75 1.5 3 6 ng
(HC+LC)2
HC+LC
HC
LC
135
100
75
63
48
35
25
20
17
B
Reducing
Anti-Mouse IgG, F(ab')2
fragment specific

**Figure 89. Immunoblot evaluation of mAb expression in selected transformants at different culture stages.** Primary antibody anti-mouse IgG, F(ab')2 fragment specific. 12 µl of 80X concentrated culture supernatant were loaded in each lane. Two exposures are shown for each immunoblot: the one showed above corresponds to seconds of exposure and the one showed below corresponds to minutes of exposition. m5C3: purified monoclonal 5C3 antibody expressed in mammalian cells, $(HC+LC)_2$: complete mAb, HC: heavy chain, LC: light chain. **A)** Immunoblot under non-reducing conditions. **B)** Immunoblot under reducing conditions.

## *3. In vitro mAb characterization*

To further characterize the expressed mAb, several culture batches of 20 L were performed. Culture supernatant from these cultures was concentrated by TFF (30 kDa) to achieve higher concentrations of mAb to allow *in vitro* characterization assays.

### 3.1 mAb evaluation by immunoblot in reducing and non-reducing conditions

Firstly, mAb concentrated by TFF was evaluated by immunoblot to corroborate the integrity of the mAb after the concentration process.

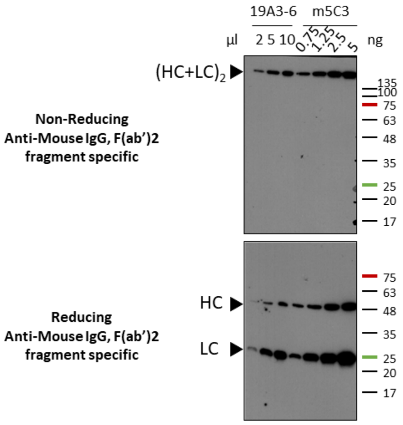

19A3-6
m5C3
µl
2 5 10
0.75
1.25
2.5
5
ng
(HC+LC)2
Non-Reducing
Anti-Mouse IgG, F(ab')2
fragment specific
135
100
75
63
48
35
25
20
17
Reducing
Anti-Mouse IgG, F(ab')2
fragment specific
HC
LC
75
63
48
35
25
20
17

**Figure 90. Immunoblot evaluation of concentrated mAb by TFF.** Upper panel corresponds to immunoblot under non-reducing conditions and lower panel corresponds to immunoblot under reducing conditions. Primary antibody anti-mouse IgG, F(ab')2 fragment specific. Several volumes of concentrated culture supernatant from transformant 19A3-6 were loaded. Concentration was performed by TFF from 30L to 250 ml (120X). m5C3: purified monoclonal 5C3 antibody expressed in mammalian cells, $(HC+LC)_2$: complete mAb, HC: heavy chain, LC: light chain.

## 3.2 Evaluation of recognition of Ag by ELISA

To determine whether the m5C3 antibody expressed in Chlamydomonas is able of binding its target antigen S100A4, ELISA assays were performed with the non-purified concentrated culture supernatant containing recombinant 5C3 and compared to purified 5C3 from hybridoma cells. Interestingly, concentrated culture supernatant containing Chlamydomonas expressed 5C3 was able to bind to S1004.

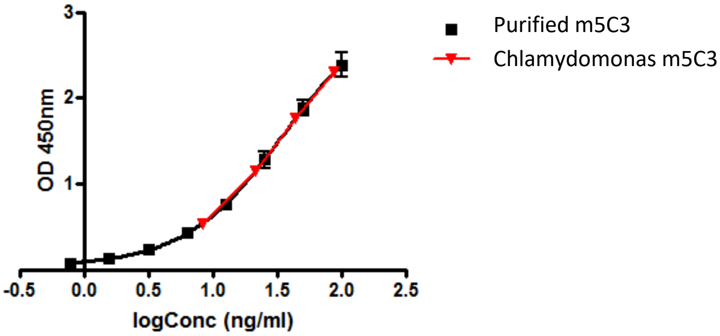


**Figure 91. Analysis of human S100A4 recognition by ELISA.** m5C3: m5C3 expressed in hybridoma cells. Chlamydomonas m5C3: concentrated culture supernatant containing m5C3 expressed in Chlamydomonas.

# Discussion

In the present chapter we report the expression of a fully-assembled mAb in Chlamydomonas nucleus, to our knowledge not previously described. Although the expression of a fully-assembled mAb has been reported in Chlamydomonas chloroplast (Tran et al. 2009), this organelle does not glycosylate and thus the biological activity of mAbs expressed in chloroplast would always be limited. The expression of a fully-assembled mAb in Chlamydomonas nucleus means a step forward for Chlamydomonas as a biotechnological platform.

Expression of a fully assembled monoclonal antibody (mAb) requires the expression and assembly of the heavy chain (HC) and the light chain (LC). Since free HC are not exported (Morrison and Scharff 1975) it is particularly important to express both proteins simultaneously. Several strategies may be employed for the expression of multiple recombinant genes: the use of multiple vectors with diverse selection markers, the use of a single vector with a polycistronic mRNA or the use of a single vector producing multiple RNAs. However, the use of multiple vectors results in different expression levels because of random genomic insertion, making necessary the performance of large screenings due to the low probability of having both genes expressed at the same level. Even though expression of separate proteins from dicistronic genes has been achieved with viral FMDV 2A sequences (Rasala et al. 2012), the efficiency of processing becomes a limitation as seen in the present work (*Chapter 1*) and reported (Rasala et al. 2012; Kong et al. 2015; López-Paz et al. 2017). Although expression of two proteins from dicistronic mRNAs containing a non-structured junction sequence between the two ORFs has been described, the expression of the downstream gene resulted lower than the expression of the upstream gene (Onishi and Pringle 2016). Finally, the use of a single vector involving several copies of a strong regulatory region introduces the risk of silencing (Ceruttti et al. 1997; Schubert 2004). Our expression strategy is based on a single microalgae expression vector containing three different regulatory regions (5’ and 3’) to co-express three different genes simultaneously and at high levels from the same vector: a resistance to antibiotic and two target proteins. The expression strategy developed in the present work emerges as a tool, not only for the expression of a fully assembled mAb but for the expression of protein complexes or components of metabolic pathways that need to be expressed simultaneously at high levels. Co-expression of subunits is an important requirement in the study and production of protein complexes. When subunits are expressed separately they may not be soluble, active or stable, and therefore a system to co-express proteins in a stoichiometric manner is a highly desirable tool. In addition, it may be useful in metabolic engineering where the objective is to obtain a modified strain overexpressing different genes with a determined objective (e.g. higher growth or the production of a specific metabolite).

Since our objective was to select a transformant expressing a fully assembled mAb, we needed a high-throughput screening leading to transformants expressing both chains simultaneously. We developed a high-throughput screening based on sandwich ELISA (Enzyme-Linked Immunosorbent Assay) after an initial selection of transformants resistant to antibiotic (indicating the insertion of heterologous DNA cassette). Sandwich ELISA was designed to be performed with total extracts from colonies grown in 96-well plates, lysed by a freeze-thaw cycle. This screening allowed us to process a high number of transformants to detect those expressing both chains assembled as a mAb. However, since the secondary antibody used in the sandwich ELISA detects the F(ab’)2 region (present in both chains), there is the possibility that the assay detects transformants expressing only one chain. Due to that, sandwich ELISA results have been always corroborated by immunoblot.

Analysis of the expression of mAb in Chlamydomonas selected transformants by sandwich ELISA and immunoblot showed that mAb accumulates in culture medium. In addition, it has been shown by immunoblot that this accumulation in medium corresponds exclusively to the fully-assembled mAb without detectable degradation bands. Furthermore, mAb expression was quantified by immunoblot in selected transformants and the highest expressing transformant of the fully-assembled mAb was 19A3-6, with an expression level of 1 µg/L. However, sandwich ELISA results indicated that UVM4-2 was the highest expressing transformant. It has been also seen by immunoblot that this transformant expresses an excess of LC. Thus, this excess of LC may be unspecifically recognised by ELISA sandwich explaining the inconsistency between sandwich ELISA and immunoblot results. It is interesting to highlight that UVM4 resulted in higher expression levels than CC-124 since the latter needed to be transformed with an additional copy of HC to express both chains to detectable levels.

Analysis of mAb expression in Chlamydomonas transformants showed that LC is overexpressed relative to HC, presumably due to poor expression or instability of this chain compared to LC. In fact, the initial transformant isolated from CC-124 strain needed to be transformed with an additional copy of HC because expression of HC was undetectable. The retransformation of this transformant with a second copy of HC increased the expression of both HC and fully-assembled mAb. It has been reported that LC:HC ratio plays an important role in mAb expression and quality: having LC in excess is not only beneficial for mAb expression, it also keeps aggregate and fragment levels low (Ho et al. 2013). Since an excess of LC is found in our isolated transformants, it would be interesting to further study the accumulation of HC gene copies to determine the LC:HC optimal ratio that maximize mAb expression in Chlamydomonas.

It has been shown that the use of alternative signal peptides affects the efficiency of protein secretion (Knappskog et al. 2007; Kober et al. 2013; Molino et al. 2018) and thereby signal peptide selection affects the overall yield of secreted proteins in an expression system. An study in CHO cells showed that when two signal peptides are linked to different antibodies the impact of the signal peptides on production of the antibodies can be very different (Haryadi et al. 2015). In the same study, by optimizing HC and LC signal peptides, a 2-fold increase in antibody yield was achieved compared to the original signal peptides with Rituxan (one of the top selling antibody therapeutics). These results demonstrate the importance of signal peptide optimization for the production of recombinant antibodies and suggest that, in addition to the signal peptide, a portion of the variable region of the antibody also affects the secretion efficiency (Haryadi et al. 2015). In the present work the native murine secretion peptide from mAb was maintained and it successfully drove secretion reinforcing the idea of the existence of a conserved secretion mechanism (Pool 2005). Optimization of signal peptides on mAb expression in Chlamydomonas is an interesting further step, for example testing the effect of Chlamydomonas native signal peptides.

We demonstrate the expression of a fully-assembled mAb in Chlamydomonas nucleus that recognises its antigen. It would be interesting to test if the Fc part of the antibody is also functional by testing its binding to receptors. Differences in glycosylation may cause conformational changes of the Fc domain, which could change its binding affinity to Fc receptors resulting in changes of immune effector functions (Tao and Morrison 1989; Liu 2015). Preliminary results obtained have shown no interaction between receptors and Chlamydomonas expressed mAb (results not shown). However, the assays were performed with the concentrated culture supernatant that may contain interfering components. Thus, these assays should be performed with purified m5C3. Furthermore, the characterization of the glycosylation of the expressed mAb is a key next step. Analysis of glycosylation is important because of the risk of immunogenicity and the impact of glycosylation patterns on pharmacokinetics (Goetze et al. 2011).

Furthermore, it would be interesting to expand the potential of the host by expressing a human or humanized mAb model since the majority of the mAb in clinical development are human or humanized.

It is important to highlight the concentration up to 1000-fold of culture supernatant from Chlamydomonas achieved by tangential flow filtration (TFF) because indicates an easy downstream processing of Chlamydomonas secreted proteins. Although expression levels achieved in plants are higher (Zeitlin et al. 2013) the downstream processing is costly because implies cells grinding. Thus, the ease of downstream of secreted proteins in Chlamydomonas may mean the difference between Chlamydomonas and plants as a recombinant protein host together with the ease of culture in closed bioreactors, no risk of crops displacement and no risk of gene flow.

# Methodology

### *Chlamydomonas strains and culture conditions*

Algal cells were cultivated under standard conditions as previously stated (see *Chapter 1, Methodology*, page 129).

### *Construction of expression cassettes*

All plasmids were built on the pGEM7Z(-) backbone. Three cis regulatory regions were used in this study to drive expression of GOI. On one hand, *HSP70A/RBCS2* chimeric promoter (Schroda et al., 2000) further improved by insertion of the first intron of *RBCS2* downstream (Sizova et al., 2001), abbreviated as AR in the present study, was used in combination with 3'UTR of *RBCS2*. On the other hand, *RPL23* (Cre04.g211800) cis regulatory elements (5' and 3') and ferredoxin 1 (*FDX1*, also called PETF) (Cre14.g626700) cis regulatory regions (5' and 3') were used (López-Paz et al. 2017).

The sequences coding for m5C3 were provided by Leitat Biomed (Patent WO2011157724 A1). The sequences of HC and LC of 5C3 mAb were codon-optimized with the IDT web tool (https://eu.idtdna.com/CodonOpt) for Chlamydomonas nuclear expression and synthesized by GeneArt (Life Technologies, USA), murine secretion sequence was maintained. HC was synthesized with restriction sites EcoRI/XhoI. LC was synthesized with restriction sites EcoRI/NdeI. *FDX1* regulatory regions (Cre14.g626700) were amplified from CC-124 strain (5' region - F421/R422 and 3' region - F423/R424) and cloned (5' - KpnI/BamHI, 3' – BamHI/SacI) into vector pARshBle2A (containing AR chimeric promoter (XbaI/EcoRI) fused to Ble$^R$ gene and 3'UTR *RBCS2* (NdeI/KpnI) in pGEM7Z(-), see *methodology chapter 1, table 12, page 131*) to obtain vector pFDX-incomplete. Then, HC was cloned into vector pFDX-incomplete (EcoRI/XhoI) obtaining the vector pFDX-HC-zeo$^R$. *RPL23* regulatory regions (Cre04.g211800) were amplified from CC-124 strain (5’ region - F409/R410 and 3’ region - F411/R412) and cloned (5’- XbaI/EcoRI and 3’- NdeI/KpnI) in pAR-ARSss-glucTEVhEGF (see *methodology chapter 1, table 12, page 130*) replacing AR chimeric promoter for 5’ region of *RPL23* and 3’UTR *RBCS2* for 3’ region of *RPL23* to obtain a vector named pPL-incomplete. Then, LC was cloned into vector pPL-incomplete (EcoRI/NdeI) obtaining vector pPL-LC. Finally, the LC gene driven by *RPL23* regulatory regions from vector pPL-LC (XbaI/SmaI, blunted) was cloned XbaI (blunted) into pFDX-HC-zeo$^R$ obtaining the vector FM-zeo$^R$. Vector FM-zeo$^R$ carries three cassettes: one for the expression of an antibiotic resistance for initial selection of transformants (Ble$^R$) driven by AR chimeric promoter and 3’UTR *RBCS2*, one for the expression of the heavy chain (HC) driven by *FDX1* regulatory regions and one for the expression of the light chain (LC) driven by *RPL23* regulatory regions.

The 5’ cis regulatory region of *RPL23* (1004 bp) contains the predicted promoter and 5’ UTR that is predicted to contain an intron. For the *RPL23* 3’ cis regulatory region the 3’UTR of *RPL23* is also predicted to contain an intron and adjacent DNA is used (total of 725 bp).

To obtain the vector HC-paro$^R$, gene *AphVIII* (resistance to paromomycin) driven by AR chimeric promoter and 3’UTR *RCBS2* was cloned from vector psi103 (https://www.chlamycollection.org/product/psi103-aphviii/ - XbaI/KpnI) replacing shBle$^R$ gene (resitance to zeocin) and FDX regulatory regions into pFDX-HC-zeo$^R$.

**Table 17. Sequences of oligonucleotides used for vectors construction in this study.**

| Oligonucleotides names | Oligonucleotides sequence (5’-3’) |
|---|---|
| F409 – 5’ RPL23 F | GGCCTCTAGAGCACGTCCATAATGAAAGGTC |
| R410- 5’ RPL23 R | CGGAAGAATTCACTCTGCGCGCAGACAAGAG |

| Oligonucleotides names | Oligonucleotides sequence (5’-3’) |
|---|---|
| F411- 3’UTR RPL23 F | GGGCATATGTAAAGGTGAGGCGGGGGTTCG |
| R412- 3’UTR RPL23 R | CCAGGTACCCCAGGCGCAGCTTCGCC |
| F421 – 5’ FDX1 F | TTAGGTACCTGGAGGGGTAATCCGTGGG |
| R422 - 5’ FDX1 R | AAAGGATCCGCCAGATCTGCCGAATTCTTTTGAAGGTTGTGAGTGG |
| F423 - 3’UTR FDX1 F | TTCGGGATCCGGCCTCGAGGCGCTTTCGCCATCTGC |
| R424 - 3’UTR FDX1 R | AAAGAGCTCCTCTCCTCAGAACGCGGAG |

### *Protein isolation and Immunoblot analyses*

Culture supernatant was separated from cells as indicated previously and supernatant was concentrated by centrifugation at 12000 gs through a membrane of 10 kDa (Vivaspin 500, 10 kDa MWCO PES, VS0101, Sartorius) unless otherwise stated. Immunoblots under reducing conditions were performed as previously stated (see *Chapter 1, Methodology*, page 140). Immunoblots under non-reducing conditions were performed following the same protocol but using laemmli loading buffer without reducing agent (120 mM tris pH 6.8, 4% SDS, 20% glycerol, bromophenol blue). The antibodies used were: Peroxidase AffiniPure Goat Anti-Mouse IgG, F(ab’)2 fragment specific (115-035- 006, Jackson Immunoresearch, 1:10 000) or goat anti-mouse IgG HRP conjugated (926-80010, LiCor, 1:25 000).

Purified m5C3 from hybridoma cells (kindly provided by Leitat Biomed, obtained as reported (Hernández et al. 2013)) was used as a reference for quantification by densitometry using Fiji software.

### *Sandwich ELISA*

Greiner CELLSTAR® 96 well plates (M0187, Sigma) were coated with 1 µg/mL of Goat Anti-Mouse IgG, Fc-γ Fragment Specific (115-005- 071, Jackson Immunoresearch) in PBS and incubated at 4°C, 16 h. After five washes with PBX (PBS containing 0.1% Tween-20, pH 7.5), 100 µl of sample (total cell extract or culture supernatant, indicated in each figure) were added and the plate was incubated at 37°C for 1 h. After five washes with PBX, the Peroxidase AffiniPure Goat Anti-Mouse IgG, F(ab')2 fragment specific (115-035-006, Jackson Immunoresearch, 1:20 000) was added and incubated at 37°C 1h. After five washes (100 µl PBX) 50 µl of Luminata Forte ELISA HRP Substrate (ELLUF0100, Merck) were added and luminescence was recorded by a 96-well microplate luminometer (Berthold LB 96V MicroLumat Plus Luminometer, Berthold technologies). Light emission was recorded for 1 s. Purified m5C3 from hybridoma cells (provided by Leitat Biomed (Hernández et al. 2013)) was used as standard curve.

Total extracts were used in initial Sandwich ELISA screening of transformants. To do so, initial clones (resistant to 15 µg/ml zeocin) were grown in 96 well plates to mid-log phase and frozen at -80ºC in presence of protein inhibitors. After thawing the cells at 37ºC for 10 min, 100 µl of total cell extract (including both intracellular content from lysed cells and culture supernatant) were analysed by sandwich ELISA previous separation of cell debris by centrifugation.

Culture supernatant was separated from cells by centrifugation in presence of 0.005% tween 20 (4000 g 10 min) when sandwich ELISA was performed only with supernatant.

### *Tangential flow filtration (TFF)*

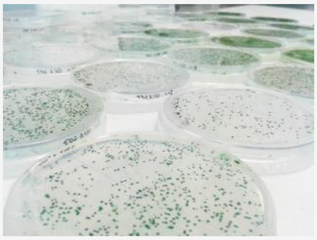
Between 30 and 40 L of culture supernatant from a culture at late exponential phase of growth (OD750nm approximately at 1.5) was separated from cells by centrifugation and 0.005% tween 20 was added. Culture medium was filtered through a 1 µm cut-off membrane (PE, filtres Peiró). Then, culture medium was concentrated by TFF with hollow fiber of 30 kDa (mPES Minikross Sampler, total length of 47.0 cm, effective length of 41.5 cm and surface of 1600 $cm^2$, Spectrum). The retentate, holding all particles bigger than 30 kDa, was continuously cycled in order to reduce its final volume to 200 ml. This concentration process needed 8 hours of TFF process. Once concentrated to 200 ml, concentrated culture supernatant was centrifuged 10000 g 10 min to eliminate precipitated particles and cell debris. 9 batches were concentrated 150X as explained and maintained at -20ºC. Once 9 batches had been concentrated as explained (approximately 270 L in 1800 ml) these 150X concentrated culture supernatant was concentrated to 200 ml with the same TFF system and hollow fiber (5 hours of process). Finally, this concentrated culture supernatant was centrifuged again to eliminate precipitated particles at 10000 g 10 min. Thus, at the end, 200 ml of a culture supernatant concentrated 1350X relative to initial culture volume were obtained.

### *Recognition of Ag by ELISA.*

This work was performed by the Biomed division of Leitat.

Corning® 96 Well EIA/RIA Assay Microplate (CLS3590, Sigma) were coated with 3 µg/mL of human S100A4 (expressed and purified in *E.coli*, (Hernández et al. 2013)) in PBS and incubated at 4°C, 16 h. After five washes with PBS several concentrations of purified m5C3 and dilutions of concentrated culture supernatant from Chlamydomonas (1/50, 1/100, 1/200, 1/400, 1/800) were added and the plate was incubated at 37°C for 1 h. After five washes with PBS, the Peroxidase AffiniPure Goat Anti-Mouse IgG, Fc-γ Fragment Specific (115-

035-071, Jackson Immunoresearch, 1:20 000) was added and incubated at room temperature for 45 min. After five washes (100 µl PBS), the ELISA was developed by the addition of Tetramethylbenzidine substrate (Sigma) followed by incubation for 30 min at RT before stopping the reaction with 1 M of HCl. Absorbance at 450 nm was recorded using a Multiskan Ascent spectrophotometer (Thermo Corporation). Using GraphPad prism 7.04 a standard curve was constructed by plotting absorbance values versus purified m5C3 known concentrations and unknown concentrations of concentrated Chlamydomonas culture were interpolated using this standard curve.



# Discussion and future prospects

The development of improved recombinant protein production hosts is a need to address costs and demands of a range of markets such as the cosmetic, industrial, research or therapeutic (Fletcher et al. 2007; Ferrer-Miralles et al. 2009). The development of RP expression hosts moves forward not to a unique advantageous host but to a variety of hosts with distinct characteristics that will be suitable depending on the target product and/or target application. Lowering the costs of RP production is desirable for several applications that cannot hold the costs of production in mammalian cells. Also, the biobetter concept is a key concept for the development of recombinant technologies: alternative recombinant hosts may improve the properties of products already marketed. An example is the first plant biopharmaceutical approved by the FDA in 2012 (Mor 2015). Together with the continuous improvement of current expression systems available, several alternative recombinant hosts have been proposed.

Microalgae are emerging as an alternative RP expression host, being Chlamydomonas the best characterized microalgal species. Despite the fact that high expression levels have been achieved in its chloroplast, there are applications for which PTMs are crucial (for example antibody glycosylation has a critical role in antibody effector function (Tao and Morrison 1989; Liu 2015)) and the chloroplast cannot perform PTM. Furthermore, RP expressed in the chloroplast remain in this organelle and cannot be secreted, involving the need for a complex extraction (Mayfield et al. 2007). Despite all the advantages of Chlamydomonas as RP expression host (*Introduction, page 12*) and the successful expression of RP in its chloroplast, the low expression levels in its nucleus represent a limitation to its use for RP production. This thesis aims to help in consolidating this host as a versatile RP expression platform by testing several strategies designed to reduce or overcome those limitations. The design of a secreted fusion protein comprised of gLuciferase as a reporter gene and hEGF as our GOI allowed the development of a high-throughput screening that enables the detection of the highest expressing clones. Versions of the fusion protein including different regulatory and stabilizing regions have been expressed in various Chlamydomonas strains resulting in expression levels ranging from 1 to 100 µg RP /L of culture. Furthermore, Chlamydomonas native stabilizing sequences have been studied and five were tested. One of this five stabilizing sequences resulted in a significant increase in RP yields although expression levels did not surpass previously described sequences from plants tested in Chlamydomonas. Several of the strategies assessed resulted in an increase of expression levels. It is expected that combination of these strategies will result in a further increase in recombinant protein expression levels. In addition, further research on the described native stabilizing region of Chlamydomonas would be a consequent next step of the present study.

A scalable purification process from culture medium has been developed consisting of a concentration step of culture medium for posterior IMAC and processing by TEV protease. This scalable purification process has been validated with different fusion proteins that were successfully recovered by affinity chromatography from Chlamydomonas culture medium and cleaved by TEV protease releasing discrete hEGF. However, the fusion protein leading to the highest expression levels reported in the present work (100 µg RP/L) could not be cleaved by TEV protease likely due to the proximity of the $(SP)_{20}$ glycomodule. Further studies to achieve the expression to these levels while allowing the cleavage of the fusion protein must be performed.

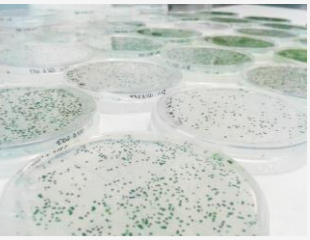
The next step regarding hEGF is to achieve high purity levels enabling the functional testing of processed hEGF. To that end, the optimization of a subsequent purification step is proposed. It is important to note that the low expression levels hamper the efficient purification of RP. To maximize RP yields, apart from optimizing RP expression itself, Chlamydomonas culture should be optimized. Chlamydomonas is usually cultured in batch with TAP medium, entering the stationary phase of growth without reaching high densities and resulting in 1-2 g/L of dry weight (Fields et al. 2018). Despite a few advances towards this end (Chen and Johns 1995; Fields et al. 2018), a comprehensive study to extend the growth of Chlamydomonas is lacking: exhausted nutrients, secondary metabolites inhibiting growth and medium optimization have to be analysed. Optimization of Chlamydomonas growth will enhance the host by maximizing its productivities and allowing more efficient purification processes.

A fully-assembled monoclonal antibody that recognises its antigen has been expressed, secreted and accumulated in Chlamydomonas culture medium. To our knowledge it has not been described previously and thus means a step forward for Chlamydomonas as a recombinant protein host. An expression strategy and a high throughput screening have been developed leading to the selection of transformants expressing a fully-assembled mAb. Our expression strategy is based on a single microalgae expression vector containing three different regulatory regions (5’ and 3’) that have previously drove transgene expression to high levels to co-express three different genes simultaneously from the same vector: a selectable marker and two target proteins. This expression strategy emerges as a tool not only for the expression of a fully assembled mAb but for the expression of other protein complexes where different genes must be constitutively expressed to high levels. In addition, it may be useful in metabolic engineering where the objective is to obtain a modified strain overexpressing different genes with a determined objective (e.g. higher growth or the production of a specific metabolite).

The achieved concentration of Chlamydomonas culture medium by TFF (up to 1000-fold) indicates an easier downstream processing of Chlamydomonas secreted proteins when compared to plants. Although expression levels in plants are higher (Zeitlin et al. 2013) the downstream processing is costly because implies cells grinding. Thus, the ease of downstream of secreted proteins in Chlamydomonas may represent the difference between Chlamydomonas and plants as a RP host together with the fact that is a more ecologically conscious alternative (no possibility of gene flow and no crops displacement).

During the course of this work, we unexpectedly found that secreted components from Chlamydomonas possess proliferative activity on fibroblast cells independently of the presence of recombinant protein. Furthermore, the induction of proliferation by these secreted components acts synergistically with hEGF. Extracellular polymeric substances (EPS) of other organisms have been shown to have high value in diverse applications: anti-adhesive materials, antitumoral agents, anti-inflammatory, antivirals and skin care products (Xiao and Zheng 2016). Due to that, it is expected that the extracellular content of Chlamydomonas holds an unexplored potential. Nowadays, microalgae products are mainly derived from their biomass. Thus, huge volumes of medium are wasted after biomass harvesting of batch cultures because medium recycling to save costs involves the risk of growth inhibition. Thereby, the waste of high volumes of medium involves environmental pollution and higher production costs. The generation of high value products from waste medium would be meaningful for environmental and commercial reasons and therefore the exploitation of Chlamydomonas culture medium becomes even more interesting.

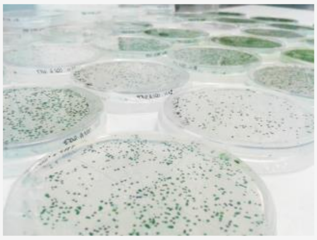
The work reported is a step forward for Chlamydomonas as a biotechnological platform. The cassettes and related high-throughput screenings developed expand the molecular toolbox available for Chlamydomonas nucleus RP expression. The developed strategies allowed the expression of a fusion protein comprised of a small growth factor and a luciferase reporter with yields ranging from 1 to 100 µg/L of culture. Furthermore, a tetrameric protein (a mAb consisting of two light chains and two heavy chains) has been correctly expressed, folded, secreted and accumulated in medium. In addition, the development and validation of a scalable purification method to obtain the processed target protein from culture medium confirms feasibility of a simple downstream processing of expressed and secreted RP that may represent the advantage in front of other expression hosts. The reported proliferative activity induced by Chlamydomonas secreted components as well as the synergistic effect seen when tested together with hEGF unveils the potential of secreted components of Chlamydomonas. The work reported validates Chlamydomonas as a versatile recombinant protein host.



# Conclusions

1. The strategy based on a fusion protein comprising shBle and hEGF led to the selection of transformants expressing a variety of RP products due to partial insertions of the cassette. Thus, this strategy could not be validated for the selection of high expressing clones of the proteins tested.
2. Clones expressing fusion proteins comprising Ble$^R$ gene duplicated their expression level through the maintenance in presence of zeocin for two months. This will enhance expression of transgenes for which the strategy shBle:GOI is suitable.
3. The strategy based on a fusion protein consisting of *Gaussia princeps* luciferase, TEV protease recognition sequence and hEGF enables the high throughput selection of the highest expressing transformants of the full-length fusion protein. The use of gLuc:TEV:hEGF allows the expression, secretion and accumulation of the fusion protein in medium. Stage of culture corresponding to maximum expression is protein dependent.
4. The use of different regulatory regions, secretion peptides and stabilizing regions resulted in RP (gLuc:hEGF) expression yields ranging from 1 to 100 µg/L.
5. We have identified a Chlamydomonas glycomodule sequence that results in higher expression levels compared to fusion protein without glycomodule.
6. Although UVM4 may be a preferred strain due to its increased RP expression, the observed higher degradation of RP and higher viscosity of its concentrated culture supernatant may limit its use at the industrial scale.
7. Preliminary results suggest that increased transgene copy number in haploid strains correlate with higher RP protein production. No differences have been observed in RP expression in diploid cells relative to haploid cells.

8. Two candidate strains for improved nuclear RP expression have been isolated by means of a UV mutagenesis screening.
9. We have developed and validated an optimized scalable purification process from culture medium consisting of a concentration step for posterior IMAC and processing by TEV protease.
10. Components present in culture supernatant from Chlamydomonas at stationary phase of growth induce proliferation of murine fibroblasts. The proliferation induced by Chlamydomonas secreted components acts synergistically with recombinant hEGF.
11. We have developed an expression strategy and a high-throughput screening that allows the selection of transformants expressing a fully assembled monoclonal antibody.
12. A fully-assembled and functional monoclonal antibody has been expressed from Chlamydomonas nucleus, secreted and accumulated in culture medium.
13. The achieved 1000-fold concentration of Chlamydomonas culture supernatant by TFF indicates that a simple downstream processing of expressed and secreted RP is feasible.



# Appendixes

# Appendix 1

## Recombinant protein expression and purification from E.coli to allow quantification based on densitometry of immunoblots

Since expression levels from Chlamydomonas nucleus were expected to be in the range of µg/L, densitometry based on immunoblot was chosen for quantification of the protein of interest. To that end, recombinant gLuc:3HA and shBle:3HA:6xhis were expressed in *E.coli* and purified. Plasmids pET22b-Ble-6xhis-3HA, pET22b-3FLAG-Ble-6xhis-3HA and pET22b-gLuc-3HA were transformed into *E.coli* strain Rosetta™ 2(DE3)pLysS (see table 18 for vector details). A pool of colonies resistant to carbenicillin (and thus carrying and expressing the heterologous vector transformed) were grown and expression was induced. Soluble protein was extracted and purified by IMAC in a Äkta system (ÄKTApurifier, GE) with $Co^{2+}$ columns (28-9537-66 HiTrap TALON Crude, GE Healthcare). Quantification of the purified protein was made by Bradford and purity was checked by Coomassie Brilliant Blue stained SDS page (figure 92). These purified proteins were used to make standard curves for quantification by densitometry of immunoblots throughout the present study.

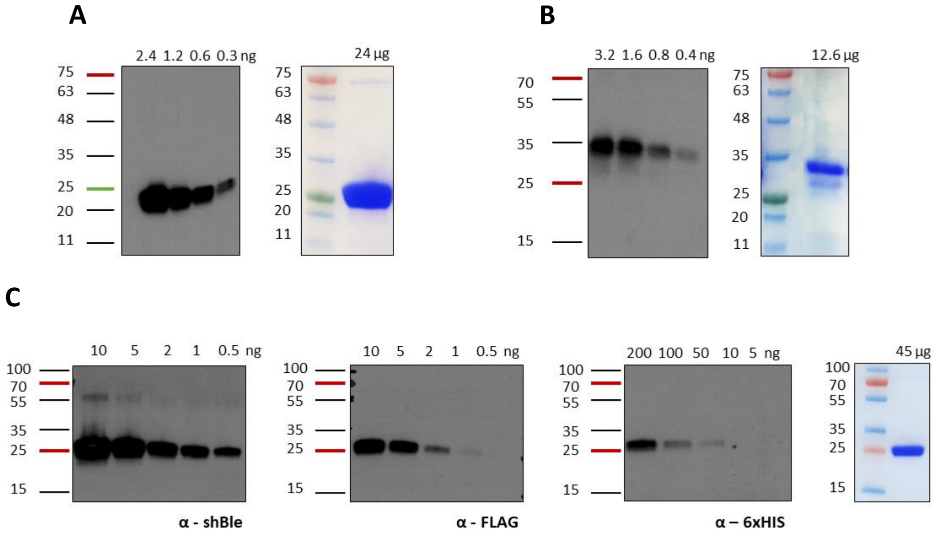
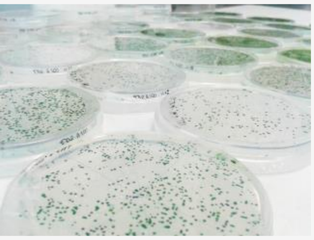

A
2.4 1.2 0.6 0.3 ng
75
63
48
35
25
20
11
24 µg
75
63
48
35
25
20
11
B
3.2 1.6 0.8 0.4 ng
70
55
35
25
15
12.6 µg
75
63
48
35
25
20
11
C
10 5 2 1 0.5 ng
100
70
55
35
25
15
α - shBle
10 5 2 1 0.5 ng
100
70
55
35
25
15
α - FLAG
200 100 50 10 5 ng
100
70
55
35
25
15
α – 6xHIS
45 µg
100
70
55
35
25
15

**Figure 92. Immunoblot and coomassie stained SDS-page of purified shBle:3HA:6xhis, gLuc:3HA and 3FLAG:shBLE:3HA:6xhis proteins expressed in *E. coli*. A)** purified shBle:3HA:6xhis. Left panel: immunoblot against shBle protein. Right panel: Coomassie stained SDS-page of purified protein**. B)** Purified gLuc:3HA. Left panel: immunoblot against gLuc protein. Right panel: Coomassie stained SDS-page of purified protein**. C)** 3FLAG:shBLE:3HA:6xhis**.** Left panel: immunoblot against shBle protein, against FLAG tag and against 6his tag. Right panel: Coomassie stained SDS-page of purified protein**.**

**Table 18. Detail of vectors for expression in *E.coli* used in the present study.**

| Vector | Content | Cloning |
|---|---|---|
| **pET22b-Ble-6xhis-3HA** | Ble:6xhis:3HA | Ble (EcoRI/HindIII) and 3HA (HindIII/XhoI) introduced to pET22b |
| **pET22b-3FLAG-Ble-6xhis-3HA** | 3FLAG:Ble:6xhis:3HA | FLAG introduced to pET22b-Ble-6xhis-3HA (EcoRI/NcoI) |
| **pET22b-gLuc-3HA** | gLuc:3HA | gluc introduced to pET22 (EcoRI/XhoI) |



# Bibliography

# Bibliography

Adam M, Murali B, Glenn NO, Potter SS (2008) Epigenetic inheritance based evolution of antibiotic resistance in bacteria. BMC Evol Biol 8:52

Ahmad ZA, Yeap SK, Ali AM, et al (2012) ScFv antibody: Principles and clinical application. Clin Dev Immunol 2012:

Ahmadzadeh V, Farajnia S, Feizi MAH, Nejad RAK (2014) Antibody Humanization Methods for Development of Therapeutic Applications. Monoclon Antib Immunodiagn Immunother 33:67–73

Allen GC, Hall G, Michalowski S, et al (1996) High-Level Transgene Expression in Plant Cells : Effects of a Strong Scaffold Attachment Region from Tobacco. Plant Cell 8:899–913

Allen GC, Hall GE, Childqb LC, et al (1993) Scaffold Attachment Regions lncrease Reporter Gene Expression in Stably Transformed Plant Cells. Plant Cell 5:603–613

Almo SC, Love JD (2014) Better and faster: Improvements and optimization for mammalian recombinant protein production. Curr. Opin. Struct. Biol. 26:39–43

Aw R, Polizzi KM, Bill RM, et al (2012) Can too many copies spoil the broth? Microb Cell Factories 2013 121 12:11–23

Baba M, Suzuki I, Shiraiwa Y (2011) Proteomic Analysis of High-CO2-Inducible Extracellular Proteins in the Unicellular Green Alga, Chlamydomonas reinhardtii. Plant Cell Physiol 52:1302–1314

Babcock J, Wilcox C, Huttinga H (2012) Partial Replacement of Chemically Defined CHO Media with Plant-Derived Protein Hydrolysates. BioPharm Int Suppl 23:1–6

Ballez JS, Mols J, Burteau C, et al (2004) Plant protein hydrolysates support CHO-320 cells proliferation and recombinant IFN-γ production in suspension and inside microcarriers in protein-free media. Cytotechnology 44:103–114

Bañuelos-Hernández B, Monreal-Escalante E, González-Ortega O, et al (2017) Algevir: An expression system for microalgae based on viral vectors. Front Microbiol 8:

Barahimipour R, Neupert J, Bock R (2016) Efficient expression of nuclear transgenes in the green alga Chlamydomonas: synthesis of an HIV antigen and development of a new selectable marker. Plant Mol Biol 90:403–418

Barahimipour R, Strenkert D, Neupert J, et al (2015) Dissecting the contributions of GC content and codon usage to gene expression in the model alga Chlamydomonas reinhardtii. Plant J 84:704–717

Barnes LM, Bentley CM, Dickson AJ (2003) Stability of protein production from recombinant mammalian cells. Biotechnol Bioeng 81:631–639

Barrera DJ, Rosenberg JN, Chiu JG, et al (2015) Algal chloroplast produced camelid VHH antitoxins are capable of neutralizing botulinum neurotoxin. Plant Biotechnol J 13:117–124

Bayne A-C V., Boltz D, Owen C, et al (2013) Vaccination against Influenza with Recombinant Hemagglutinin Expressed by Schizochytrium sp. Confers Protective Immunity. PLoS One 8:e61790

Beltrán-López JI, Romero-Maldonado A, Monreal-Escalante E, et al (2016) Chlamydomonas reinhardtii chloroplasts express an orally immunogenic protein targeting the p210 epitope implicated in atherosclerosis immunotherapies. Plant Cell Rep 35:1133–1141

Benvenuto E, Ordàs RJ, Tavazza R, et al (1991) “Phytoantibodies”: a general vector for the expression of immunoglobulin domains in transgenic plants. Plant Mol Biol 17:865–874

Berthold P, Schmitt R, Mages W (2002) An engineered Streptomyces hygroscopicus aph 7" gene mediates dominant resistance against hygromycin B in Chlamydomonas reinhardtii. Protist 153:401–12

Bertran K, Thomas C, Guo X, et al (2015) Expression of H5 hemagglutinin vaccine antigen in common duckweed (Lemna minor) protects against H5N1 high pathogenicity avian influenza virus challenge in immunized chickens. Vaccine 33:3456–3462

Brown AJ, Sweeney B, Mainwaring DO, James DC (2014) Synthetic promoters for CHO cell engineering. Biotechnol Bioeng 111:1638–1647

Brown LE, Sprecher SL, Keller LR (1991) Introduction of exogenous DNA into Chlamydomonas reinhardtii by electroporation. Mol Cell Biol 11:2328–32

Butler M (2005) Animal cell cultures: Recent achievements and perspectives in the production of biopharmaceuticals. Appl Microbiol Biotechnol 68:283–291

Capon DJ, Chamow SM, Mordenti J, et al (1989) Designing CD4 immunoadhesins for AIDS therapy. Nature 337:525–531

Carnemolla B, Borsi L, Balza E, et al (2013) Enhancement of the antitumor properties of interleukin-2 by its targeted delivery to the tumor blood vessel extracellular matrix Enhancement of the antitumor properties of interleukin-2 by its targeted delivery to the tumor blood vessel extracellular matr. Blood 99:1659–1665

Casas-Mollano JA, Jeong B-R, Xu J, et al (2008) The MUT9p kinase phosphorylates histone H3 threonine 3 and is necessary for heritable epigenetic silencing in Chlamydomonas. Proc Natl Acad Sci U S A 105:6486–6491

Casas-Mollano JA, van Dijk K, Eisenhart J, Cerutti H (2007) SET3p monomethylates histone H3 on lysine 9 and is required for the silencing of tandemly repeated transgenes in Chlamydomonas. Nucleic Acids Res 35:939–950

Cerutti H, Johnson AM, Gillham NW, Boynton JE (1997) A eubacterial gene conferring spectinomycin resistance on Chlamydomonas reinhardtii: Integration into the nuclear genome and gene expression. Genetics 145:97–110

Ceruttti H, Johnson AM, Gillham NW, Boynton JE (1997) Epigenetic Silencing of a Foreign Gene in Nuclear Transformants of Chlamydomonas. Plant Cell Online 9:925–945

Chen CY, Kao AL, Tsai ZC, et al (2017) Expression of Synthetic Phytoene Synthase Gene to Enhance β-Carotene Production in Scenedesmus sp. CPC2. Biotechnol J 12:1–7

Chen F, Johns MR (1995) A strategy for high cell density culture of heterotrophic microalgae with inhibitory substrates. J Appl Phycol 7:43–46

Chen X, Zaro JL, Shen W-C (2013) Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev 65:1357–69

Chen Y, Wang Y, Sun Y, et al (2001) Highly efficient expression of rabbit neutrophil peptide-1 gene in Chlorella ellipsoidea cells. Curr Genet 39:365–370

Chronopoulou E, Harper JC (2015) IVF culture media: Past, present and future. Hum Reprod Update 21:39–55

Cohen S (1962) Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. J Biol Chem 237:1555–1562

Collén A, Saloheimo M, Bailey M, et al (2005) Protein production and induction of the unfolded protein response in Trichoderma reesei strain Rut-C30 and its transformant expressing endoglucanase I with a hydrophobic tag. Biotechnol Bioeng 89:335–344

Corchero JL, Gasser B, Resina D, et al (2013) Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. Biotechnol. Adv. 31:140–153

Cox KM, Sterling JD, Regan JT, et al (2006) Glycan optimization of a human monoclonal antibody in the aquatic plant Lemna minor. Nat Biotechnol 24:1591–1597

Cox MMJ (2012) Recombinant protein vaccines produced in insect cells. Vaccine 30:1759–1766

Cross FR, Umen JG (2015) The Chlamydomonas cell cycle. Plant J 82:370–392

Czajkowsky DM, Hu J, Shao Z, Pleass RJ (2012) Fc-fusion proteins: New developments and future perspectives. EMBO Mol. Med. 4:1015–1028

Danielli R, Patuzzo R, Ruffini PA, et al (2015) Armed antibodies for cancer treatment: A promising tool in a changing era. Cancer Immunol. Immunother. 64:113–121

David F, Steinwand M, Hust M, et al (2011) Antibody production in Bacillus megaterium: Strategies and physiological implications of scaling from microtiter plates to industrial bioreactors. Biotechnol J 6:1516–1531

De Buck S, Windels P, De Loose M, Depicker A (2004) Single-copy T-DNAs integrated at different positions in the Arabidopsis genome display uniform and comparable β-glucuronidase accumulation levels. Cell Mol Life Sci 61:2632–2645

De Muynck B, Navarre C, Boutry M (2010) Production of antibodies in plants: Status after twenty years. Plant Biotechnol J 8:529–563

de Neve M, De Loose M, Jacobs A, et al (1993) Assembly of an antibody and its derived antibody fragment in Nicotiana and Arabidopsis. Transgenic Res 2:227–237

De Paepe A, De Buck S, Nolf J, Depicker A (2012) High frequency of single-copy T-DNA transformants produced after floral dip in CRE-expressing arabidopsis plants. Methods Mol Biol 847:317–333

De Pourcq K, De Schutter K, Callewaert N (2010) Engineering of glycosylation in yeast and other fungi: Current state and perspectives. Appl. Microbiol. Biotechnol. 87:1617–1631

Decker EL, Reski R (2007) Moss bioreactors producing improved biopharmaceuticals. Curr Opin Biotechnol 18:393–398

Decker EL, Reski R (2008) Current achievements in the production of complex biopharmaceuticals with moss bioreactors. Bioprocess Biosyst Eng 31:3–9

Deepa K, Rodionov RN, Weiss N, Parani M (2013) Transgenic expression and functional characterization of human platelet derived growth factor BB (hPDGF-BB) in Tobacco (Nicotiana tabacum L.). Appl Biochem Biotechnol 171:1390–1404

Demain AL, Vaishnav P (2009) Production of Recombinant Proteins by Microbes and Higher Organisms. In: Comprehensive Biotechnology, Second Edition. pp 333–345

Demirbas A (2010) Use of algae as biofuel sources. Energy Convers Manag 51:2738–2749

Dietz-Pfeilstetter A (2010) Stability of transgene expression as a challenge for genetic engineering. Plant Sci. 179:164–167

Dong B, Hu H-H, Li Z-F, et al (2017) A novel bicistronic expression system composed of the intraflagellar transport protein gene ift25 and FMDV 2A sequence directs robust nuclear gene expression in Chlamydomonas reinhardtii. Appl Microbiol Biotechnol.

Donnelly MLL, Hughes LE, Luke G, et al (2001) The “cleavage” activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring “2A-like” sequences. J Gen Virol 82:1027–1041

Dorokhov YL (2007) Gene silencing in plants. Mol Biol 41:579–592

Dougherty WG, Cary SM, Parks TD (1989) Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. Virology 171:356–364

Duilio A, Madonna S, Tutino ML, et al (2004a) Promoters from a cold-adapted bacterium: Definition of a consensus motif and molecular characterization of UP regulative elements. Extremophiles 8:125–132

Duilio A, Tutino ML, Marino G (2004b) Recombinant protein production in Antarctic Gram-negative bacteria. Methods Mol Biol 267:225–237

Dumas P, Bergdoll M, Cagnon C, Masson JM (1994) Crystal structure and site-directed mutagenesis of a bleomycin resistance protein and their significance for drug sequestering. EMBO J 13:2483–2492

Ebersold WT (1967) Chlamydomonas reinhardi: heterozygous diploid strains. Sci (New York, NY) 157:447–449

Ecker DM, Jones SD, Levine HL (2015) The therapeutic monoclonal antibody market. MAbs 7:9–14

Ehrenberg L, Eriksson G, Eriksson G, et al (1971) Hydroxyproline Heterooligosaccharides in Chlamydomonas. 73:1970–1973

Eichler-Stahlberg A, Weisheit W, Ruecker O, Heitzer M (2009) Strategies to facilitate transgene expression in Chlamydomonas reinhardtii. Planta 229:873–883

Feng S, Feng W, Zhao L, et al (2014) Preparation of transgenic Dunaliella salina for immunization against white spot syndrome virus in crayfish. Arch Virol 159:519–525

Ferrer-Miralles N, Domingo-Espín J, Corchero J, et al (2009) Microbial factories for recombinant pharmaceuticals. Microb Cell Fact 8:17

Fields FJ, Ostrand JT, Mayfield SP (2018) Fed-batch mixotrophic cultivation of Chlamydomonas reinhardtii for high-density cultures. Algal Res 33:109–117

Fischer N, Rochaix JD (2001) The flanking regions of PsaD drive efficient gene expression in the nucleus of the green alga Chlamydomonas reinhardtii. Mol Genet Genomics 265:888–894

Fischer R, Stoger E, Schillberg S, et al (2004) Plant-based production of biopharmaceuticals. Curr. Opin. Plant Biol. 7:152–158

Fischer S, Handrick R, Otte K (2015) The art of CHO cell engineering: A comprehensive retrospect and future perspectives. Biotechnol. Adv. 33:1878–1896

Fletcher SP, Muto M, Mayfield SP (2007) Optimization of recombinant protein expression in the chloroplasts of green algae. Adv. Exp. Med. Biol. 616:90–98

Fragoso G, Hernández M, Cervantes-Torres J, et al (2017) Transgenic papaya: a useful platform for oral vaccines. Planta 245:1037–1048

Fuhrmann M, Hausherr A, Ferbitz L, et al (2004) Monitoring dynamic expression of nuclear genes in Chlamydomonas reinhardtii by using a synthetic luciferase reporter gene. Plant Mol Biol 55:869–881

Fuhrmann M, Oertel W, Hegemann P (1999) A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in Chlamydomonas reinhardtii. Plant J 19:353–361

Gemmill TR, Trimble RB (1999) Overview of N- and O-linked oligosaccharide structures found in various yeast species. Biochim. Biophys. Acta - Gen. Subj. 1426:227–237

Geng D, Wang Y, Wang P, et al (2003) Stable expression of hepatitis B surface antigen gene in Dunaliella salina (Chlorophyta). J Appl Phycol 15:451–456

Georgianna DR, Hannon MJ, Marcuschi M, et al (2013) Production of recombinant enzymes in the marine alga Dunaliella tertiolecta. Algal Res 2:2–9

Gimble JM, Katz AJ, Bunnell BA (2007) Adipose-derived stem cells for regenerative medicine. Circ Res 100:1249–1260

Gimpel JA, Hyun JS, Schoepp NG, Mayfield SP (2015) Production of recombinant proteins in microalgae at pilot greenhouse scale. Biotechnol Bioeng 112:339–345

Girod PA, Nguyen DQ, Calabrese D, et al (2007) Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. Nat Methods 4:747–753

Giuliani M, Parrilli E, Ferrer P, et al (2011) Process optimization for recombinant protein production in the psychrophilic bacterium Pseudoalteromonas haloplanktis. Process Biochem 46:953–959

Goetze AM, Liu YD, Zhang Z, et al (2011) High-mannose glycans on the Fc region of therapeutic IgG antibodies increase serum clearance in humans. Glycobiology 21:949–959

Gomord V, Fitchette AC, Menu-Bouaouiche L, et al (2010) Plant-specific glycosylation patterns in the context of therapeutic protein production. Plant

Biotechnol. J. 8:564–587

Gong Y, Hu H, Gao Y, et al (2011) Microalgae as platforms for production of recombinant proteins and valuable compounds: Progress and prospects. J Ind Microbiol Biotechnol 38:1879–1890

Guilak F, Butler DL, Goldstein SA, Baaijens FPT (2014) Biomechanics and mechanobiology in functional tissue engineering. J. Biomech. 47:1933–1940

Guillemette T, van Peij NN, Goosen T, et al (2007) Genomic analysis of the secretion stress response in the enzyme-producing cell factory Aspergillus niger. BMC Genomics 8:158

Hajimiri M, Shahverdi S, Kamalinia G, Dinarvand R (2015) Growth factor conjugation: Strategies and applications. J Biomed Mater Res - Part A 103:819–838

Haq TA, Mason HS, Clements JD, Arntzen CJ (1995) Oral immunization with a recombinant bacterial antigen produced in transgenic plants. Science 268:714–716

Harris EH (2001) Chlamydomonas as a model organism. Mol Biol 52:363–406

Haryadi R, Ho S, Kok YJ, et al (2015) Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. PLoS One 10:

Heitzer M, Eckert A, Fuhrmann M, Griesbeck C (2007) Chapter 5. Influence of Codon Bias on the Expression. Transgenic Microalgae as Green Cell Factories 46–53

Heitzer M, Zschoernig B (2007) Construction of modular tandem expression vectors for the green alga Chlamydomonas reinhardtii using the Cre/lox-system. Biotechniques 43:324–332

Hempel F, Lau J, Klingl A, Maier UG (2011) Algae as protein factories: Expression of a human antibody and the respective antigen in the diatom phaeodactylum tricornutum. PLoS One 6:

Hempel F, Maier UG (2012) An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. Microb Cell Fact 11:

Hempel F, Maurer M, Brockmann B, et al (2017) 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 16:

Hernández JL, Padilla L, Dakhel S, et al (2013) Therapeutic Targeting of Tumor Growth and Angiogenesis with a Novel Anti-S100A4 Monoclonal Antibody. PLoS One 8:

Hernández M, Rosas G, Cervantes J, et al (2014) Transgenic plants: A 5-year update on oral antipathogen vaccine development. Expert Rev. Vaccines 13:1523–1536

Hiatt A, Caffferkey R, Bowdish K (1989) Production of antibodies in transgenic plants. Nature 342:76–78

Ho SCL, Koh EYC, van Beers M, et al (2013) Control of IgG LC: HC ratio in stably transfected CHO cells and study of the impact on expression, aggregation, glycosylation and conformational stability. J Biotechnol 165:157–166

Hobbs SL, Warkentin TD, DeLong CM (1993) Transgene copy number can be positively or negatively associated with transgene expression. Plant Mol Biol 21:17–26

Houdebine LM (2009) Production of pharmaceutical proteins by transgenic animals. Comp Immunol Microbiol Infect Dis 32:107–121

Hu J, Deng X, Shao N, et al (2014) Rapid construction and screening of artificial microRNA systems in Chlamydomonas reinhardtii. Plant J 79:1052–1064

Innocentin S, Guimarães V, Miyoshi A, et al (2009) Lactococcus lactis expressing either Staphylococcus aureus fibronectin-binding protein A or Listeria monocytogenes internalin A can efficiently internalize and deliver DNA in human epithelial cells. Appl Environ Microbiol 75:4870–4878

Jeong Br B, Wu-Scharf D, Zhang C, Cerutti H (2002) Suppressors of transcriptional transgenic silencing in Chlamydomonas are sensitive to DNA-damaging agents and reactivate transposable elements. Proc Natl Acad Sci U S A 99:1076–81

Jia B, Jeon CO (2016) High-throughput recombinant protein expression in Escherichia coli: current status and future perspectives. Open Biol 6:1–17

Jin H, Cantin GT, Maki S, et al (2011) Soluble periplasmic production of human granulocyte colony-stimulating factor (G-CSF) in Pseudomonas fluorescens. Protein Expr Purif 78:69–77

Jinkerson RE, Jonikas MC (2015) Molecular techniques to interrogate and edit the Chlamydomonas nuclear genome. Plant J 82:393–412

Joseph BC, Pichaimuthu S, Srimeenakshi S (2015) An Overview of the Parameters for Recombinant Protein Expression in Escherichia coli. J Cell Sci Ther 06:

Jun H, Kieselbach T, Jönsson LJ (2011) Enzyme production by filamentous fungi: analysis of the secretome of Trichoderma reesei grown on unconventional carbon source. Microb Cell Fact 10:68

Kaplon H, Reichert JM (2018) Antibodies to watch in 2018. MAbs 10:183–203

Karnieli O, OM F, JG A, et al (2017) A consensus introduction to serum replacements and serum-free media for cellular therapies. Cytotherapy 19:155–169

Kieliszewski MJ, Lamport DTA (1994) Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. Plant J. 5:157–172

Kikuchi Y, Itaya H, Date M, et al (2008) Production of Chryseobacterium proteolyticum protein-glutaminase using the twin-arginine translocation pathway in Corynebacterium glutamicum. Appl Microbiol Biotechnol 78:67–74

Kikuchi Y, Itaya H, Date M, et al (2009) TatABC overexpression improves Corynebacterium glutamicum Tat-dependent protein secretion. Appl Environ Microbiol 75:603–607

Kilian O, Benemann CSE, Niyogi KK, Vick B (2011) High-efficiency homologous recombination in the oil-producing alga Nannochloropsis sp. Proc Natl Acad Sci 108:21265–21269

Kim D-H, Kim YT, Cho JJ, et al (2002) Stable Integration and Functional Expression of Flounder Growth Hormone Gene in Transformed Microalga, Chlorella ellipsoidea. Mar Biotechnol 4:0063–0073

Kim DG, La HJ, Ahn CY, et al (2011) Harvest of Scenedesmus sp. with bioflocculant and reuse of culture medium for subsequent high-density cultures. Bioresour Technol 102:3163–3168

Kindle KL (1990) High-frequency nuclear transformation of Chlamydomonas reinhardtii. Proc. Natl. Acad. Sci. USA 87:1228–1232

Kleinman HK, Klebe RJ, Martin GR (1981) Role of collagenous matrices in the adhesion and growth of cells. J. Cell Biol. 88:473–485

Knappskog S, Ravneberg H, Gjerdrum C, et al (2007) The level of synthesis and secretion of Gaussia princeps luciferase in transfected CHO cells is heavily dependent on the choice of signal peptide. J Biotechnol 128:705–715

Kober L, Zehe C, Bode J (2013) Optimized signal peptides for the development of high expressing CHO cell lines. Biotechnol Bioeng 110:1164–1173

Kong F, Yamasaki T, Kurniasih SD, et al (2015) Robust expression of heterologous genes by selection marker fusion system in improved Chlamydomonas strains. J Biosci Bioeng 120:239–245

Kong F, Yamasaki T, Ohama T (2014) Expression levels of domestic cDNA cassettes integrated in the nuclear genomes of various Chlamydomonas reinhardtii strains. J Biosci Bioeng 117:613–616

Korz DJ, Rinas U, Hellmuth K, et al (1995) Simple fed-batch technique for high cell density cultivation of Escherichia coli. J Biotechnol 39:59–65

Kronholm I, Bassett A, Baulcombe D, Collins S (2017) Epigenetic and Genetic Contributions to Adaptation in Chlamydomonas. Mol Biol Evol 34:2285–2306

Kumar A, Falcao VR, Sayre RT (2013) Evaluating nuclear transgene expression systems in Chlamydomonas reinhardtii. Algal Res 2:321–332

Kurniasih SD, Yamasaki T, Kong F, et al (2016) UV-mediated Chlamydomonas mutants with enhanced nuclear transgene expression by disruption of DNA methylation-dependent and independent silencing systems. Plant Mol Biol 92:629–641

Kwak M, Park WK, Shin SE, et al (2017) Improvement of biomass and lipid yield under stress conditions by using diploid strains of Chlamydomonas reinhardtii. Algal Res 26:180–189

Lauersen KJ, Baier T, Wichmann J, et al (2016) Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga Chlamydomonas reinhardtii. Metab Eng 38:331–343

Lauersen KJ, Berger H, Mussgnug JH, Kruse O (2013a) Efficient recombinant protein production and secretion from nuclear transgenes in Chlamydomonas reinhardtii. J Biotechnol 167:101–110

Lauersen KJ, Huber I, Wichmann J, et al (2015a) Investigating the dynamics of recombinant protein secretion from a microalgal host. J Biotechnol 215:62–71

Lauersen KJ, Kruse O, Mussgnug JH (2015b) Targeted expression of nuclear transgenes in Chlamydomonas reinhardtii with a versatile, modular vector toolkit. Appl Microbiol Biotechnol 99:3491–3503

Lauersen KJ, Vanderveer TL, Berger H, et al (2013b) Ice recrystallization inhibition mediated by a nuclear-expressed and -secreted recombinant ice-binding protein in the microalga Chlamydomonas reinhardtii. Appl Microbiol Biotechnol 97:9763–72

LaVallie E, DiBlasio EA, Kovacic S, et al (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. Nat Biotechnol 11:187–193

Laxa M (2017) Intron-Mediated Enhancement: A Tool for Heterologous Gene Expression in Plants? Front Plant Sci 7:

Le Loir Y, Azevedo V, Oliveira SC, et al (2005) Protein secretion in Lactococcus lactis : an efficient way to increase the overall heterologous protein production. Microb Cell Fact 4:2

Lee E, Lim ZR, Chen HY, et al (2018) Defined Serum-Free Medium for Bioreactor Culture of an Immortalized Human Erythroblast Cell Line. Biotechnol J 1700567:1–12

Lee HO, Davidson JM, Duronio RJ (2010) Endoreplication: polyploidy with purpose. Genes Dev 2461–2477

Lemaire SD, Miginiac-Maslow M (2004) The thioredoxin superfamily in Chlamydomonas reinhardtii. Photosynth. Res. 82:203–220

Li H, Sethuraman N, Stadheim TA, et al (2006) Optimization of humanized IgGs in glycoengineered Pichia pastoris. Nat Biotechnol 24:210–215

Li Y, Liu D, López-Paz C, et al (2016) A new class of cyclin dependent kinase in chlamydomonas is required for coupling cell size to cell division. Elife 5:1–28

Lindroos B, Boucher S, Chase L, et al (2009) Serum-free, xeno-free culture media maintain the proliferation rate and multipotentiality of adipose stem cells in vitro. Cytotherapy 11:958–972

Liu L (2015) Antibody glycosylation and its impact on the pharmacokinetics and pharmacodynamics of monoclonal antibodies and Fc-fusion proteins. J. Pharm.

Sci. 104:1866–1884

López-Paz C, Liu D, Geng S, Umen JG (2017) Identification of Chlamydomonas reinhardtii endogenous genic flanking sequences for improved transgene expression. Plant J 92:1232–1244

Loyevsky M, Mompoint F, Yikilmaz E, et al (2003) Expression of a recombinant IRP-like Plasmodium falciparum protein that specifically binds putative plasmodial IREs. Mol Biochem Parasitol 126:231–238

Lumbreras V, Stevens DR, Purton S (1998) Efficient foreign gene expression in Chlamydomonas reinhardtii mediated by an endogenous intron. Plant J 14:441–447

Ma JKC, Hiatt A, Hein M, et al (1995) Generation and assembly of secretory antibodies in plants. Science (80- ) 268:716–719

Ma JKC, Hikmat BY, Wycoff K, et al (1998) Characterization of a recombinant plant monoclonal secretory antibody and preventive immunotherapy in humans. Nat Med 4:601–606

Makrides SC (1999) Components of vectors for gene transfer and expression in mammalian cells. Protein Expr Purif 17:183–202

Mandalam RK, Palsson B (1998) Elemental balancing of biomass and medium composition enhances growth capacity in high-density Chlorella vulgaris cultures. Biotechnol Bioeng 59:605–611

Martínez-González L, Rosales-Mendoza S, Soria-Guerra RE, et al (2011) Oral immunization with a lettuce-derived Escherichia coli heat-labile toxin B subunit induces neutralizing antibodies in mice. Plant Cell Tissue Organ Cult 107:441–449

Martino MM, Briquez PS, Maruyama K, Hubbell JA (2015) Extracellular matrix-inspired growth factor delivery systems for bone regeneration. Adv Drug Deliv Rev 94:41–52

Mayfield SP, Franklin SE (2005) Expression of human antibodies in eukaryotic micro-algae. In: Vaccine. pp 1828–1832

Mayfield SP, Franklin SE, Lerner RA (2003) Expression and assembly of a fully active antibody in algae. Proc Natl Acad Sci 100:438–442

Mayfield SP, Manuell AL, Chen S, et al (2007) Chlamydomonas reinhardtii chloroplasts as protein factories. Curr. Opin. Biotechnol. 18:126–133

Merchant SS, Prochnik SE, Vallon O, et al (2007) The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. Science (80- ) 318:245–250

Merten OW, Kallel H, Manuguerra JC, et al (1999) The new medium MDSS2N, free of any animal protein supports cell growth and production of various viruses. Cytotechnology 30:191–201

Miletic S, Hünerberg M, Kaldis A, et al (2017) A Plant-Produced Candidate Subunit Vaccine Reduces Shedding of Enterohemorrhagic Escherichia coli in Ruminants. Biotechnol J 12:1–9

Miyoshi A, Bermúdez-Humarán LG, Ribeiro LA, et al (2006) Heterologous expression of Brucella abortus GroEL heat-shock protein in Lactococcus lactis. Microb Cell Fact 5:14

Mlynarova L, Hricová A, Loonen A, Nap JP (2003) The Presence of a Chromatin Boundary Appears to Shield a Transgene in Tobacco from RNA Silencing. Plant Cell 15:2203–2217

Mlynarova L, Keizer L, Stiekema WJ, Nap JP (1996) Approaching the Lower Limits of Transgene Variability. Plant Cell 8:1589–1599

Moabbi AM, Agarwal N, El Kaderi B, Ansari A (2012) Role for gene looping in intron-mediated enhancement of transcription. Proc Natl Acad Sci 109:8505–8510

Molino JVD, de Carvalho JCM, Mayfield SP (2018) Comparison of secretory signal peptides for heterologous protein expression in microalgae: Expanding the secretion portfolio for Chlamydomonas reinhardtii. PLoS One 13:

Mor TS (2015) Molecular pharming's foot in the FDA's door: Protalix's trailblazing story. Biotechnol. Lett. 37:2147–2150

Morello E, Bermúdez-Humarán LG, Llull D, et al (2007) Lactococcus lactis, an efficient cell factory for recombinant protein production and secretion. J Mol Microbiol Biotechnol 14:48–58

Morrison SL, Scharff MD (1975) Heavy Chain-Producing Variants of a Mouse Myeloma Cell Line. J Immunol 655–659

Mosmann T (1983) Rapid colorimetric assay for cellular growth and survival:

Application to proliferation and cytotoxicity assays. J Immunol Methods 65:55–63

Nelson BK, Cai X, Nebenführ A (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J 51:1126–1136

Neupert J, Karcher D, Bock R (2009) Generation of Chlamydomonas strains that efficiently express nuclear transgenes. Plant J 57:1140–1150

Newman SM, Boynton JE, Gillham NW, et al (1990) Transformation of Chloroplast Ribosomal-Rna Genes in Chlamydomonas - Molecular and Genetic-Characterization of Integration Events. Genetics 126:875–888

Nielsen J (2013) Production of biopharmaceutical proteins by yeast: Advances through metabolic engineering. Bioengineered 4:207–211

Nose M, Wigzell H (1983) Biological significance of carbohydrate chains on monoclonal antibodies. Proc Natl Acad Sci U S A 80:6632–6636

Nyyssönen E, Penttilä M, Harkki A, et al (1993) Efficient production of antibody fragments by the filamentous fungus Trichoderma reesei. Biotechnology (N Y) 11:591–5

Onishi M, Pringle JR (2016) Robust Transgene Expression from Bicistronic mRNA in the Green Alga Chlamydomonas reinhardtii. Genes|Genomes|Genetics 6:4115–4125

Ortega‐berlanga B, Bañuelos‐hernández · Bernardo, Rosales‐mendoza S (2018) Efficient Expression of an Alzheimer's Disease Vaccine Candidate in the Microalga Schizochytrium sp. Using the Algevir System. Mol Biotechnol 60:362–368

Owen M, Gandecha A, Cockburn B, Whitelam G (1992) Synthesis of a functional anti-phytochrome single-chain Fv protein in transgenic tobacco. Nat Biotechnol 10:790–794

Peremarti A, Twyman RM, Gómez-Galera S, et al (2010) Promoter diversity in multigene transformation. Plant Mol. Biol. 73:363–378

Petersen K, Leah R, Knudsen S, Cameron-Mills V (2002) Matrix attachment regions (MARs) enhance transformation frequencies and reduce variance of transgene expression in barley. Plant Mol Biol 49:45–58

Pirie CM, Hackel BJ, Rosenblum MG, Wittrup KD (2011) Convergent potency of internalized gelonin immunotoxins across varied cell lines, antigens, and targeting moieties. J Biol Chem 286:4165–4172

Plucinak TM, Horken KM, Jiang W, et al (2015) Improved and versatile viral 2A platforms for dependable and inducible high-level expression of dicistronic nuclear genes in Chlamydomonas reinhardtii. Plant J 82:717–729

Pool MR (2005) Signal recognition particles in chloroplasts, bacteria, yeast and mammals. Mol Membr Biol 22:3–15

Porro D, Gasser B, Fossati T, et al (2011) Production of recombinant proteins and metabolites in yeasts. Appl Microbiol Biotechnol 89:939–948

Porro D, Sauer M, Branduardi P, Mattanovich D (2005) Recombinant Protein Production in Yeasts. Mol Biotechnol 31:245–260

Potvin G, Zhang Z (2010) Strategies for high-level recombinant protein expression in transgenic microalgae: A review. Biotechnol. Adv. 28:910–918

Pulz O (2001) Photobioreactors: Production systems for phototrophic microorganisms. Appl Microbiol Biotechnol 57:287–293

Ramos-Martinez EM, Fimognari L, Sakuragi Y (2017) High-yield secretion of recombinant proteins from the microalga Chlamydomonas reinhardtii. Plant Biotechnol J 15:1214–1224

Rasala, B.A., Mayfield SP (2011) The microalga Chlamydomonas reinhardtii as a platform for the production of human protein therapeutics. Bioengineered 2:50–54

Rasala BA, Barrera DJ, Ng J, et al (2013) Expanding the spectral palette of fluorescent proteins for the green microalga Chlamydomonas reinhardtii. Plant J 74:545–556

Rasala BA, Chao S-S, Pier M, et al (2014) Enhanced Genetic Tools for Engineering Multigene Traits into Green Algae. PLoS One 9:e94028

Rasala BA, Lee PA, Shen Z, et al (2012) Robust expression and secretion of xylanase1 in Chlamydomonas reinhardtii by fusion to a selection gene and processing with the FMDV 2A peptide. PLoS One 7:

Rasala BA, Mayfield SP (2015) Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. Photosynth Res 123:227–39

Reski R, Parsons J, Decker EL (2015) Moss-made pharmaceuticals: From bench to bedside. Plant Biotechnol J 13:1191–1198

Retallack DM, Jin H, Chew L (2012) Reliable protein production in a Pseudomonas fluorescens expression system. Protein Expr. Purif. 81:157–165

Retallack DM, Schneider JC, Mitchell J, et al (2007) Transport of heterologous proteins to the periplasmic space of Pseudomonas fluorescens using a variety of native signal sequences. Biotechnol Lett 29:1483–1491

Roberts K, Gurney-Smith M, Hills GJ (1972) Structure, composition and morphogenesis of the cell wall of Chlamydomonas reinhardi. VI. The flagellar collar. J Ultrastruct Res 599–613

Romanova N, Noll T (2017) Engineered and natural promoters and chromatin-modifying elements for recombinant protein expression in CHO cells. Biotechnol J 1700232:1700232

Rosales-Mendoza S (2016) Algae-Based Biopharmaceuticals. Springer International Publishing, Cham

Rosales-Mendoza S, Paz-Maldonado LMT, Soria-Guerra RE (2012) Chlamydomonas reinhardtii as a viable platform for the production of recombinant proteins: Current status and perspectives. Plant Cell Rep. 31:479–494

Ruecker O, Zillner K, Groebner-Ferreira R, Heitzer M (2008) Gaussia-luciferase as a sensitive reporter gene for monitoring promoter activity in the nucleus of the green alga Chlamydomonas reinhardtii. Mol Genet Genomics 280:153–162

Sapra P, Shor B (2013) Monoclonal antibody-based therapies in cancer: Advances and challenges. Pharmacol. Ther. 138:452–469

Schiff LJ (2005) Review: production, characterization, and testing of banked mammalian cell substrates to produce biological products. Vitr Cell Dev Biol - Anim 41:65–70

Schmidt-Hansen B, Klingelhöfer J, Grum-Schwensen B, et al (2004) Functional significance of metastasis-inducing S100A4(Mts1) in tumor-stroma interplay. J Biol Chem 279:24498–24504

Schmidt FR (2004) Recombinant expression systems in the pharmaceutical industry. Appl Microbiol Biotechnol 65:363–372

Schroda M, Blöcker D, Beck CF (2000) The HSP70A promoter as a tool for the improved expression of transgenes in Chlamydomonas. Plant J 21:121–131

Schubert D (2004) Silencing in Arabidopsis T-DNA Transformants: The Predominant Role of a Gene-Specific RNA Sensing Mechanism versus Position Effects. Plant Cell Online 16:2561–2572

Scranton MA, Ostrand JT, Georgianna DR, et al (2016) Synthetic promoters capable of driving robust nuclear gene expression in the green alga Chlamydomonas reinhardtii. Algal Res 15:135–142

Shao N, Bock R (2008) A codon-optimized luciferase from Gaussia princeps facilitates the in vivo monitoring of gene expression in the model alga Chlamydomonas reinhardtii. Curr Genet 53:381–388

Shi X, Jarvis DL (2007) Protein N-glycosylation in the baculovirus-insect cell system. Curr Drug Targets 8:1116–25

Shields RL, Lai J, Keck R, et al (2002) Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity. J Biol Chem 277:26733–26740

Shimogawara K, Fujiwara S, Grossman A, Usuda H (1998) High-efficiency transformation of Chlamydomonas reinhardtii by electroporation. Genetics 148:1821–1828

Shpak E, Leykam JF, Kieliszewski MJ (1999) Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. Proc Natl Acad Sci U S A 96:14736–14741

Sinclair AM, Elliott S (2005) Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. J Pharm Sci 94:1626–1635

Sizova I, Fuhrmann M, Hegemann P (2001) A Streptomyces rimosus aphVIII gene coding for a new type phosphotransferase provides stable antibiotic resistance to Chlamydomonas reinhardtii. Gene 277:221–229

Specht E, Miyake-Stoner S, Mayfield S (2010) Micro-algae come of age as a platform for recombinant protein production. Biotechnol. Lett. 32:1373–1383

Specht EA, Mayfield SP (2014) Algae-based oral recombinant vaccines. Front Microbiol 5:

Spolaore P, Joannis-Cassan C, Duran E, Isambert A (2006) Commercial applications of microalgae. J Biosci Bioeng 101:87–96

Stammen S, Müller BK, Korneli C, et al (2010) High-yield intra- And extracellular protein production using bacillus megaterium. Appl Environ Microbiol 76:4037–4046

Staudacher E, Altmann F, Wilson IBH, März L (1999) Fucose in N-glycans: From plant to man. Biochim. Biophys. Acta - Gen. Subj. 1473:216–236

Stevens DR, Purton S, Rochaix J-D (1996) The bacterial phleomycin resistance geneble as a dominant selectable marker inChlamydomonas. Mol Gen Genet MGG 251:23–30

Strohl WR (2015) Fusion Proteins for Half-Life Extension of Biologics as a Strategy to Make Biobetters. BioDrugs 29:215–239

Tabata Y (2003) Tissue Regeneration Based on Growth Factor Release. Tissue Eng 9:5–15

Tan KY, Teo KL, Lim JF, et al (2015) Serum-free media formulations are cell line-specific and require optimization for microcarrier culture. Cytotherapy 17:1152–1165

Tao MH, Morrison SL (1989) Studies of aglycosylated chimeric mouse-human IgG. Role of carbohydrate in the structure and effector functions mediated by the human IgG constant region. J Immunol 143:2595–2601

Tekoah Y, Tzaban S, Kizhner T, et al (2013) Glycosylation and functionality of recombinant β-glucocerebrosidase from various production systems. Biosci Rep 33:771–781

Tran M, Henry RE, Siefker D, et al (2013a) Production of anti-cancer immunotoxins in algae: Ribosome inactivating proteins as fusion partners. Biotechnol Bioeng 110:2826–2835

Tran M, Van C, Barrera DJ, et al (2013b) Production of unique immunotoxin cancer therapeutics in algal chloroplasts. Proc Natl Acad Sci 110:E15–E22

Tran M, Zhou B, Pettersson PL, et al (2009) Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. Biotechnol Bioeng 104:663–673

Ülker B, Allen GC, Thompson WF, et al (1999) A tobacco matrix attachment region reduces the loss of transgene expression in the progeny of transgenic tobacco plants. Plant J 18:253–263

Uusitalo JM, Helena Nevalainen KM, Harkki AM, et al (1991) Enzyme production by recombinant *Trichoderma reesei* strains. J Biotechnol 17:35–49

Vain P, Finer KR, Engler DE, et al (1996) Intron-mediated enhancement of gene expression in maize (Zea mays L.) and bluegrass (Poa pratensis L.). Plant Cell Rep 2:489–494

Vain P, Worland B, Kohli A, et al (1999) Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. Plant J 18:233–242

Valdés R, Reyes B, Alvarez T, et al (2003) Hepatitis B surface antigen immunopurification using a plant-derived specific antibody produced in large scale. Biochem Biophys Res Commun 310:742–747

Van Der LindenY RHJ, Frenken LGJ, Geus B De, Harmsen MM (1999) Comparison of physical chemical properties of llama V HH antibody fragments and mouse monoclonal antibodies. 1431:37–46

Van Der Valk J, Mellor D, Brands R, et al (2004) The humane collection of fetal bovine serum and possibilities for serum-free cell and tissue culture. In: Toxicology in Vitro. pp 1–12

Vanier G, Stelter S, Vanier J, et al (2017) Alga-Made Anti-Hepatitis B Antibody Binds to Human Fcγ Receptors. Biotechnol J 1700496:1–10

Walker TL, Purton S, Becker DK, Collet C (2005) Microalgae as bioreactors. Plant Cell Rep. 24:629–641

Wallick SC, Kabat EA, Morrison SL (1988) Glycosilation of a VH residue of a monoclonal antibody against a(1-6) dextran increases its affinity for antigen. J Exp Med 168:1099–1109

Walsh G (2014) Biopharmaceutical benchmarks 2014. Nat Biotechnol 32:992–1000

Walsh G, Jefferis R (2006) Post-translational modifications in the context of therapeutic proteins. Nat. Biotechnol. 24:1241–1252

Weiner I, Atar S, Schweitzer S, et al (2018) Enhancing heterologous Expression in Chlamydomonas reinhardtii by Transcript Sequence Optimization. Plant J accepted:

Werner R, Mergenhagen D (1998) Mating Type Determination of Chlamydomonas reinhardtii by PCR. Plant Mol Biol Report 16:295-299–299

Wijffels RH, Barbosa MJ (2010) An Outlook on Microalgal Biofuels. Science (80-) 329:796–799

Woessner JP, Goodenough UW (1992) Zygote and vegetative cell wall proteins in Chlamydomonas reinhardtii share a common epitope, (SerPro)X. Plant Sci 83:65–76

Wong WKR, Lam E, Huang RC, et al (2001) Applications, and efficient large-scale production, of recombinant human epidermal growth factor. Biotechnol Genet Eng Rev 18:51–71

Wright a, Tao MH, Kabat E a, Morrison SL (1991) Antibody variable region glycosylation: position effects on antigen binding and carbohydrate structure. EMBO J 10:2717–23

Wu-Scharf D, Jeong B, Zhang C, Cerutti H (2000) Transgene and transposon silencing in Chlamydomonas reinhardtii by a DEAH-box RNA helicase. Science 290:1159–1162

Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22:1393–1398

Xiao R, Zheng Y (2016) Overview of microalgal extracellular polymeric substances (EPS) and their applications.

Xiong W, Li X, Xiang J, Wu Q (2008) High-density fermentation of microalga Chlorella protothecoides in bioreactor for microbio-diesel production. Appl Microbiol Biotechnol 78:29–36

Xu J, Tan L, Goodrum KJ, Kieliszewski1 MJ (2007) High-Yields and Extended Serum Half-Life of Human Interferon a2b Expressed in Tobacco Cells as Arabinogalactan-Protein Fusions. Biotechnol Bioeng 97:997–1008

Yeh P, Landais D, Lemaitre M, et al (1992) Design of yeast-secreted albumin derivatives for human therapy: Biological and antiviral properties of a serum albumin-CD4 genetic conjugate. Appl Biol Sci 89:1904–1908

Zeitlin L, Bohorov O, Bohorova N, et al (2013) Prophylactic and therapeutic testing of Nicotiana-derived RSV-neutralizing human monoclonal antibodies in the cotton rat model. MAbs 5:263–269

Zhang C, Wu Scharf D, Jeong B, Cerutti H (2002) A WD40 repeat containing protein, similar to a fungal co repressor, is required for transcriptional gene silencing in Chlamydomonas. Plant J 31:25–36

Zhang R, Patena W, Armbruster U, et al (2014) High-Throughput Genotyping of Green Algal Mutants Reveals Random Distribution of Mutagenic Insertion Sites and Endonucleolytic Cleavage of Transforming DNA. Plant Cell 26:1398–1409

Zhang YH, Robinson DG (1990) Cell-wall synthesis in Chlamydomonas reinhardtii: an immunological study on the wild type and wall-less mutants cw2 and cw15. Planta 180:229–236

Zhou J, Yang Y, Wang X, et al (2013) Enhanced transgene expression in rice following selection controlled by weak promoters. BMC Biotechnol 13:29

# *Patents*

Solazyme, Inc. Microalgal polysaccharide compositions. Inventor: Anna Coragliotti, Scott Franklin, Anthony G. Day, Stephen M. Decker. US 8,927,522 B2. 2015-01-06.

Lipotec SA, POLYMARIS BIOTECHNOLOGY. Exopolysaccharide for the treatment and/or care of the skin, mucous membranes, hair and/or nails. Inventor: Anthony Courtois, Bertrand Thollas, Raquel Delgado, Juan Cebrian, Albert Soley. EP2646115 B1. 2010-11-30.

Lipotec SA. Antarticine-nf3 for the treatment and re-epithelialisation of wounds. Inventor: Antonio Parente, Josep Garces, Jesus Guinea, Josep Maria Garcia, Ricardo Casaroli, Manuel Reina, Senen Vilaro. EP1402898 B1. 2001-06-14.

Latvijas UnivLatvijas Universitate. Method for the isolation of glycoprotein-rich fungal extract and its use in anti-ageing cosmetic formulations. Inventor: Anna Ramata-Stunda, Vizma NIKOLAJEVA, Zaiga Petrina, Inese Cakstina, Lita Grine, Ilze Blake, Indrikis Muiznieks. EP2875805 B1. 2013-11-26.

Lykera Biomed SA. S100a4 antibodies and therapeutic uses thereof. Inventor: José Luis HERNÁNDEZ, Jaume Adan, Josep Maria MARTÍNEZ, Marc MASA, Ramon Messeguer, Francesc Mitjans, Sheila Dakhel, Antonio Coll, Rosa Mª HERVAS, Carme Calvis, Laura PADILLA, Lourdes Tatiana Roque, Laura BARBERÀ, Manuel RIVAS, Luis Angel Gomez. WO2011157724 A1. 2010-06-14.