Development and Optimization of Novel Platforms for the Production of Recombinant Proteins

Gabriel Potvin

A thesis submitted to
the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the
PhD degree in Chemical Engineering

Department of Chemical and Biological Engineering
Faculty of Engineering
University of Ottawa

Abstract

As the worldwide demand for recombinant proteins and valuable metabolites continues to grow, and as the biological toolset at our disposal continues to expand, the development of novel, robust, and effective platforms for the production of these bioproducts represents an area of ever-increasing interest. Although many such bioprocesses are currently economically viable, many more, though holding considerable promise, remain uncompetitive. The development of novel, more productive systems increases the versatility and industrial applications of bioprocesses.

The work described in this thesis explores several aspects of bioprocessing, both on the upstream side, concerned with the development of novel recombinant protein expression platforms or the isolation of novel genes with products possessing characteristics of interest, and on the downstream side, through the improvement of fermentation-based bioprocesses.

Thirty-six homoplasmic recombinant strains of the microalga *Chlamydomonas reinhardtii* were developed having integrated genes for phytase or xylanase under the control of *psbA* and *psbD* promoters, codon optimized using novel algorithms, at two different genetic loci, in chloroplasts, to be used as novel animal feed additives. Enzyme production was characterized, and results, when compared to other published work in this field, may provide insight into the factors impacting recombinant protein production in microalgae.

Using a "bio-prospecting approach", the microflora of the digestive tract of a Canadian beaver was screened for cellulase-producing microorganisms. Although the screening approach did successfully identify a novel β-glucosidase gene from an isolated strain of *Bacillus thuringiensis*, the sequence was not significantly different from those already characterized.

Two bioprocessing studies were performed to improve recombinant protein production in *Pichia pastoris*. In the first, the composition of standard Basal Salt Medium (BSM) was systematically optimized for the production of recombinant phytase, and the optimized media produced significantly more enzyme than the standard one, while also containing significantly reduced concentrations of KH_2PO_4 and $MgSO_4 \cdot 7H_2O$ (27.9 g/l and 4.8 g/l respectively), lowering the price of process inputs. The second was based on the screening of unconventional carbon sources for candidates that could sustain the growth and enzyme production using the same *P. pastoris* strain. Fructose and ethanol have shown to be viable alternatives to glucose or glycerol as sole carbon sources, and provide flexibility in terms of process design.

Résumé

Puisque la demande mondiale pour les produits recombinants continue d'augmenter, et que les outils biologiques à notre disposition deviennent de plus en plus puissants et versatils, le développement de nouveaux systèmes robustes et efficacies pour la production de bioproduits demeure un domaine de recherché d'intérêt considérable. Bien que plusieurs bioprocédés soient présentement économiquement viables, plusieurs autres procédés, malgré leur potentiel, ne peuvent pas présentement faire compétition avec les alternatives traditionnelles. Le dévelopement de nouveaux biosystèmes efficaces et puissants augmente les options disponibles lors de design de bioprocédés industriels.

Plusieurs aspects reliés aux bioprocédés, y compris le développement de nouvelles plateformes pour la production de protéines recombinantes, l'identification de nouveaux gènes codant pour des enzymes ayant des caractéristiques intéressantes, et le développement et l'optimisation de procédés de fermentation sont discutés dans cette œuvre.

Trente-six lignées recombinantes homoplasmiques de la micro-algue *Chlamydomonas reinhardtii*, ayant incorporé dans le génome du chloroplaste des gènes pour une phytase ou une xylanase contrôlés pas les promoteurs des gènes *psbA* et *psbD* et dont les codons furent optimisés à l'aide de nouveaux algorithmes, furent développées pour être utilisées comme suppléments nutritionnels pour les animaux monogastriques. L'expression enzymatique fut caractérisée, et les résultats, lorsque comparés avec ceux d'autres œuvres publiées dans ce domaine, pourraient contribuer à la compréhension des différents facteurs qui affectent l'expression protéique dans les micro-algues.

En suivant une approche « d'exploration biologique », la microflore du système digestif d'un castor canadien fut analysée pour identifier des microorganismes qui produisent des enzymes cellulolytiques. Bien que cette approche ait permis l'identification d'un nouveau gène pour une β-glucosidase d'une lignée de *Bacillus thuringiensis*, la séquence génétique n'est pas assez différente de celles déjà caractérisées.

Deux études sur l'amélioration de la production d'enzymes recombinantes à l'aide de la levure *Pichia pastoris* furent effectuées. La première avait comme objectif l'optimisation systématique de la composition du milieu de croissance BSM (*Basal Salt Medium*) pour la production de phytase. La composition optimisée a permis la production d'une quantité significativement plus élevée que le milieu standard, et contient de plus petites quantités de

KH_2PO_4 et $MgSO_4 \cdot 7H_2O$ (27.9 g/l et 4.8 g/l respectivement), ce qui réduit les coûts de production associés au procédé. La deuxième avait pour but l'identification de sources de carbone non-traditionnelles qui peuvent appuyer la croissance et la production d'enzyme dans la même lignée de *P. pastoris*. Les résultats démontrent que le fructose et l'éthanol sont des alternatives viables au glucose et au glycérol, et leur utilisation permet le design de bioprocédés flexibles et versatiles.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ANN	Artificial Neural Network
AOX	Alcohol Oxidase
ATCC	American Type Culture Collection
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
BSM	Basal Salt Medium
CAI	Codon Adaptation Index
CCC	Chlamydomonas Culture Center
CER	Carbon dioxide Evolution Rate
CGMCC	China General Microbiological Culture Collection
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
D	Dilution rate
DCW	Dry Cell Weight
DO	Dissolved Oxygen
EC	Enzyme Commission (number)
EDTA	Ethylenediaminetetraacetic Acid
ER	Endoplasmic Reticulum
FIA	Flow Injection Analysis
FID	Flame Ionization Detectors
FTIR	Fourier Transform Infrared spectroscopy
GAPDH	Glyceraldehyde 3-phosphate Dehydrogenase
GC	Gas Chromatography
GFP	Green Fluorescent Protein
GRAS	Generally Regarded As Safe
HDGS	Homology-Dependent Gene Silencing
HPLC	High Pressure Liquid Chromatography
HRP	Horseradish Peroxidase
ICP	Inductively Coupled Plasma spectroscopy
MFR	Methanol Feeding Rate
mRNA	Messenger RNA

mt+	Mating Type (+)
mt-	Mating Type (-)
MWF	Multi-Wavelength Fluorescence
NCBI	National Center for Biotechnology Information
OD_{xxx}	Optical Density at xxx wavelength
OLFB	Oxygen-Limited Fed-Batch
OUR	Oxygen Uptake Rate
PCA	Principle Component Analysis
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PI	Proportional-Integral (control)
PID	Proportional-Integral-Derivative (control)
PMSF	Phenylmethanesulfonylfluoride
RBCS	Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO)
ROS	Reactive Oxygen Species
rRNA	Ribosomal RNA
SIA	Sequential Injection Analysis
TCP	Total Cell Protein
TLFB	Temperature-Limited Fed-Batch
tRNA	Transfer RNA
TSP	Total Soluble Protein
UTR	Untranslated Region
WCW	Wet Cell Weight
WSSV	White Spot Syndrome Virus

Acknowledgements

It is with the apprehension that the inadequacy of my words may fail to do justice to this task, and with the tormenting realization that the names of deserving friends and colleagues, betrayed by a memory currently, if perhaps reprehensibly, pre-occupied more with technical considerations than thoughts of appreciation, will inevitably, if unintentionally, be omitted from these remarks, that I must attempt to convey here my heartfelt gratitude to all those whose support and camaraderie have made the efforts expended in the pursuit of this degree much less onerous.

Thanks must first, as much as a matter of protocol than as a genuine recognition of indebtedness, be extended to my supervisor Dr. Jason Zhang, for his staunch and unwavering support.

To fail to acknowledge the collective technical Cerberus that is Louis Tremblay, Gérard Nina and Franco Ziroldo would be unforgivable, as they have all unfailingly gone above the call of duty (and perhaps occasionally beyond the limits of the reasonable), in their assistance in all of my undertakings. To say that they have been invaluable to the completion of this work is in no way an exaggeration.

Heartfelt gratitude is particularly offered to Joanne Gamage and Krystle Talbot, as well as to Alison Reiche, Ayla Ahmad, Rong Ran, Zhi Li, Shuping Zhong, and Zhiliang Yang. One could not hope for better or more capable colleagues, let alone friends.

I have had the privilege to work with many talented students over the years that must be acknowledged here, in particular Carol Pietka, Amanda Defela, André Richard, Tanis McMahon, Camille Tremblay-Laganiere and Mathieu Paquette. They without a doubt leave their marks in the research presented here.

I must also acknowledge the generous financial support provided by the National Science and Engineering Research Council, the Ontario Ministry of Research and Innovation, and the Ontario Ministry of Training, Colleges and Universities, whose research funding and scholarships have made this work possible

Finally, it remains to thanks family and friends, too numerous to enumerate, who have always steadfastly offered encouragement and without whose support none of this work would have been completed.

Copyright Statement

All material contained in this thesis that has previously been published has been published in journals that allow their use in graduate theses. Full references for the original publication are provided where required.

Co-Author Contributions

Everything presented in this document is original and entirely the product of the author's work, with the exception of the contributions made by co-authors, which are listed below. All of the presented material is used for the purpose of this thesis with their full knowledge and consent, and in accordance with the rules established by the University of Ottawa.

Ayla Ahmad: Assisted in the research, proof-reading and editing of the published review paper that forms the basis of Chapter 5 of this thesis. She is listed as a co-author on the original published article.

Zhi Li: Assisted in performing the bioprocessing experiments, including sample collection and process monitoring, that form the basis for Chapter 6 of this thesis. He will be listed as a co-author on any publication that results from this work.

Amanda Defela: Performed part of the cell culture experiments that form the basis of Chapter 7 of this thesis, and was involved in the proof-reading of parts of the document presented here. She will be listed as a co-author on any publication that results from this work.

Carol Pietka: Performed the plasmid assembly and enzyme activity characterization experiments described in section 4.4.4 of this thesis, and was involved in the proof-reading of parts the document presented here. She will be listed as a co-author on any publication that results from this work.

Chapter 1 - Introduction

Fuelled by a continued exponential growth in the worldwide demand for recombinant proteins and bioproducts, widely used in the therapeutics, cosmetics and food industries, considerable efforts are allocated to the development of novel efficient, robust, and economically viable bioprocesses for their production.

Bioprocess engineering research, the main focus of this thesis, can be divided into two broad categories, concerning themselves with either the ‘upstream’ or ‘downstream’ aspects of the bioprocesses of interest. In this thesis, upstream considerations pertain to the engineering, development, or modification of the microorganism(s) around which a bioprocess is based with the objectives of either developing a novel strain with given beneficial characteristics, or improving existing strains, typically to improve biomass, metabolite, or heterologous protein yields. The most common approaches include the development of recombinant strains with genetic networks and regulatory elements allowing for the controlled production of the bioproducts of interest, the metabolic engineering of a strain, to maximize the production of a valuable metabolite or to streamline a metabolism for another end-goal, and the development of intricate and tightly regulated biological networks or systems that can be implemented in more complex bioprocesses, for example in a synergetic mixed culture in which different strains are modified to be symbiotic or complementary.

Downstream research, more traditionally associated with chemical engineering, is defined here as anything related to the process itself, usually in terms of its design and operation, with the objectives of maximizing its efficiency and economic viability. This typically involves the optimization of operating parameters, which for bioreactors notably includes operating pH and temperature, agitation rates, aeration rates, dissolved oxygen concentrations, induction times and residual nutrient concentrations, the optimization of the composition of growth medium, which can have a dramatic impact on the performance of the microbial strains, as well as any processing considerations such as purification methods, raw materials or substrate sources, or process logistics.

The present document outlines bioprocess research in both of the aforementioned categories performed in the framework of a Ph.D. research program in the Department of Chemical and Biological Engineering at the University of Ottawa, under the supervision of Dr. Jason Zhang.

Taken as a whole, this research program's main objectives were the development and characterization of novel recombinant protein production platforms, with particular focus on microalgae (*Chlamydomonas reinhardtii*), yeast (*Pichia pastoris*) and bacteria (*Escherichia coli*), and the optimization and characterization of *P. pastoris* bioreactor-based fermentation bioprocesses. Specific aspects of this research are individually described in their corresponding chapters, allowing the readers interested in a particular topic to independently consult the relevant section, and those more interested in a broad range of bioprocess engineering approaches to refer to the work as a whole.

This document is divided into eight chapters, including this introduction, each prepared as a research article, as well as two appendices with supplementary information. Chapters 2 and 3 provide in-depth literature reviews of relevant fields, Chapters 4 and 5 discuss work performed on upstream aspects of recombinant protein production, and Chapters 5 to 7 are focused on downstream aspects of yeast-based systems.

Chapters 2 and 3 consist of literature reviews regarding the current status of the fields of interest of this work, with the former consisting of a literature review and discussion of recent progress in the use of microalgae for the production of recombinant proteins, oils and other valuable metabolites, and the latter of an extensive literature review on the bioprocessing of *P. pastoris* for the production of recombinant proteins, both previously published and updated as needed.

Chapters 4 and 5 describe the work associated with the development and improvement of upstream process considerations. Chapter 4 describes the work done towards the development and characterization of recombinant strains of the microalgae *C. reinhardtii* producing phytase or xylanase, to be used as value-added animal feed additives or as general protein production platforms. Chapter 5 describes 'bio-prospecting' work in which the microflora of the digestive tract of a Canadian beaver (*Castor Canadensis*) was screened for cellulase-producing microorganisms, and the sequence of a novel cellulase gene was isolated and analyzed.

Chapters 6 and 7 describe the work performed towards the development, optimization and characterization of bioreactor-based fermentation systems for the pGAP-mediated production of recombinant phytase with *P. pastoris*. Chapter 6 describes the systematic optimization of BSM growth medium to increase extracellular recombinant phytase yields, while Chapter 7 describes

the development and characterization of novel systems for the cultivation of recombinant *P. pastoris* that use unconventional carbon sources.

The final chapter of this work summarizes conclusions, and offers recommendations regarding possible future directions of interest for this research. All material is appropriately referenced in its respective section. The appendices provide additional information related to the work described in Chapter 4, with full sequences for all designed gene constructs found in Appendix A.1, and a full list of oligonucleotides used in this work found in Appendix A.2.

Chapter 2 – Literature review on the development and use of transgenic microalgae for the production of recombinant proteins

Authors: Potvin G and Zhang Z

The following chapter consists of the paper *Strategies for high-level recombinant protein expression in transgenic microalgae: A review*, published in Biotechnology Advances (Potvin and Zhang, 2010), with some minor edits to language and formatting, updated references and information where needed to preserve the relevance and accuracy of this work, and an additional section (2.8) providing an update on the progress achieved and emerging trends in this field since its publication. It provides an overview of available transformation methods, approaches shown to be effective for the development of high-yield microalgal (primarily *Chlamydomonas reinhardtii*) strains, and areas of interest for future research. To this day, although progress continues, and the pool of strategies at one's disposal to increase recombinant protein expression in microalgae continues to expand, the development of standard, effective and consistent methods or protocols for the development of high-level recombinant production strains, analogous to those, for example, that are available for well-established systems such as *Escherichia coli*, remains an elusive objective of this field.

2.1 Abstract

Microalgae represent the "best of both worlds" as expression platforms, combining the high growth rate and ease of cultivation of microorganisms with the ability to perform post-translational processing of plants. The development of economically viable microalgal expression systems is, however, hindered by generally low recombinant protein yields. Although there are still many obstacles to overcome before microalgae become standard expression systems, considerable progress has been made in recent years in regards to elucidating the causes for these low yields and in the development of strategies to improve them. Transgenes have successfully been expressed in both nuclear and chloroplast microalgal genomes, although at economically viable levels only in the latter. The present review describes recent progress in genetic manipulation of microalgae, outlines strategies to increase protein yields and presents some interesting avenues of research that remain to be explored and emerging trends in the field.

2.2 Introduction

Harnessing the benefits of microalgae is by no means a novel concept. Historical records recount that Chinese civilizations used *Nostoc*, *Arthrospira* and *Aphanizomenon* blue-green algae as alternative food sources to stave off famine as early as 2000 years ago, a strategy later employed by Aztec civilizations in the 14^{th}-16^{th} centuries. The large-scale cultivation of microalgae, however, only began shortly after WWII when the United States, Japan and Germany were facing dwindling food reserves and were forced to consider alternate sources of protein for booming populations. The current world production of raw microalgal biomass exceeds 5000 tonnes, generating an estimated 1-1.25 billion USD in revenue every year (Pulz and Gross 2004, Walker et al. 2005a, Spolaore et al. 2006), with production set to grow exponentially in the next decade.

Today microalgae are still mainly used as nutritional supplements for human and animal consumption due to their high protein and vitamin contents. They have, however, also elicited interest as production systems for a host of valuable naturally-produced compounds (Harun et al. 2010). The most notable of these applications is the cultivation of the green algae *Haematococcus pluvialis* for its high-level (50 mg/g DCW) production of astaxanthin, a carotenoid pigment widely used as a feed additive in fish aquaculture, and prized by the pharmaceutical and cosmetic industries for its antioxidant properties (Hyunsuk et al. 2005, Kathiresan and Sarada 2009).

Given their relatively high oil content, microalgae are also of considerable interest for the biodiesel industry (Williams and Laurens 2010). Although microalgal oil contents as high as 75% DCW have been reported, such contents are highly unusual and obtained only under specific cultivation conditions, and therefore do not apply to current industrial systems. The microalgal systems used in practical applications typically have oil contents closer to 20-40% DCW. It has been argued that microalgae may present the only economically and environmentally viable feedstock for biofuel production. Estimates based on oil contents of 70% DCW, which some believe to eventually be attainable in industrial processes as research progresses, place the cost of microalgae-derived biodiesel at 0.72 USD/L, which currently exceeds the viability of alternate biofuel feedstocks (Chisti 2007). Considerable progress has been achieved in regards to increasing the oil yield of microalgae (Courchesne et al. 2009), and several reviews on microalgae-based biodiesel production have recently been published (Gressel

2008, Li et al. 2008, Meng et al. 2009, Sialve et al. 2009, Brennan and Owende 2010, Greenwell et al. 2010, Mata et al. 2010, Smith et al. 2010, Williams and Laurens 2010, Gong and Jiang 2011, Zhu et al. 2013, Rashid et al. 2014, Ho et al. 2014). Microalgae cultivation methods are reviewed by Carvalho et al. (2006), Eriksen (2008) and Xu et al. (2009).

In addition to their value as platforms for the production of naturally-produced compounds, transgenic microalgae have been eliciting considerable interest over the last decade as recombinant protein expression systems, as they combine the rapid growth and ease of cultivation inherent to many microorganisms with the ability of plant cells to perform post-transcriptional and post-translational modifications. Research into transgenic microalgae is fuelled by the worldwide demand for recombinant proteins and other bioproducts, the market for which is growing exponentially. This market was valued at over 100 billion USD at the end of 2010, with no signs of slowing down (Walsh 2010, Sanchez and Demain 2011).

2.3 Current State of Bioreactor Technology

Bacterial and yeast-based bioreactors are the most widely used systems for the production of recombinant proteins, as they are very well characterized, their genomes easy to manipulate, and their cultivation simple and inexpensive (Swartz 2001). Bacteria, however, do not perform the post-transcriptional and post-translational modifications, including glycosylation, phosphorylation and disulfide bond formation, required for the correct folding and assembly of more complex proteins. Although eukaryotic yeasts can perform these modifications, their profiles are often unsuitable for therapeutic proteins destined for animal or human consumption. Recombinant proteins in yeast are usually hyperglycosylated, which alters immunogenic epitopes, and the high-mannose glycosylation performed in such systems results in low *in vivo* half-life of proteins, both factors which compromise the therapeutic activity of the products (Wildt and Gerngross 2005).

To overcome these difficulties, mammalian, insect, or plant cell bioreactors are used. Several recombinant eukaryotic proteins have been correctly synthesized, processed, and harvested in these cell-based reactors, and have been approved for use in therapeutic applications (Wurm 2004, Boehm 2007, Weathers et al. 2010). Mammalian cell-based bioreactors are, however, very expensive to develop and maintain, and are plagued by complex nutrient requirements, poor oxygen and nutrient distribution, waste accumulation, contamination by pathogens, and high

sensitivity of cells to shear stress (Wurm 2004, Zhang et al. 2010). Compared to mammalian cells, insect cells are easier to culture, are more tolerant to osmolarity changes and by-product accumulation, and baculovirus-infected insect cells generally lead to higher recombinant protein expression levels. They do, however, also have complex nutrient requirements, and baculovirus-infection systems, due to their lytic operating mechanism, lead to the release of endogenous proteases that significantly lower product yields (Ikonomou et al. 2003).

Plant-based reactors, although much less expensive than their mammalian and insect counterparts and resistant to most animal-infecting pathogens, have slow growth cycles and are linked with concerns relating to environmental contamination by genetically modified plants. Although differences are observed in glycosylation profiles between animal and plant cells, in many cases the stability, correct folding and resistance to proteases of mammalian-derived products expressed in plants are not significantly affected. Plant-derived proteoglycans have, however, been linked to allergic reactions, which is a major concern for recombinant proteins destined for human therapeutic applications (Decker and Reski 2004).

Although transgenic microalgal technology is still in its relative infancy, microalgae may represent the "best of both worlds" by combining the simple and inexpensive growth requirements and capabilities for post-transcriptional and -translational processing of plants, with the rapid growth rate and potential for high-density culture of microorganisms (Walker et al. 2005a). Unicellular photosynthetic green algae are most commonly used for protein production as they only require inexpensive salt-based media, carbon dioxide and light for growth. Most green algae are also classified as GRAS (Generally Regarded As Safe), making purification and processing of expressed products much less onerous for many targeted applications. Contrary to transgenic plants which must be strictly contained to avoid the transfer of transgenic material to surrounding wild-type flora by airborne vectors, microalgae can be cultivated in open facilities as no such transfer can occur and contamination is of very little concern. On the economics side, based on recombinant antibody production studies, the cost of production per gram of functional antibody is $150, $0.05 and $0.002 (USD) in mammalian, plant and microalgal bioreactor systems respectively, making the latter system very economically attractive (Mayfield et al. 2003). A comparison of different recombinant protein expression systems is shown in Table 2.1.

Despite the recent surge of interest and successful transformation of a myriad of microalgal species, transgenic strains belonging to the *Chlamydomonas, Chlorella, Volvox, Haematococcus*

and *Dunaliella* genera remain the most widely used and studied (Griesbeck et al. 2006, Raja et al. 2008, Rosenberg et al. 2008, Purton et al. 2014, Rasala and Mayfield 2015), and many obstacles remain to be overcome before microalgae can be considered standard expression systems. The large majority of current work is performed with *Chlamydomonas reinhardtii*, as it is the best characterized of the microalgal species, and more recently in *Dunaliella* halotolerant strains (Rasala and Mayfield 2015). Considerable progress has been made in metabolic engineering towards increasing the expression of naturally produced compounds, with varying levels of success (Rosenberg et al. 2008, Blatti et al. 2013, Bellou et al. 2014, Dubini and Ghirardi 2015). Although recombinant protein production is notably hindered by low expression levels, the continuing development of genetic engineering tools for microalgae has allowed the expression of fully functional antibodies (Franklin and Mayfield 2005, Tran et al. 2009), therapeutics (Boehm 2007, Weathers et al. 2010), and bactericides (Li and Tsai 2009) at economically viable levels. Despite this progress, however, no wide-ranging system or protocol leading to high-level expression has been established. Please refer to section 2.8 for a more detailed update on the current state of this technology.

Table 2.1. Comparison of different recombinant protein expression systems.

	System Characteristics							
	Molecular				Operational			
System	**Glycosylation**	**Gene Size**	**Sensitivity to Shear Stress**	**Recombinant Product Yield**	**Production Time**	**Cost of Cultivation**	**Scale-up Costs**	**Cost for Storage**
Bacteria	None	Unknown	Medium	Medium	Short	Medium	High	Low (-20ºC)
Yeast	Incorrect	Unknown	Medium	High	Medium	Medium	High	Low (-20ºC)
Insect[a]	Correct, but depends on strain and product	Limited	High	Medium to High	Long	High	High	High (Liq. N_2)
Mammalian Cells	Correct	Limited	High	Medium to High	Long	High	High	High (Liq. N_2)
Plant Cells	Correct[b]	Unlimited	N/A	High	Long	Low	Very Low	Low (Room Temp.)
Unicellular Microalgae	Correct[b]	Unlimited	Low	Generally Low	Short	Very Low	Low	Low (Room Temp.)

[a] (Ikonomou 2003, Tomiya 2009)

[b] Nuclear expressed proteins only, no glycosylation in chloroplast

In the following sections, transformation methods for both nuclear and chloroplast microalgal genomes, strategies to increase recombinant protein yields, and potential research directions of interest and emerging trends are reviewed.

2.4 Genetic Transformation Methods

The relatively recent study of transgenic microalgae was only made possible following the development of efficient techniques for the delivery of DNA to target microalgal genomes. These transformation methods have been briefly reviewed before (León-Bañares et al. 2004, Walker et al. 2005a, Griesbeck et al. 2006). Beyond targeted application-based optimization, these methods have not changed significantly since their initial development, and are still being applied in recent work. An overview of these available methodologies is compiled here.

2.4.1 Cell Wall Deficient Strains

Successful transformations were achieved in wild-type microalgae using all the methods described below, although often at very low efficiencies. The use of cell wall-deficient strains, or the removal of the cell walls from wild-type strains, greatly increases the number of transformants recovered following transformation.

Protocols for cell wall removal have been developed which facilitate the study of microalgae. These protocols involve the mating of mating type plus (mt+) and mating type minus (mt-) gametes of *C. reinhardtii*. The specific cell-cell recognition resulting from flagellar interaction leads to the release of enzymes, autolysin or lysin, which cause cell wall degradation. These enzymes can be purified and used as a pre-treatment to transformation. A detailed protocol for production and purification of these enzymes is given by Buchanan and Snell (1998) and a detailed study of the mating process was more recently reported by Hoffmann and Beck (2005).

2.4.2 Particle Bombardment

Bombardment of target cells with DNA-coated metallic particles is a widespread, simple, effective and highly reproducible transformation method. This method has been successfully employed for the transformation of most standard cellular expression systems, and it is therefore not surprising that it is also useful for the study of microalgae. The main drawback of the particle bombardment method is the cost of the required specialized equipment and consumables.

Although the number of transformants recovered following particle bombardment can be low, it remains the most effective method for the transformation of chloroplasts, as it allows for the delivery of multiple copies of recombinant DNA through both the cellular and chloroplast membranes, increasing the chance for a successful integration event to occur (Boynton and Gillham 1993).

This method has been shown to be effective for the stable nuclear (Mayfield and Kindle 1990) and chloroplast (Boynton et al. 1988, El-Sheekh 2000) transformation of *C. reinhardtii*, the transformation of *Volvox carteri* (Schiedlmeier et al. 1994), *Chlorella sorokiana* (Dawson et al. 1997), *ellipsoidea* (Chen et al. 1998) and *kessleri* (El-Sheekh 1999) species, transient transformation of *H. pluvialis* (Teng et al. 2002) and the stable nuclear transformation of the diatom *Phaeodactylum tricornutum* (Apt et al. 1996). Recent work has shown that the particle bombardment method is also effective for the transformation of more complex algal species, such as the multi-cellular *Gonium pectorale* (Lerche and Hallmann 2009).

2.4.3 Glass Beads Method

A simple and effective transformation method consists of agitating cell wall-deficient microalgal cells with recombinant DNA, polyethylene glycol (PEG), which greatly increases transformation efficiency, and glass beads. Despite the drop in cell viability to 25% following agitation with the beads, a nuclear transformation efficiency of 10^3 transformants/μg DNA was achieved using this method (Kindle 1990) and an efficiency of 50 transformants/μg DNA was achieved for the transformation of *C. reinhardtii* chloroplasts (Kindle et al. 1990). Compared to the particle bombardment method, the glass beads method is simpler, more efficient for nuclear transformations, and much less expensive as it does not require specialized equipment. A recent study showed that the glass beads method is also more efficient than particle bombardment for the transformation of *Dunaliella salina* (Feng et al. 2009).

A similar protocol, using silicon carbon whiskers instead of glass beads to pierce cells, has also successfully been used (Dunahay 1993, Wang et al. 1995). The cell viability following agitation is much improved, but due to low transformation efficiencies, high cost of materials, and health and safety concerns associated with the handling of the whiskers, the glass beads are generally preferred.

2.4.4 Electroporation

The effectiveness of microalgal electroporation, or the induction of macromolecular uptake by exposing cell walls to high intensity electrical field pulses, was first reported by Brown et al. (1991). Electroporation specifically disrupts lipid bilayers, leading to efficient molecular transport across the plasma membrane (Azencott et al. 2007). Efficient electroporation-mediated transformation was achieved in both wild-type and cell wall deficient strains (Brown et al. 1991). The transformation efficiency of electroporation is two orders of magnitude higher than the glass beads method, and only requires relatively simple equipment (Shimogawara et al. 1998). Important parameters affecting the effectiveness of electroporation include field strength, pulse length, medium composition, temperature and membrane characteristics (Brown 1991) as well as the concentration of DNA (Wang et al. 2007a).

Electroporation was successfully used for the transformation of *Dunaliella salina* (Geng et al. 2004, Sun et al. 2005, Wang et al. 2007b, Sun et al. 2008, Feng et al. 2009, Feng et al. 2013), *viridis* (Sun et al. 2006) and *tertiolecta* (Walker et al. 2005b) species, *C. reinhardtii* (Tang et al. 1995, Shimogawara et al. 1998, Kovar et al. 2002, Ladygin 2003, Ladygin 2004, Jeon et al. 2013), *Chlorella* species (Chow and Tung 1999, Wang et al. 2007a, Liu et al. 2013), and *Nannochloropsis* species (Chen et al. 2008, Li and Tsai 2008, Kilian et al. 2011, Yu et al. 2014, Li et al. 2014), *Scenedmus* species (Guo et al. 2013), *Karenia* species (Bahi et al. 2011) and *Phaeodactylum tricornutum* (Zhang et al. 2014).

2.4.5 Agrobacterium tumefaciens-mediated transformation

Transformation by the tumour-inducing *Agrobacterium tumefaciens* is another efficient means of delivering genetic material, although this method has so far been mainly used to modify plant cells. Transformation results from the stimulation of cell division by products encoded by T-DNA transferred from *Agrobacterium* to the target cell. The T-DNA and virulence (*vir*) regions, are located on the tumour inducing plasmid (pTi). The *vir* system processes and transfers any DNA between the short flanking repeats that delimit the T-DNA, making *Agrobacterium* an efficient DNA delivery system (Akhond and Machray 2009).

Using the *Agrobacterium*-mediated transformation method, *C. reinhardtii* was successfully transformed with *uidA* (β-glucuronidase), *gfp* (Green Fluorescent Protein) and *hpt* (hygromycin phosphotransferase) reporter genes, with a fifty-fold increase in resulting transformants

compared to the glass beads method (Kumar et al. 2004). GUS (β-glucuronidase), *gfp* and *hpt* genes were also successfully integrated in *H. pluvialis* using this method (Kathiresan et al. 2009). Although results are currently sparse regarding transformation of microalgae with *Agrobacterium*, the method holds considerable promise, and given its success in plant transformations warrants further study. A detailed protocol for the transformation of *H. pluvialis* was devised and studied by Kathiresan and Sarada (2009). More recently, protocols for the efficient transformation of other microalgae species from the *Parachlorella* (Rathod et al. 2013), *Chlorella*, *Ankistrodesmus* and *Scenedesmus* (Sanitha et al. 2013), *Chlamydomonas* (Praheesh et al. 2014) and *Isochrysis* (Prasad et al. 2014) families were developed.

2.4.6 Gene Copy Number

As further discussed in subsequent sections, expression levels of foreign genes in microalgae, in addition to generally being low, are inconsistent and difficult to predict. A significant part of this variation in expression levels arises from inconsistencies in the number of transgene copies integrated within a particular genome. Multiple gene integration patterns, notably complex ones such as inverted or tandem repeats, are associated with low-level transgene expression. The reduction in expression levels is due to homology-dependent gene silencing (HDGS), which affects expression when multiple copies are present, whether at a single locus or at unlinked sites. Single-copy transformants, which generally have higher and more predictable expression levels, are therefore desirable. Silencing occurs at the transcriptional or post-transcriptional level, and is believed to have arisen as a defense mechanism of plants against viruses and as a means of regulating gene expression. Several reviews on transgene silencing and related expression variation in plants have previously been published (Muskens et al. 2000, Baulcombe 2004, Butaye et al. 2005, Angaji et al. 2010, Teixeira and Colot 2010 Marenkova and Deineko 2010, Tuteja et al. 2012).

The transformation method used may impact the number of integrated transgene copies in target cells, although given the variability of reported results, the number of transgene integrations may depend on numerous factors. Although direct DNA-transfer methods such as particle bombardment or glass bead shearing generally lead to a large number of integrated gene copies which may increase silencing effects, by varying the amount of delivered DNA, single or low number copy integration can be achieved, making these methods fairly flexible (Yao et al.

2006, Jayaraj et al. 2008, Lowe et al. 2009). The number of transgene copies integrated following electroporation-mediated transformation is highly variable, but this transformation method has been shown to lead to low-copy transformants. In addition to the electroporation conditions discussed previously, copy integration also depends on the quality of the target tissue and pretreatments it may undergo prior to transformation (Sorokin et al. 2000). *Agrobacterium*-mediated transformation generally leads to low copy number integration, and leads to a higher proportion of single-copy transformants (Butaye et al. 2005, Wu et al. 2008, Zale et al. 2009, Oltmanns et al. 2010).

2.5 Selection Markers

Transformation protocols are useful only if effective markers are available to select successful transformants. The majority of selectable markers confer a resistance to antibiotics or operate by complementation of metabolic mutants. In the former method, genes conferring antibiotic resistance, whether hybridized to the foreign DNA of interest or co-transformed on a distinct vector, are most commonly used for selection of microalgal transformants. The other strategy relies on metabolic or photosynthetic rescue of microalgal mutants with wild-type gene constructs, with transformant selection protocols based on cultivation conditions. The latter method may be particularly useful for chloroplast transformations, where integration of genetic material occurs by homologous recombination. In such cases, hybrid foreign DNA constructs containing wild-type genes can not only rescue microalgal mutants in which that gene is knocked-out, thus allowing for selection, but specifically targets adjacent regions for foreign DNA integration.

Although lists of selectable markers in microalgae have been compiled in past reviews (León-Bañares et al. 2004, Walker et al. 2005a, Griesbeck et al. 2006), which will not be copied here, novel markers have since been developed.

Phytoene desaturase (PDS), a rate-limiting enzyme involved in the production of carotenoids by green algae, is inhibited by certain herbicides. The induction of point-mutation in PDS makes the transgenic strain resistant to the herbicide norflurazon, and this has been shown to be an effective selectable marker in *H. pluvialis* (Steinbrenner and Sandmann 2006) and *Chlorella zofingiensis* (Huang et al. 2008).

The *ARG9* gene from *C. reinhardtii* encodes a plastid *N*-acetyl ornithine aminotransferase, an enzyme involved in arginine synthesis. The integration of an *ARG9* cassette in the plastid chromosome of the nuclear *arg9* mutant restores arginine prototrophy, making ARG9 a novel selectable marker for plastid transformations (Remacle et al. 2009).

These two markers, in addition to those listed in previously published reviews, are still the most commonly used (Rasala and Mayfield 2015). Progress in the development of a greater number of novel selection markers is, however, one of the areas that is particular needed for the continued advancement of microalgal technology (Purton et al. 2013)

Microalgal chloroplasts present an attractive platform for the expression of recombinant therapeutic or nutritional products at high levels, making marker-free systems highly desirable in such applications. Marker genes are engineered for high level expression for effective protection of cells for selection. Once homoplasmic transformation is achieved, 5-18% of the total soluble protein can consist of marker gene products, which lowers the maximum yield of the target protein. If the recombinant algae are destined for human or animal consumption, unnecessary DNA, including genes conferring resistance to antibiotics, is undesirable. Marker removal can be achieved by homology-based excision, excision by phage site-specific recombinases, transient co-integration of the marker gene or the cotransformation-segregation approach. Excellent reviews of these processes by Day and Goldschmidt-Clermont (2011) and Yau and Stewart (2013) have recently been published.

2.6 Nuclear versus Chloroplast Genomes

Although recombinant protein expression in the microalgal nuclear, mitochondrial and chloroplast genomes has been achieved, commercially viable expression levels (for most applications) have only been reported in the latter, and for relatively simple proteins.

Nuclear expression of foreign proteins remains very low, for reasons that are as yet not well understood. Positional effects, RNA silencing, a prohibitively compact chromatin structure and non-conventional epigenetic effects have been proposed as possible causes. The latter is supported by a recent study describing a protocol for the selection of highly expressed nuclear transgenes following UV-induced mutations of transformed strains. Using this protocol, yields of foreign proteins accounting for 0.2% total soluble protein (TSP) were achieved, which is quite high for nuclear expression (Neupert et al. 2009, Rasala and Mayfield 2014). Like all eukaryotic

genomes, post-transcriptional and -translational processing is performed, and post-translational targeting to specified downstream organelles is possible, making nuclear expression necessary for complex protein expression despite low yields.

Chloroplasts are generally preferred for foreign protein expression in microalgae due to high-expression levels and, contrary to nuclear transformation, the possibility of targeted insertion of sequences by homologous recombination. Contary to other prokaryotic genomes, microalgal chloroplasts are also able to produce chaperone molecules and perform simple post-translational modifications like disulfide bond formation. Based on plant studies, chloroplast proteolytic pathways are limited, and the chloroplast envelope may protect foreign protein from degradation, thus increasing their overall yield (Faye and Daniell 2006). The highest reported protein expression level in chloroplasts is slightly over 10% TSP (Surzycki et al. 2009) although the large majority of yields are < 2% TSP and lower (Manuell et al. 2007, Rasala and Mayfield 2015).

2.7 Factors Affecting Protein Expression and Strategies for its Increase

The development of economically viable microalgal expression systems is currently hindered by low and inconsistent recombinant protein yields. Recent efforts towards yield improvement have concentrated, usually as independent parameters, on the study of promoters, UTR sequences and fusion between native and recombinant peptides in microalgal chloroplasts. The regulation of recombinant protein expression is a complex system consisting of interacting elements. Although the extent of interdependence between different factors is not completely understood, several strategies and mechanisms of particular interest have been proven to increase recombinant protein yields in microalgae. Although the systematic study of some of these factors has historically been prohibitively time consuming, the ongoing shift from traditional gene isolation from known organisms and subsequent modifications to the *de novo* design and synthesis of genes provides us with new strategic avenues for the design of time-efficient and comprehensive studies. Recent progress and development on key factors affecting recombinant protein yields are reviewed here. The factors discussed here are still the most relevant areas of interest in this field of research, although a more up-to-date list of successfully expressed recombinant proteins is provided in section 2.8.

2.7.1 Codon Optimization

It is well established that the genomes of different organisms, and the different genomes of single organisms, employ codon biases as mechanisms for optimizing and regulating protein expression (Gustafsson et al. 2004). As is the case for most heterologous genes, optimizing the codon usage of microalgae-destined transgenes to reflect this bias increases their expression efficiency by increasing their translation rates, and may decrease their susceptibility to silencing (Heitzer et al. 2007). In prokaryotic genomes, such as those from microalgal chloroplasts, codon bias is the single most important determinant of protein expression (Lithwick and Margalit 2003, Surzycki et al. 2009, Purton et al. 2013), and adjustment of codons in transgenes is necessary for high level (i.e. commercially viable) expression (Franklin et al. 2002, Mayfield et al. 2003, Mayfield and Schultz 2004).

The Codon Adaptation Index (CAI) is used as a quantitative tool to predict heterologous gene expression levels based on their codon usage. As the chloroplast, mitochondrial and nuclear genomes of microalgae may exhibit different codon biases, as is the case of *C. reinhardtii*, genome-specific CAI values should be used for optimal translation.

Underscoring the importance of codon optimization in biotechnological applications, several free software and web applications have recently been developed to estimate CAI values and optimize the codon usage of sequences. E-CAI, (http://genomes.urv.es/CAIcal/E-CAI) determines whether differences in CAI between sequences are significant or arise from biases in G+C or amino-acid composition (Puigbò et al. 2008a). The online OPTIMIZER application (http://genomes.urv.es/OPTIMIZER) optimizes the codon usage of provided sequences using pre-constructed usage tables based on either a ‘one amino acid-one-codon’ basis, Monte-Carlo algorithms or a novel algorithm for optimization with minimal changes (Puigbò et al. 2007). CAI-cal (http://genomes.urv.es/CAIcal) provides an integrated set of tools for the optimization of codon usage (Puigbò et al. 2008b). Some gene design software packages such as *Gene Composer* include gene optimization functionalities (Lorimer et al. 2009). A list of older codon optimization software has been compiled by Villalobos et al. (2006). It should also be noted that all companies providing gene synthesis services use proprietary and sophisticated optimization software that considers the factors described above, while also minimizing the formation of secondary structures in both DNA and RNA constructs. A database of CAI and codon usage

indices for most sequenced species is available online at http://www.kazusa.or.jp/codon/ (Nakamura et al. 2000).

Codon optimization is an effective and necessary step in gene sequence optimization, and one relatively simple to address with recent advances in DNA synthesis technology, but it is not the only factor to be considered. An excellent review by Welch et al. (2009) describes important factors and useful strategies for the *de novo* design of genes optimized for recombinant protein production. It must be noted, however, that although it is clear that using codons that are preferentially used in native genes generally improves translational efficiency, the rules governing the optimal choices are still poorly understood, and effects related to chosen codon pairs or codon repeats may significantly affect expression levels (Franklin et al. 2002, Wu et al. 2011, Purton et al. 2013).

2.7.2 Transformation-associated genotypic modifications

Transgenes are inserted in the chloroplast genome by homologous recombination, which implies that each transformant obtained should be identical if using a single integration vector. Identical recombinant protein expression profiles for each transformant are therefore expected.

Surzycki and coworkers (2009) however, have observed protein yields varying from 0.88 to 20.9 % total cell protein (TCP), the latter being the highest yield reported to date, in transgenic lines obtained from a single biolistic transformation. They associated this variation to genotypic modifications resulting from the transformation process, dubbed *transformosomes*. The observed expression levels of transgenic proteins, which are to date low, may therefore depend more on these modifications than on the selection of promoters, UTRs or insertion sites. Although the mechanisms behind transformosomes are as yet unconfirmed, and indeed somewhat controversial, they may be due to additional insertions of the transgenic gene in the nuclear genome, which would interfere with proteins regulating recombinant yields (Surzycki et al. 2009), or due to HDGS following the insertion of variable numbers of gene copies (see section 2.4.6). Nuclear insertion may also interfere with the expression of chloroplast-bound genes necessary for photosynthesis or with proteins involved in their targeting and transfer. Extensive screening of transformants is thus recommended to isolate the most productive ones for further process development.

2.7.3 Endogenous Enhancer and Regulatory Elements

Inserting introns from native genes in heterologous sequences under the control of that gene's promoter has been shown to increase protein yields. A study by Eichler-Stahlberg et al. (2009) showed that inserting three introns from native *C. reinhardtii* RBCS2 chloroplast gene in recombinant codon-optimized luciferase and erythropoietin increased expression upwards of 400% compared to base levels. Although each individual RBCS intron had a positive effect on expression, their integration in their physiological order and number produced a synergetic effect. Expression of recombinant genes in the nuclear genome of *C. reinhardtii* also improved following the insertion of the first RBCS2 intron, which has been shown to contain an enhancer element (Lumbreras et al. 1998, Berthold et al. 2002).

A number of *C. reinhardtii* chloroplast gene products regulate the translation of their own mRNA through feedback inhibition (Wostrikoff et al. 2004, Minai et al. 2006). This may partially explain the low expression levels of heterologous genes in microalgal chloroplasts compared to tobacco chloroplast expression systems, in which this inhibition is not observed. Manuell et al. (2007) showed that the product of the endogenous *psbA* chloroplast gene, D1 protein, does inhibit the expression of recombinant M-SAA under the control of the psbA 5'UTR. In this case, however, competition between endogenous and recombinant psbA-mediated constructs for limited transcription and translation factors reduced protein accumulation. They report a maximum yield of just over 5% TSP.

The fusion of recombinant products to native proteins has also resulted in an increase of protein yield. Muto et al. (2009) fused the endogenous Rubisco LSU protein to a recombinant luciferase through a cleavable domain. This resulted in a 33-fold increase in luciferase expression compared to luciferase expressed alone, and near-wild-type Rubisco expression levels. Rasala et al. successfully produced and secreted xylanase fused to an endogenous gene fragment and selection marker, linked with a protease cleavage site in the nuclear genome of *C. reinhardtii* (Rasala 2012). These results indicate that recombinant protein accumulation in algal chloroplasts can be enhanced by fusion with a native protein. The usefulness of engineering proteolytic processing sites to liberate recombinant proteins from the native ones has also been demonstrated, which would simplify product purification in such applications.

2.7.4 Sensitivity to Proteases

The level of foreign protein accumulation results from a balance between rates of protein synthesis and degradation, the latter of which is increasingly found to impact recombinant product yields. Proteolytic enzymes, which are essential for endogenous protein processing, may lead to the degradation of foreign proteins after synthesis, or interfere with their correct assembly and post-translational modification. Proteolysis may also lead to inconsistent results and to difficulties in downstream processing or purification due to degraded or non-functional protein fragments. Only limited information on the impact of degradation on yields in microalgae is available, but many studies on proteolysis in plants have been reported. In *C. reinhardtii*, proteolytic degradation is one of the principal factors affecting recombinant protein yield (Surzycki et al. 2009, Purton et al. 2013).

Several strategies are available to minimize proteolytic degradation of foreign proteins in plants, which may also be applicable in microalgal systems. The more interesting strategies are briefly reviewed here, although additional work is required to evaluate their effectiveness in microalgal systems specifically. An excellent review of proteolytic degradation of foreign proteins in plants has been published by Doran (2006).

For nuclear-expressed proteins, degradation can be minimized by targeting protein synthesis to the ER rather than to the cytosol (Conrad and Fiedler 1998), a strategy that led to a 10^4-fold increase in recombinant growth factor expression in tobacco (Wirth et al. 2004). As the plant cell ER contains very few proteases, in applications where protein secretion or modification in the Golgi are not required, proteins can be retained in the ER using KDEL or HDEL sequences, which may have the added effect of enhancing proper folding and stability of certain proteins (Nuttall et al. 2002). Alternatively, the co-expression of protease inhibitors has proven useful in increasing recombinant protein yields in plants, without affecting normal growth and development (Van der Vyver et al. 2003).

Proteins not requiring post-translational modifications beyond disulfide bond formation can be expressed in the chloroplasts of algae. In plant chloroplasts, proteolytic pathways, although present and necessary for processing, are limited, and may interfere less with recombinant protein accumulation (Bock 2001). Chloroplasts could potentially therefore serve as a protective envelope for expressed foreign proteins during long-term storage.

Many proteins do require post-translational modifications not performed in the prokaryotic plastids. The large majority of plant chloroplast proteins are encoded in the nuclear genome and imported from the cytoplasm (Faye and Daniell 2006, Jarvis 2008), which is also the case in microalgae (Specht and Mayfield 2014). A possible strategy to minimize proteolytic degradation of proteins requiring modification would therefore be to target the nuclear-expressed proteins to the chloroplast for storage. The simplified chloroplast protein import mechanisms are illustrated in Figure 2.1. According to the classic, and dominant, mechanism, chloroplast proteins synthesized by cytosolic ribosomes are targeted to the chloroplast through interactions between an N-terminal transit sequence and the *Toc* and *Tic* complexes in the chloroplast membranes (Fig. 2.1a). In addition to the post-translational targeting of proteins to chloroplasts by this mechanism, certain proteins (e.g. CAH1 and NPP1) are also co-translationally synthesized on membrane-bound ribosomes and inserted in the ER through N-terminal signal peptides (Levitan et al. 2005). These proteins can then be targeted to the chloroplast directly or undergo further processing in the Golgi apparatus prior to delivery to the chloroplast via the secretory pathway (Fig. 2.1b) (Villarejo et al. 2005, Radhamony and Theg 2006, Nanjo et al. 2006, Kitajima et al. 2009, Hummel et al. 2010). Although the existence of this mechanism is well supported by experimental data, specific details on endomembrane-mediated chloroplast targeting and import mechanisms remain to be fully elucidated. A more detailed account on this subject is outside the scope of this work, but additional details are provided in excellent reviews by Faye and Daniell (2006), Inaba and Schnell (2008), Jarvis (2008), and Faso et al. (2009).

2.8 Update and Recent Developments

Since this review paper was published in 2010, microalgal technology has continued to develop, with continued progress in the expression of a variety of recombinant proteins, as well as some interesting trends in research directions and product development, briefly outlined here. Preceding sections of this chapter have been updated where necessary to maintain the relevance and accuracy of their contents, and supplementary information is presented here.

2.8.1. Production of recombinant proteins in Chlamydomonas reinhardtii

Over the last few years, several new recombinant proteins have been expressed in both the nuclear and chloroplast genomes of *C. reinhardtii*, bringing the total to approximately 50, and a

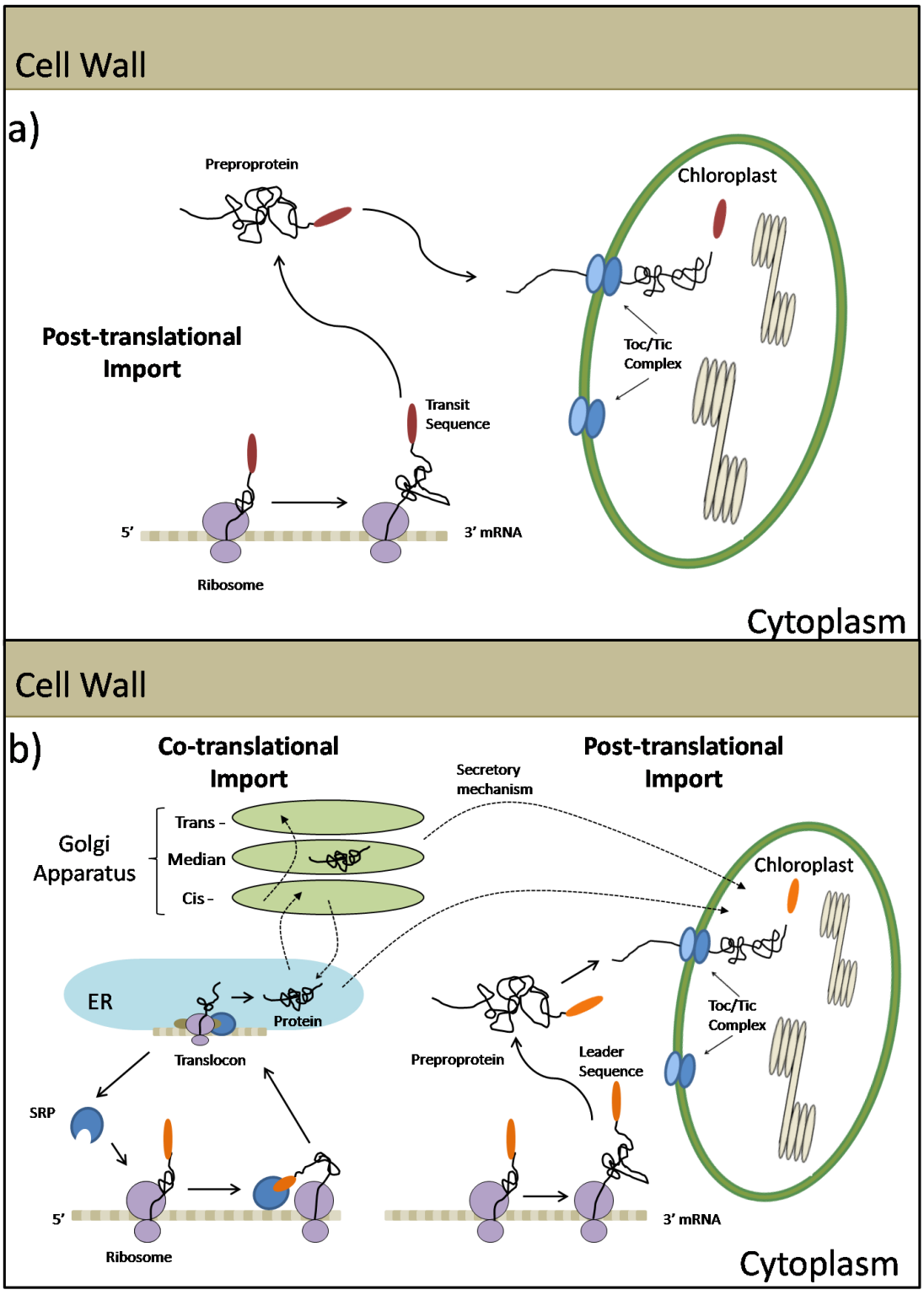

Figure 2.1 *Main chloroplast protein import mechanisms in plants. a) Classic post-translational import mechanism. b) Simultaneous co-translational and post-translational mechanisms. Additional work is required to determine if these pathways are reproduced in microalgae and if they can be utilized for protective storage of recombinant proteins.*

few other species of microalgae have also proved to be viable candidates for the production of recombinant products.

Recombinant products expressed in *C. reinhardtii* chloroplasts include human protein therapeutics and antibodies (VEGF and HMGB1, involved in wound repair, and antibodies – Rasala et al. 2010), components for White Spot Syndrome Virus (WSSV) (Surzycki et al. 2009), *Staphylococcus aureus* (Dreesen et al. 2010), HPV (Demurtas et al. 2013), and malaria (Gregory et al. 2012, Jones et al. 2013) vaccines, and industrial enzymes including phytase and xylanase (Yoon et al. 2011, Georgianna et al. 2013).

A surge of interest for the expression of recombinant products in the nuclear genome has also occurred, leading to the successful nuclear production of erythropoietin (Eichler-Stahlbert et al. 2009), vaccine components (Dauvillée et al. 2010), xylanase (Rasala et al. 2012) and nutritional selenium supplements (Hou et al. 2013). The yields of nuclear-expressed proteins remain extremely low however, with reported levels as low as 3 ng/mg of TCP, and not exceeding 0.2% of TSP, due to gene silencing and positioning effects that remain poorly understood. Promising developments in targeted gene inactivation (Sizova et al. 2013) and gene insertion using CRISPR technology (Belhaj et al. 2013) may help to reduce the impact of these factors in the future.

Although the molecular tools for the manipulation of *C. reinhardtii* are the most abundant by far, several other species of microalgae, most notably *Dunaliella salina* (Feng et al. 2014) and *Dunaliella tertiolecta* (Georgianna et al. 2013) have attracted considerable interest as viable protein expression platforms, being very well studied and also classified as GRAS. *D. salina* in particular, being highly halotolerant, is of interest for the production of proteins that agglomerate and form insoluble inclusion bodies at low salt concentrations. Cultivating recombinant strains in highly saline environments can significantly increase the solubility of these proteins (Akbari et al. 2014). A comprehensive review of the different aspects involved in strain engineering (as opposed to recombinant protein production), which is outside the direct scope of this work, is provided by Purton et al. (2013).

2.8.2. Recent Trends and Future Research Directions

Although, as previously discussed, microalgae have the potential to be very versatile protein production platforms, two notable specific areas of interest have emerged in recent years, namely

the use of microalgae for the production of immunotoxins and antibodies, and as the basis for oral vaccines.

2.8.2.1 Immunotoxins

Microalgal chloroplasts have genomes that though prokaryotic, also express chaperone molecules and are able to perform simple post-translational modifications such as disulfide bond formation. This allows for the expression of immunotoxins, a class of molecules composed of eukaryotic toxins linked to endocytosing antibodies, and which allow for the targeted killing of specific cell types, a characteristic of particular interest for the development of anti-cancer treatments. These endotoxins cannot, however, be produced in conventional mammalian or yeast systems, and they specifically target eukaryotic protein translation machinery, which prevents cell proliferation, and prokaryotic systems such as *E.coli* cannot properly fold the proteins, which therefore require expensive and complicated *ex vivo* processing (Rasala and Mayfield, 2013). Microalgal chloroplasts, however, provides a prokaryotic environment with the necessary post-translation processing, which would allow for the production of these immunotoxins. This was successfully demonstrated as a proof of concept experiment through the expression of CD22, a B-cell surface protein, linked to exotoxins from *Gelonium multiflorum* or *Pseudomonas aeruginosa*, both of which significantly inhibited tumor growth and improved the survival of a mice model (Tran et al. 2013ab). The viable expression of these properly-folded cytotoxic protein hybrids opens up some promising avenues of research.

2.8.2.2 Microalgae as vaccines

Given the advantageous properties of microalgal chloroplasts described in the previous section, microalgae have generated considerable interest as platforms for the production of vaccine components and antibodies (see section 2.8.1), and microalgal lysates have successfully been demonstrated as potential vaccines both through subcutaneous and oral administration.

Oral vaccination has considerable advantages over more conventional parenteral administration methods, including the capacity to elicit both systemic and mucosal immune responses, the removal of the need for trained medical personnel to administer the vaccine, the potential for much lower costs, and less resistance from the general population to treatment. The approval of effective, unprocessed, microalgae-based vaccines for common illnesses would also

make the production and distribution of these treatments simple and inexpensive, which would have tremendous benefits for health around the world, particularly in poorer countries.

The feasibility of such vaccines has been demonstrated in a number of proof-of-concept experiments. A WSSV vaccine antigen, VP28, was successfully expressed in *D. salina*. When these microalgae were orally administered to crayfish, their survival rate significantly increased when contracting this disease (Feng et al. 2013). Oral administration of a *S. aureus* protein-expressing *C. reinhardtii* strain to mice elicited a strong immune response and protected the mice against lethal doses of the bacterium (Dreesen et al. 2010). Malaria is responsible for several hundreds of thousands of deaths annually. The oral administration of recombinant *C. reinhardtii* expressing malaria antigens to mice proved effective at protecting them against the disease and significantly improved their survival rates. Although it is too early for any algae-based oral vaccines to be approved for distribution, these preliminary results are encouraging and will no doubt lead to interesting research. A comprehensive review of algal vaccines is outside the scope of this work, but the interested reader may refer to the excellent one published by Specht and Mayfield (2014).

2.9 Conclusions and Future Prospects

The complex interplay of many factors, including enhancer elements, regulatory mechanisms, competition for available transcription and translation factors, codon dependency, transformation-associated events, sensitivity to proteases, and protein localization and gene silencing, underscores the difficulties in establishing a standard system for recombinant protein production in microalgae. These difficulties are further emphasized as attempts to transfer transgenic technology to new and more unconventional microalgal species are undertaken. Although the systematic study and optimization of some of these parameters is, sometimes prohibitively, difficult and time-consuming using traditional methods, as the *de novo* gene design and synthesis become less expensive and more readily available, the design of more efficient and comprehensive experiments will become possible.

The factors discussed in the present paper deal with the 'upstream' side of process engineering and design. Once transgenic strains are developed, the cultivation operating parameters, including growth media composition, temperature, pH, CO_2 concentration, agitation and illumination, can also significantly affect protein yields, whether directly through molecular

effects on cellular mechanisms, or indirectly through their effect on microalgal growth and maximum supportable cell density. Cooperation among researchers on both upstream and downstream bioprocess design will further facilitate the development of economically viable transgenic microalgae-based bioreactors.

2.10 References

Akbari F, Eskandani M, Khosroushahi AY. The potential of transgenic green microalgae: a robust photobioreactor to produce recombinant therapeutic proteins. World J. Microbiol. Biotechnol. 2014; 30, 2783-96

Akhond MAY, Machray GC. Biotech crops: technologies, achievements and prospects. Euphytica. 2009; 166, 47-59

Angaji SA, Hedayati SS, Poor RH, Poor SS, Shiravi S, Madani S. Application of RNA interference in plants. Plant Omics. 2010; 3, 77-84.

Apt KE, Kroth-Pancic PG, Grossman AR. Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. Mol. Gen. Genet. 1996; 252, 572-9

Azencott HR, Peter GF, Prausnitz MR. Influence of the Cell Wall on Intracellular Delivery to Algal Cells by Electroporation and Sonication. Ultrasound in Med. Biol. 2007; 33, 1805-17

Bahi MM, Tsaloglou MN, Mowlem M, Morgan H. Electroporation and lysis of marine microalga *Karenia brevis* for RNA extraction and amplification. J. Royal Society Interface. 2011; 8, 601-8.

Baulcombe D. RNA silencing in plants. Nature. 2004; 431, 356-63.

Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods. 2013; 9, 39

Bellou S, Baeshen MN, Elazzazy AM, Aggeli D, Sayegh F, Aggelis G. Microalgal lipids biochemistry and biotechnological perspectives. Biotechnol. Adv. 2014; 32, 1476-93

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

Blatti JL, Michaud J, Burkart MD. Engineering fatty acid biosynthesis in microalgae for sustainable biodiesel. Curr. Opinion Chem. Biol. 2013; 17, 496-505.

Bock R. Transgenic plastids in basic research and plant biotechnology. J. Mol. Biol. 2001; 312, 425-38

Boehm R. Bioproduction of Therapeutic Proteins in the 21st Century and the Role of Plants and Plant Cells as Production Platforms. Ann. N.Y. Acad. Sci. 2007; 1102, 121-34

Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, et al. Chloroplast Transformation in *Chlamydomonas* with High Velocity Microprojectiles. Sci. 1988; 240, 1534-8

Boynton JE, Gillham NW. Chloroplast Transformation in Chlamydomonas. Methods Enzymol. 1993; 217, 510-36.

Brennan L, Owende P. Biofuels from microalgae – A review of technologies for production, processing and extractions of biofuels and co-products. Renewable Sustainable Energy Rev. 2010; 14, 557-77.

Brown LE, Sprecher SL, Keller LR. Introduction of Exogenous DNA into *Chlamydomonas reinhardtii* by electroporation. Mol. Cell. Biol. 1991; 11, 2328-32

Buchanan MJ, Snell WJ. Biochemical Studies on Lysin, a Cell Wall Degrading Enzyme Released During Fertilization in *Chlamydomonas*. Exp. Cell. Res. 1988; 179, 181-93

Butaye KMJ, Cammue BPA, Delauré SL, De Bolle MFC. Approaches to minimize variation of transgene expression in plants. Mol. Breeding. 2005; 16, 79-91.

Carvalho AP, Meireles LA, Malcata FX. Microalgal Reactors: A Review of Enclosed System Designs and Performances. Biotechnol. Prog. 2006; 22, 1490-506

Chen Y, Li WB, Bai QH, Sun YR. Study on transient expression of GUS gene in *Chlorella ellipsoidea* (Chlorophyta), by using biolistic particle delivery system. Chin. J. Oceanol. Limnol. 1998; 16 Suppl, 47-9

Chen HL, Li SS, Huang R, Tsai HJ. Conditional production of a functional fish growth hormone in the transgenic line of *Nannochloropsis oculata* (Eustigmatophyceae). J. Phycol. 2008; 44, 768-76.

Chisti Y. Biodiesel from Microalgae. Biotechnol. Adv., 2007; 25, 294-306

Chow KC, Tung WL. Electrotransformation of *Chlorella vulgaris*. Plant Cell Rep. 1999; 18, 778-80

Conrad U, Fiedler U. Compartment-specific accumulation of recombinant immunoglobulins in plant cells: an essential tool for antibody production and immunomodulation of physiological functions and pathogen activity. Plant Mol. Biol. 1998; 38, 101-9

Courchesne NMD, Parisien A, Wang B, Lan CQ. Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. J. Biotechnol. 2009; 141, 31-41

Dawson HN, Burlingame R, Cannons AC. Stable Transformation of *Chlorella*: Rescue of Nitrate Reductase-Deficient Mutants with the Nitrate Reductase Gene. Curr. Microbiol. 1997; 35, 356-62

Day A, Goldschmidt-Clermont M. The chloroplast transformation toolbox: selectable markers and marker removal. Plant Biotechnol. J. 2011; 9, 540-53

Decker EL, Reski R. The moss bioreactor. Curr. Opinion Plant Biol. 2004; 7, 166-70

Demurtas OC, Massa S, Ferrante P, Venuti A, Franconi R, Giuliano G. A *Chlamydomonas*-derived human papillomavirus 16 E7 vaccine induces specific tumor protection. PLoS One. 2013; 8, e61473

Doran PM. Foreign protein degradation and instability in plants and plant tissue cultures. Trends Biotechnol. 2006; 24, 426-32

Dreesen IAJ, Charpin-El Hamri G, Fussenegger M. Heat-stable oral alga-based vaccine protects mice from *Staphylococcus aureus* infection. J. Biotechnol. 2010; 145-273-80

Dubini A, Ghirardi ML. Engineering photosynthetic organisms for the production of biohydrogen. Photosynthesis Res. 2015; 3, 241-53.

Dunahay TG. Transformation of *Chlamydomonas reinhardtii* with Silicon Carbide Whiskers. Biotechniques. 1993; 15, 452-60

Eichler-Stahlberg A, Weisheit W, Ruecker O, Heitzer M. Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. Planta. 2009; 229, 873-83.

El-Sheekh MM. Stable transformation of the intact cells of *Chlorella kessleri* with high velocity microprojectiles. Biologia Plantarum. 1999; 42, 209-16

El Sheekh MM. Stable Chloroplast Transformation in *Chlamydomonas reinhardtii* using Microprojectile Bombardment. Folia Microbiol. 2000; 45, 496-504

Eriksen NT. The technology of microalgal culturing. Biotechnol. Lett. 2008; 30, 1525-36.

Faso C, Boulaflous A, Brandizzi F. The plant Golgi apparatus: Last 10 years of answered and open questions. FEBS Lett. 2009; 583, 3752-7.

Faye L, Daniell H. Novel pathways for glycoprotein import into chloroplasts. Plant Biotechnol. J. 2006; 4, 275-9

Feng S, Xue L, Liu H, Lu P. Improvement of efficiency of genetic transformation for *Dunaliella salina* by glass beads method. Mol. Bio. Rep. 2009; 36, 1433-9

Feng S, Feng W, Zhao L, Gu H, Li Q, Shi K, Guo S, Zhang N. Preparation of transgenic *Dunaliella salina* for immunization against white spot syndrome virus in crayfish. Arch Virol. 2014; 159, 519-25

Franklin S, Ngo B, Efuet E, Mayfield SP. Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. Plant J. 2002; 30, 733-44

Franklin S, Mayfield SP. Recent developments in the production of human therapeutic proteins in eukaryotic algae. Expert Opinion Biol. Therap. 2005; 5, 225-35

Geng DG, Han Y, Wang YQ, Wang P, Zhang LM, Li WB, et al. Construction of a system for the stable expression of foreign genes in *Dunaliella salina*. Acta Bot. Sin. 2004; 46, 342-6

Georgianna DR, Hannon MJ, Marcuschi M, Wu S, Botsch K, Lewis AJ, Hyun J, Mendez M, Mayfield SP. Production of recombinant enzymes in the marine alga *Dunaliella tertiolecta*. Algal Res. 2013; 2, 2-9

Gong YM, Jiang ML. Biodiesel production with microalgae as feedstock: from strains to biodiesel. Biotechnol. Lett. 2011; 1269-84.

Greenwell HC, Laurens LML, Shields RJ, Lovitt RW, Flynn KJ. Placing microalgae on the biofuels priority list: a review of the technological challenges. J. R. Soc. Interface. 2010; 7, 703-26

Gregory JA, Li F, Tomosada LM, Cox CJ, Topol AB, Vinetz JM, Mayfield SP. Algae-produced Pfs25 elicits antibodies that inhibit malaria transmission. PLoS One. 2012; 7, e37179

Gressel J. Transgenics are Imperative for Biodiesel Crops. Plant Sci. 2008; 174, 246-63

Griesbeck C, Kobl I, Heitzer M. *Chlamydomonas reinhardtii*. Mol. Biotechnol. 2006; 34, 213-23

Guo SL, Zhao XQ, Tang Y, Wan C, Alam MA, Ho SH, Bai FW, Chang JS. Establishment of an efficient genetic transformation system in *Scenedesmus obliquus*. J. Biotechnol. 2013; 163, 61-8

Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol. 2004; 22, 346-53

Harun R, Singh M, Forde GM, Danquah MK. Bioprocess engineering of microalgae to produce a variety of consumer products. Renewable Sustainable Energy Rev. 2010; 14, 1037-47.

Heitzer M, Eckert A, Fuhrmann M, Griesbeck C. Influence of Codon Bias on the Expression of Foreign Genes in Microalgae. In: León R, Gaván A, Fernández E, editors. Transgenic Microalgae as Green Cell Factories. Landes Bioscience/Springer Science + Business Media, LLC, 2007. p. 46-53

Ho SH, Ye XT, Hasunuma T, Chang JS, Kondo A. Perspectives on engineering strategies for improving biofuel production from microalgae – A critical review. Biotechnol. Adv. 2014; 32, 1448-59

Hoffmann XK, Beck CF. Mating-Induced Shedding of Cell Walls, Removal of Walls from Vegetative Cells and Osmotic Stress Induce Presumed Cell Wall Genes in *Chlamydomonas*. Plant Physiol. 2005; 139, 999-1014.

Hou Q, Qiu S, Liu Q, Tian J, Hu Z, Ni J. Selenoprotein-Transgenic *Chlamydomonas reinhardtii*. Nutrients. 2013; 5, 624-36.

Huang J, Liu J, Li Y, Chen F. Isolation and Characterization of the Phytoene Desaturase Gene as a Potential Selective Marker for Genetic Engineering of the Astaxanthin-Producing Green Alga *Chlorella zofingiensis* (Chlorophyta). J. Phycol. 2008; 44, 684-90

Hummel E, Osterrieder A, Robinson DG, Hawes C. Inhibition of Golgi function causes plastid starch accumulation. J. Exp. Bot. 2010; 61, 2603-14.

Hyunsuk E, Park S, Lee CG, Jin E. Gene expression profiling of Eukaryotic Microalga *Haematococcus pluvialis*. J. Microbiol. Biotechnol. 2005; 15, 1060-6

Ikonomou L, Schneider YJ, Agathos SN. Insect cell culture for industrial production of recombinant proteins. Appl. Microbiol. Biotechnol. 2003; 62, 1-20

Inaba T, Schnell DJ. Protein trafficking to plastids: one theme, many variations. Biochem. J. 2008; 413, 15-28.

Jarvis P. Targeting of nucleus-encoded proteins to chloroplasts in plants. New Phytol. 2008; 179, 257-85.

Jayaraj J, Liang GH, Muthukrishnan S, Punja ZK. Generation of low copy number and stably expressing transgenic creeping bentgrass plants using minimal gene cassette bombardment. Biologia Plantarum. 2008; 52, 215-21.

Jeon K, Suresh A, Kim YC. Highly efficient molecular delivery into *Chlamydomonas reinhardtii* by electroporation. Korean J. Chem. Eng. 2013; 30, 1626-30

Jones CS, Luong T, Hannon M, Tran M, Gregory JA, Shen Z, Briggs SP, Mayfield SP. Heterologous expression of the C-terminal antigenic domain of the malaria vaccine candidate Pfs48/45 in the green algae *Chlamydomonas reinhardtii*. Appl. Microbiol. Biotechnol. 2013; 97, 1987-95.

Kathiresan S, Sarada R. Towards genetic improvement of commercially important microalga *Haematococcus pluvialis* for biotech applications. J. Appl. Phycol. 2009; 21, 553-58

Kathiresan S, Chandrashekar A, Ravishankar A, Sarada R. *Agrobacterium*-mediated transformation in the green alga *Haematococcus pluvialis* (Chlorophyceae volvocales). J. Phycol. 2009; 45, 642-9

Kilian O, Benemann CSE, Niyogi KK, Vick B. High-efficienty homologous recombination in the oil-producing alga *Nannochloropsis sp.* Proc. Nat. Academy Sci USA. 2011; 52, 21265-9.

Kindle KL. High Frequency Nuclear Transformation of *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA. 1990; 87, 1228-32

Kindle KL, Richards KL, Stern DB. Engineering the chloroplast genome: Techniques and capabilities for chloroplast transformation in *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA. 1990; 88, 1721-25

Kitajima A, Asatsuma S, Okada H, Hamada Y, Kaneko K, Nanjo Y, et al. The Rice alpha-Amylase Glycoprotein is Targeted from the Golgi Apparatus through the Secretory pathway to the Plastids. Plant Cell. 2009; 21, 2844-58.

Kovar JL, Zhang J, Funke RP, Weeks DP. Molecular analysis of the acetolactate synthase gene of *Chlamydomonas reinhardtii* and development of a genetically engineered gene as a dominant selectable marker for genetic transformation. Plant J. 2002; 29, 109-117

Kumar SV, Misquitta RW, Reddy VS, Rao BJ, Rajam MV. Genetic transformation of the green alga – *Chlamydomonas reinhardtii* by *Agrobacterium tumefaciens*. Plant Sci. 2004; 166, 731-8

Ladygin VG. The Transformation of the Unicellular Alga *Chlamydomonas reinhardtii* by Electroporation. Microbiol. 2003; 72, 585-91

Ladygin VG. Efficient transformation of mutant cells of *Chlamydomonas reinhardtii* by electroporation. Process Biochem. 2004; 39; 1685-91.

León-Bañares R, González-Ballester, D, Galván A, Fernández E. Transgenic microalgae as green cell-factories. Trends Biotechnol. 2004; 22, 45-52

Lerche K, Hallmann A. Stable nuclear transformation of *Gonium pectorale*. BMC Biotechnol. 2009; 9, 64

Levitan A, Trebitsh T, Kiss V, Pereg Y, Dangoor I, Danan A. Dual targeting of the protein disulfide isomerase RB60 to the chloroplast and the endoplasmic reticulum. Proc. Natl. Acad. Sci. USA. 2005; 102, 6225-30

Li Q, Du W, Liu D. Perspectives of microbial oils for biodiesel production. Appl. Microbiol. Biotechnol. 2008; 80, 749-56.

Li SS, Tsai HJ. Transgenic microalgae as a non-antibiotic bactericide producer to defend against bacterial pathogen infection in the fish digestive tract. Fish Shellfish Immunol. 2009; 26, 316-25

Lithwick G, Margalit H. Hierarchy of sequence-dependent features associated with prokaryotic translation. Genome Res. 2003; 13, 2665-73

Liu LL, Wang YQ, Zhang YC, Chen XY, Zhang P, Ma SW. Development of a new method for genetic transformation of the green alga *Chlorella ellipsoidea*. Mol. Biotechnol. 2013; 54, 211-9

Lorimer D, Raymond A, Walchli J, Mixon M, Barrow A, Wallace E, et al. Gene Composer: database software for protein construct design, codon engineering and gene synthesis. BMC Biotechnol. 2009; 9, 36

Lowe BA, Prakash NS, Way M, Mann MT, Spencer TM, Boddupalli RS. Enhanced single copy integration events in corn via particle bombardment using low quantities of DNA. Transgenic Res. 2009; 18, 831-40.

Lumbreras V, Stevens DR, Purton S. Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. Plant J. 1998; 14, 441-7

Lutz KA, Maliga P. Construction of marker-free transplastomic plants. Curr. Opin. Biotechnol. 2007; 18, 107-14

Manuell AL, Beligni MV, Elder JH, Siefker DT, Tran M, Weber A, et al. Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast. Plant Biotechnol. J. 2007; 5, 402-12

Marenkova TV, Deineko EV. Transcriptional Gene Silencing in Plants. Russian J. Genet. 2010; 46, 511-20.

Mata TM, Martins AA, Caetano NS. Microalgae for biodiesel production and other applications: A review. Renewable Sustainable Energy Rev. 2010; 14, 217-32.

Mayfield SP, Kindle KL. Stable Nuclear Transformation of *Chlamydomonas reinhardtii* by Using a *C. reinhardtii* Gene as the Selectable Marker. Proc. Natl. Acad. Sci. USA. 1990; 87, 2087-91

Mayfield SP, Franklin SE, Lerner RA. Expression and assembly of a fully active antibody in algae. Proc. Natl. Acad. Sci. USA. 2003; 100, 438-42

Mayfield SP, Schultz J. Development of a luciferase reporter gene, luxCt, for *Chlamydomonas reinhardtii* chloroplast. Plant J. 2004; 37, 449-58

Mayfield SP, Manuell AL, Chen S, Wu J, Tran M, Siefker D, et al. *Chlamydomonas reinhardtii* chloroplasts as protein factories. Curr. Opin. Biotechnol. 2007; 18, 126-33

Meng X, Yang JM, Xu X, Zhang L, Nie QJ, Xian M. Biodiesel production from oleaginous microorganisms. Renewable Energy. 2009; 34, 1-5.

Minai L, Wostrikoff K, Wollman FA, Choquet Y. Chloroplast biogenesis of photosystem II cores involves a series of assembly-controlled steps that regulate translations. Plant Cell. 2006; 18, 159-75

Muskens MWM, Vissers APA, Mol JNM, Kooter JM. Role of inverted DNA repeats in transcription and post-transcriptional gene silencing. Plant Mol. Biol. 2000; 43, 243-60.

Muto M, Henry RE, Mayfield SP. Accumulation and processing of a recombinant protein designed as a cleavable fusion to the endogenous Rubisco LSU protein in *Chlamydomonas* chloroplast. BMC Biotechnol. 2009; 9, 26

Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases : status of the year 2000. Nucleic Acids Res. 2000; 28, 292

Nanjo Y, Oka H, Ikarashi N, Kaneko K, Kitajima A, Mitsui T, et al. Rice plastidial N-glycosylated nucleotide pyrophosphatase/phosphodiesterase is transported from the ER-Golgi to the chloroplast through the secretory pathway. Plant Cell. 2006; 18, 2582-92.

Neupert J, Karcher D, Bock R. Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. Plant J. 2009; 57, 1140-50

Nuttall J, Vine N, Hadlington JL, Drake P, Frigerio L, Ma JKC. ER-resident chaperone interactions with recombinant antibodies in transgenic plants. Eur. J. Biochem. 2002; 269, 6042-51

Oltmanns H, Frame B, Lee LY, Johnson S, Li B, Wang K, et al. Generation of Backbone-Free, Low Transgene Copy Plants by Launching T-DNA from the *Agrobacterium* Chromosome. Plant Physiol. 2010; 152, 1158-1166

Pavlou AK, Reichert JM. Recombinant protein therapeutics: success rates, market trends and values to 2010. Nat. Biotechnol. 2004; 22, 1513-9

Purton S, Szaub JB, Wannathong T, Young R, Economou CK. Genetic Engineering of Algal Chloroplasts: Progress and Prospects. Russian J. Plant Physiol. 2013; 60, 491-9

Potvin G, Zhang Z. Strategies for high-level recombinant protein expression in transgenic microalgae: A review. Biotech. Adv. (2010) 28; 910-918.

Prasad B, Vadakedath N, Jeong HJ, General, T, Cho MG, Lein W. *Agrobacterium tumefaciens*-mediated genetic transformation of haptophytes (*Isochrysis* species). Appl. Microbiol. Biotechnol. 2014; 98, 8629-39

Pratheesh, PT, Vineetha M, Kurup GM. An efficient protocol for the *Agrobacterium*-mediated genetic transformation of microalgae *Chlamydomonas reinhardtii*. Mol. Biotehcnol. 2014; 56, 507-15.

Puigbò P, Guzmán E, Romeu A, Garcia-Vallvé S. OPTIMIZER : a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. 2007; 35, W126-31

Puigbò P, Bravo IG, Garcia-Vallvé S. E-CAI : a novel server to estimate an expected value of Codon Adaptation Index (eCAI). BMC Bioinformatics. 2008a; 9, 65

Puigbò P, Bravo IG, Garcia-Vallvé S. CAI-cal : A combined set of tools to assess codon usage adaptation. Biol. Direct. 2008b; 3, 38

Pulz O, Gross W. Valuable products from biotechnology of microalgae. Appl. Microbiol. Biotechnol. 2004; 65, 635-48

Radhamony RN, Theg SM. Evidence for an ER to Golgi to chloroplast protein transport pathway. Trends Cell Biol. 2006; 16, 385-7.

Raja R, Hemaiswarya S, Kumar NA, Sridhar S, Rengasamy R. A Perspective on the Biotechnological Potential of Microalgae. Crit. Rev. Microbiol. 2008; 34, 77-88

Rasala BA, Muto M, Lee PA, Mayfield SP. Production of therapeutic proteins in alagae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. Plant Biotechnol J. 2010; 8, 719-33

Rasala BA, Lee PA, Shen Z, Briggs SP, Mendez M, Mayfield SP. Robust expression and secretion of xylanase 1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing in the FMDV 2A peptide. PLoS One 2012; 7, e43349

Rashid N, Rehman MSU, Sadiq M, Mahmood T, Han JI. Current status, issues and developments in microalgae derived biodiesel production. Renewable Sust. Energy Rev. 2014; 40, 760-78

Rathod J, Prakash G, Pandit R, Lali A. *Agrobacterium*-mediated transformation of promising oil-bearing marine algae *Parachlorella kessleri*. Photosynthesis Res. 2013; 118, 141-6.

Remacle C, Cline S, Boutaffala L, Gabilly S, Larosa V, Barbieri MR, et al. The *ARG9* Gene Encodes the Plastid-Resident *N*-Acetyl Ornithine Aminotransferase in the Green Alga *Chlamydomonas reinhardtii*. Eukaryot. Cell. 2009; 8, 1460-63

Rosenberg JN, Oyler GA, Wilkinson L, Betenbaugh MJ. A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution. Curr Opinion Biotechnol. 2008; 19, 430-6

Sanchez S, Demain AL. Enzymes and bioconversions of industrial, pharmaceutical and biotechnological significance. Org. Process. Res. Dev. 2011; 15, 224-30

Sanitha M, Radha S, Fatima AA, Devi, SG, Ramya M. *Agrobacterium*-mediated transformation of three freshwater microalgal strains. Polish J. Microbiol. 2014; 63, 387-92.

Schiedlmeier B, Schmitt R, Muller W, Kirk MM, Gruber H, Mages W, Kirk DL. Nuclear transformation of *Volvox carteri*. Proc. Natl. Acad. Sci USA. 1994; 91,5080-4

Shimogawara K, Fujiwara S, Grossman A, Usuda H. High-Efficiency Transformation of *Chlamydomonas reinhardtii* by Electroporation. Genetics. 1998; 148, 1821-8

Sialve B, Bernet N, Bernard O. Anaerobic digestion of microalgae as a necessary step to make microalgal biodiesel sustainable. Biotechnol. Adv. 2009; 27, 409-16

Sizova I, Greiner A, Awasthi M, Kateriya S, Hegemann P. Nuclear gene targeting in *Chlamydomonas* using engineered zinc-finger nucleases. Plant J. 2013; 5, 873-82

Smith VH, Sturm BSM, deNoyelles FJ, Billings SA. The ecology of algal biodiesel production. Trends Ecol. Evol. 2010; 25, 301-9

Sorokin AP, Ke XY, Chen DF, Elliott MC. Production of fertile transgenic wheat plants via tissue electroporation. Plant Sci. 2000; 156, 227-33

Specht EA, Mayfield SP. Algae-based oral recombinant vaccines. Frontiers Microbiol. 2014; 5, 60

Spolaore P, Joannis-Cassan C, Duran E, Isambert A. Commercial Applications of Microalgae. J. Biosci. Bioeng. 2006; 101, 87-96

Steinbrenner J, Sandmann G. Transformation of the Green Algae *Haematoccocus pluvialis* with a Phyotene Desaturase for Accelerated Astaxanthin Biosynthesis. Appl. Environ. Microbiol. 2006; 72 , 7477-84

Sun Y, Yang Z, Gao X, Li Q, Zhang Q, Xu Z. Expression of Foreign Genes in *Dunaliella* by Electroporation. Mol. Biotechnol. 2005; 30, 185-92

Sun Y, Gao XS, Li QY, Zhang QQ, Xu ZK. Functional complementation of a nitrate reductase defective mutant of a green alga *Dunaliella viridis* by introducing the nitrate reductase gene. Gene. 2006; 377, 140-9

Sun GH, Zhang XC, Sui ZH, Mao YX. Inhibition of pds gene expression via the RNA interference approach in *Dunaliella salina* (Chlorophyta). Marine Biotechnol. 2008; 10, 219-26.

Surzycki R, Greenham K, Kitayama K, Dibal F, Wagner R, Rochaix JD, et al. Factors effecting expression of vaccines in microalgae. Biol. 2009; 37, 133-8

Swartz JR. Advances in *Escherichia coli* production of therapeutic proteins. Curr. Opin. Biotechnol. 2001; 12, 195-201

Tang DKH, Qiao SY, Wu M. Insertion mutagenesis of *Chlamydomonas reinhardtii* by electroporation and heterologous DNA. Biochem. Mol. Biol. Intl. 1995; 36, 1025-35

Teng C, Qin S, Liu J, Yu D, Liang C, Tseng C. Transient expression of *lacZ* in bombarded unicellular alga *Haematococcus pluvialis*. J. Appl. Phycol. 2002; 14, 495-500

Teixeira FK, Colot V. Repeat elements and the *Arabidopsis* DNA methylation landscape. Heredity. 2010; 105, 14-23.

Tomiya N. Humanization of recombinant glycoproteins expressed in insect cells. Trends Glycosci. Glyc. 2009; 21, 71-86.

Tran M, Zhou B, Pettersson PL, Gonzalez MJ, Mayfield SP. Synthesis and Assembly of a Full Length Human Monoclonal Antibody in Algal Chloroplasts. Biotechnol. Bioeng. 2009; 104, 663-73

Tran M, Van C, Barrera DJ, Pettersson L, Peinado CD, Bui J, Mayfield SP. Production of unique immunotoxin cancer therapeutics in algal chloroplasts. Proc. Natl. Acad. Sci. USA. 2013a; 110, E15-E22

Tran M, Henry RE, Siefker D, Van C, Newkirk G, Kim J, Bui J, Mayfield SP. Production of anti-cancer immunotoxins in algae: Ribosome inactivating proteins as fusion partners. Biotechnol. Bioeng. 2013b; 110, 2826-35

Tuteja N, Verma S, Sahoo RK, Raveendar S, Reddy INBL. Recent advances in development of marker-free transgenic plants: Regulation and biosafety concern. J. Biosci. 2012; 37, 167-97.

Van der Vyver C, Schneidereit J, Driscoll S, Turner J, Kunert K, Foyer CH. Oryzacystatin I expression in transformed tobacco produces a conditional growth phenotype and enhances chilling tolerance. Plant Biotechnol. J. 2003; 1, 101-12

Villalobos A, Ness JE, Gustafsson C, Minshull J, Govindarajan S. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Biotechnol. 2006; 7, 285

Villarejo A, Buren S, Larsson S, Dejardin A, Monne M, Rudhe C, et al. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. Nat. Cell Biol. 2005; 7, 1124-31

Walker TL, Purton S, Becker DK, Collet C. Microalgae as bioreactors. Plant Cell Rep. 2005a; 24, 629-41

Walker TL, Becker DK, Dale JL, Collet C. Towards the development of a nuclear transformation system for *Dunaliella tertiolecta*. J. Appl. Phycol. 2005b; 17, 363-8

Walsh G. Biopharmacetucial benchnmarks 2010. Nat. Biotechnol. 2010; 28, 917-24

Wang K, Drayton P, Frame B, Dunwell J, Thompson J. Whisker-Mediated Plant Transformation: an Alternative Technology. In Vitro Cell. Dev. Biol. 1995; 31, 101-4

Wang C, Wang Y, Su Q, Gao X. Transient Expression of the GUS gene in a Unicellular Marine Green Alga, *Chlorella* sp. *MACC/C95*, via Electroporation. Biotechnol. Bioprocess Eng. 2007a; 12, 180-3

Wang TY, Xue LX, Hou WH, Yang BS, Chai YR, Ji XA, et al. Increased expression of transgene in stably transformed cells of *Dunaliella salina* by matrix attachment regions. Appl. Microbiol. Biotechnol. 2007b; 76, 651-7

Weathers PJ, Towler MJ, Xu J. Bench to batch: advances in plant cell culture for producing useful products. Appl. Microbiol. Biotechnol. 2010; 85, 1339-51

Welch M, Villalobos A, Gustafsson C, Minshull J. You're one in a googol: optimizing genes for protein expression. J. R. Soc. Interface. 2009; 6, S467-76

Wildt S, Gerngross TU. The humanization of N-glycosylation pathways in yeast. Nature Rev. Microbiol. 2005; 3, 119-28

Williams PJL, Laurens LML, Microalgae as biodiesel & biomass feedstocks: Review & analysis of the biochemistry, energetic & economics. Energy Environ. Sci. 2010; 3, 554-90.

Wirth S, Calamante G, Mentaberry A, Bussmann L, Lattanzi M, Baranao L, et al. Expression of active human epidermal growth factor (hEGF) in tobacco plants by integrative and non-integrative systems. Mol. Breed. 2004; 13, 23-35

Wostrikoff K, Girard-Bascou J, Wollman FA, Choquet Y. Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of *Chlamydomonas*. EMBO J. 2004; 23, 2696-705

Wu G, Dress L, Freeland SJ. Optimal encoding rules for synthetic genes: the need for a community effort. Molec. Syst. Biol. 2007; 3, 134

Wu SJ, Wang HH, Li FF, Chen TZ, Zhang J, Jiang YJ, et al. Enhanced Agrobacterium-mediated Transformation of Embryogenic Calli of Upland Cotton via Efficient Selection and Timely Subculture of Somatic Embryos. Plant Mol. Biol. Rep. 2008; 26, 174-85.

Wu S, Xu L, Huang R, Wang Q. Improved biohydrogen production with an expression of codon-optimized *hemH* and *Iba* genes in the chloroplast of *Chlamydomonas reinhardtii*. Bioresour. Tecnol. 2011; 102, 2610-6

Wurm FM. Production of Recombinant Protein Therapeutics in Cultivated Mammalian Cells. Nature Biotechnol., 2004; 22, 1393-8

Xu L, Weathers PJ, Xiong XR, Liu CZ. Microalgal Bioreactors: Challenges and Opportunities. Eng. Life Sci. 2009; 9, 178-89.

Yao Q, Cong L, Chang JL, Li KX, Yang GX, He GY. Low copy number gene transfer and stable expression in a commercial wheat cultivar via particle bombardment. J. Exp. Bot. 2006; 57, 3737-46.

Yau YY, Stewart CN. Less is more: strategies to remove marker genes from transgenic plants. BMC Biotechnol. 2013; 13, 36

Yu GL, Hu HH, Li XB, Kong RQ. Transformation of marine oleaginous alga *Nannochloropsis gaditana* CCAP849/5 and heterologous gene integration. J. Tropical Oceanography. 2014; 33, 72-7.

Yoon SM, Kim SY, Li KF, Yoon BH, Choe S, Kuo MMC. Transgenic microalgae expressing *Escherichia coli* AppA phytase as feed additive to reduce phytate excretion in the manure of young broiler chicks. Appl. Microbiol. Biotechnol. 2001; 91, 553-63

Zale JM, Agarwal S, Loar S, Steber CM. Evidence for stable transformation of wheat by floral dip in *Agrobacterium tumefaciens*. 2009; 28, 903-13.

Zhang CY, Hu HH. High-efficiency nuclear transformation of the diatom *Phaeodactylum tricornutum* by electroporation. Marine Genomics. 2014; 16, 63-6.

Zhang H, Wang W, Quan C, Fan S. Engineering Considerations for Process Development in Mammalian Cell Cultivation. Curr. Pharm. Biotechnol., 2010; 11, 103-12

Zhu LD, Hitunen E, Antila E, Zhong JJ, Yuan ZH, Wang ZM. Microalgal biofuels: Flexible bioenergies for sustainbale development. Renewable Sust. Energy Rev. 2014; 30, 1035-46.

Chapter 3 – Literature review on the use of *Pichia pastoris* for the production of recombinant proteins

Authors: Potvin G, Ahmad A, Zhang Z

The following chapter consists of the paper *Bioprocess engineering aspects of heterologous protein production in Pichia pastoris: A review*, published in the Biochemical Engineering Journal (Potvin et al, 2012), with some minor edits to language and formatting, relevant updates to the material presented, as well as an additional section (3.10) that provides an overview of the work performed in this field since its publication. The original review was quite comprehensive, and although *Pichia pastoris* fermentation remains a commonly used industrial process, the associated bioprocess considerations are well-established, and the different aspects discussed in the original document, though updated where necessary, remain accurate and relevant.

This chapter provides an overview of the most commonly used expression constructs and newly developed systems, strategies for process monitoring, fed-batch and continuous cultivation considerations for AOX1- and pGAP- mediated strains, discusses issues related to proteolytic degradation and stability of recombinant proteins, and discusses modelling of *P. pastoris* cultivation systems.

3.1 Abstract

Pichia pastoris is currently one of the most effective and versatile systems for the expression of heterologous proteins. Its success is due to its powerful methanol-inducible alcohol oxidase 1 (AOX1) promoter, high achievable cell density, capacity to perform post-translational modifications, and pathways leading to the secretion of recombinant products. Despite its advantages, *P. pastoris* cultivation is plagued by high protease expression levels, high sensitivity to methanol levels, nutrient-deficiency when grown on defined media, difficulties in systematic study due to product-specific effects, and health and safety concerns associated with the storage of large quantities of methanol. Difficulties also arise that are specific to the cultivation methods and control strategies used.

The present review discusses several bioprocess engineering aspects related to *P. pastoris* cultivation, including the different promoters available, both constitutive and inductive, on- and off-line process parameter monitoring methods, fed-batch and continuous cultivation control

strategies, proteolytic degradation of products and methods to minimize associated yield reductions, and the different models devised to describe cell growth and protein production. The economics of *P. pastoris* cultivation, through a case-study involving recombinant phytase production, are also discussed.

3.2 Introduction

Fuelled by exponential growth in worldwide demand for heterologous proteins and bioproducts, considerable progress has been achieved in bioprocess engineering regarding the development of high-yield economically viable production systems. At the end of 2010 the global market for recombinant proteins was valued at over 100 billion USD, with no signs of slowing down (Walsh 2010, Sanchez 2011).

The methylotrophic *Pichia pastoris* has become one of the most widely studied yeasts since its development in the early 1970s, and it is reportedly one of the most useful and versatile systems for heterologous protein expression. This system is of particular industrial interest due to its powerful and tightly-regulated methanol-inducible alcohol oxidase 1 promoter (pAOX1), its capacity for foreign protein secretion, its ability to perform post-translational modifications including glycosylation and disulfide bond formation, the ability to grow on defined media at high cell densities and its strong preference for respiratory, as opposed to fermentative, growth (Macauley-Patrick 2005, Cos 2006). The genome of *P. pastoris* is relatively simple to manipulate and commercialized expression kits are available (Life Technologies 2002, Life Technologies 2013, Ahmad 2014). Foreign proteins expressed in *P. pastoris* can also be directed to secretory pathways through signal peptides, which considerably simplifies downstream product recovery (Idiris 2010). To date, several hundred recombinant proteins have been expressed in *P. pastoris* systems (Plantz 2006) with expression levels as high as 80% total cell protein (TCP) (Cregg 1993).

The yield of heterologous proteins is affected by a variety of factors at both the genetic and cultivation levels. Important factors from the former category include codon usage and GC content of the foreign gene, proteolytic cleavage, and glycosylation sites on the expressed protein. It is well established that the genomes of different species employ biases in codon usage as mechanisms for optimizing and regulating protein expression (Gustafsson 2004). Optimizing the codon sequence and GC content of a human glucocerebrosidase gene led to a 10.6- and 7.5-

fold increase in protein expression levels in *P. pastoris*, highlighting the usefulness of such strategies (Sinclair 2002). Proteolytic degradation of expressed foreign protein also has a profound impact on observed yields. The development of novel gene optimization tools is facilitating the ongoing shift from traditional, often time-prohibitive, systematic expression optimization procedures to the *de novo* design and synthesis of efficient expression systems. This makes codon optimization and the prediction of protease cleavage sites considerably simpler and time-efficient (Welch 2009). Strain-specific factors, such as promoters and gene copy number also influence the expression level (Macauley-Patrick 2005).

Foreign protein expression is also affected by cultivation-level factors, notably temperature, pH and dissolved oxygen (DO) concentration, either directly through effects on molecular mechanisms or indirectly through effects on cell growth. Although *P. pastoris* has an optimal growth temperature of 30ºC (Cos 2006), induction phase temperatures as low as 15ºC have been shown to enhance foreign protein expression without significantly affecting cell growth (Wu 2008). Lower temperature cultivation can also improve yields by decreasing protease activity in certain applications (Sirén 2006). *P. pastoris* cultures are typically grown at pH values ranging from 3 to 7 (Cregg 2000). pH values outside of this range, although having little effect on cell growth, may affect both protein stability and extracellular protease activity (Jahic 2003) and may cause the precipitation of medium salts. In oxygen-limited cultivations on glucose or glycerol, fermentative by-products such as ethanol may have a negative impact on foreign protein expression (Inan 2001) and DO levels must be maintained sufficiently high.

Several excellent reviews on many aspects of *P. pastoris* cultivation have been published in recent years (Cereghino 2002, Macauley-Patrick 2005, Gurkan 2005, Daly 2005, Cos 2006, Li 2007, Graf 2009, Silva 2009, Gao 2013, Gasser 2013, Ahmad 2014). These reviews, however, do not focus specifically on aspects of bioprocess engineering. Given recent advances in the development of *P. pastoris* as a powerful and industrially viable protein expression system, an updated review focusing on the subject would be a useful reference for any future work in process design and optimization. The present review discusses recent work in promoter isolation, fed-batch and continuous cultivation control strategies, issues relating to proteolytic degradation and progress in modeling of *Pichia pastoris* growth, as well as provides a brief overview of economic factors of different cultivation strategies, underscoring their impact on process viability.

3.3 Promoters used in *P. pastoris* systems

3.3.1 pAOX1 Promoter

The methanol-inducible pAOX1, to which the success of *P. pastoris* as a foreign protein expression system is mainly attributed, was first isolated and used for the development of vectors and genetic manipulation protocols by Salk Institute Biotechnology/Industrial Associates Inc. (SIBIA) (Macauley-Patrick 2005). Alcohol oxidase (AOX), the first enzyme in the inducible methanol utilization pathway, represents up to 35% TCP in wild-type cells grown on methanol. In cultures grown on glucose, ethanol or glycerol, its expression is undetectable (Sreekrishna 1997). Although an effective system, the methanol required for induction presents a fire hazard and a considerable health risk, especially in larger-scale operations in which large quantities of methanol must be stored on-site.

Both AOX1 and AOX2 genes code functional enzymes, with the former accounting for up to 95% of the total expressed alcohol oxidase due to the relative strength of its promoter (Macauley-Patrick 2005). Systems using pAOX1 are therefore preferred for foreign protein expression, and are by far the most common systems used, although significant expression of recombinant protein has been reported in systems using truncated versions of pAOX2 (Mochizuki 2001, Kuwae 2005).

There are three *P. pastoris* methanol utilization phenotypes. Both AOX genes are functional in Mut^+ strains, which grow at the wild-type rate on methanol, but require large quantities of it during cultivation. The sensitivity of Mut^+ strains to high methanol concentrations makes such systems difficult to control and scale-up, and AOX expression may compete for available cellular machinery with recombinant protein production processes, lowering the yield of the desired products. In Mut^S strains the AOX1 gene is deleted, and strain growth is thus limited by AOX2 expression. In some cases higher productivities were obtained in Mut^S strains compared to the wild-type strains (Cos 2005). Due to the lower growth rate, difficulties inherent to the high-density cultivation of rapid growth strains, notably oxygen limitation, are alleviated. Lower sensitivity to methanol also makes the process easier to scale-up. Mut^- strains do not express either AOX genes and consequently cannot grow on methanol (Chiruvolu 1997).

Recent progress has been achieved in increasing the effectiveness of this promoter by compiling a library of putative transcription factor binding sites within pAOX1, resulting in

modified promoter sequences with activities ranging from 6% to >160% of wild-type activity. Novel artificial promoters can be assembled by combining cis-acting elements with the basal promoter, improving protein yield and quality (Hartner 2008).

pAOX1-regulated systems are by far the most commonly used systems, although constitutive promoters are quickly gaining ground.

3.3.2 pGAP Promoter

The glyceraldehyde 3-phosphate dehydrogenase (GAPDH) constitutive promoter (pGAP) was first isolated in 1997 (Waterham 1997). It has since been used to express many heterologous proteins in cells using glucose or glycerol as growth substrates without the process control or safety concerns associated with methanol (Cos 2006, Ahmad 2014). Unlike AOX1 systems, in which protein production is limited by the methanol available during the induction phase, in GAP systems biomass and protein synthesis occur simultaneously and are directly correlated to pGAP-regulated gene dosage (Vassileva 2001). While as conditions during the induction phase must be strictly controlled in AOX1 systems, GAP systems have minimal control requirements (Boer 2000). Stable and efficient continuous cultivation using the GAP promoter can be achieved with longer protein production periods, lasting up to 30 days (Goodrick 2001). pGAP-regulated genes are well expressed when cells are grown on methanol, glucose or glycerol, and more recently on fructose and ethanol (2015, see Chapter 7).

Many heterologous proteins have successfully been expressed in pGAP-based *P. pastoris* systems, with expression levels varying significantly based on properties of the expressed proteins. Several studies have reported that pGAP is more efficient than pAOX1 for protein production (Döring 1998, Delroisse 2005), whereas others report opposite results (Sears 1998, Vassileva 2001, Kim 2009). Additional research is required to determine which factors impact the efficiency of both promoters in terms of recombinant protein expression, in order to design *de novo* optimal production systems.

Expressing proteins under the control of both pAOX1 and pGAP has proven to be an effective strategy to increase yields. Developing such systems involves sequentially transforming host cells with vectors containing pAOX1- and pGAP-regulated genes, resulting in a strain expressing the target protein both constitutively and inductively through non-competing mechanisms (Wu 2003a, Wu 2003b, He 2008).

A more detailed review on the properties of the GAP promoter in *P. pastoris* has recently been published by Zhang and coworkers (Zhang 2009), and an entire book Seidler (2013) is dedicated to this subject.

3.3.3 Alternative Promoters

3.3.3.1 Inducible Promoters

The promoter for the glutathione-dependent enzyme formaldehyde dehydrogenase (pFLD1) is an attractive alternative to pAOX1 as it allows the regulation of protein expression through induction with either methanol (carbon source) or methylamine (nitrogen source) (Shen 1998). A methanol-free fed-batch cultivation system was developed for the methylamine-regulated production of a *Rhizopus oryzae* lipase under the control of pFLD1, with resulting productivities comparable to those of pAOX1 systems (Cos 2005, Resina 2005, Resina 2009). Gelatin and GFP, under the control of pAOX1 and pFLD1 respectively, were successfully simultaneously expressed in *P. pastoris*. As both are methanol inducible but only pFLD1 is methylamine-inducible, such systems will prove useful for the simultaneous expression of co-dependent proteins or the modulation of expression of certain interacting proteins to determine their effect on each other (Duan 2009). Dihydroxyacetone synthase (DHAS) is involved in the methanol assimilation pathway, and its promoter (pDHAS) is also inducible by methanol. Due to its strong promoter DHAS can account for up to 20% TCP in cells grown on methanol (Gellissen 2000).

In addition to the methanol-optional pFLD1, inductive promoters have been isolated that do not require methanol for induction. The promoter for PEX8, a peroxisome membrane protein, is inducible by both methanol and oleic acid (Liu 1995). The isocitrate lyase gene promoter (pICL1) is inducible with ethanol, which is safer to handle and store than methanol (Menendez 2003). Although these promoters provide alternatives to methanol-inducible systems, they are not widely used due to their low expression levels compared to more conventional systems.

3.3.3.2 Constitutive Promoters

In addition to pGAP, several constitutive promoters have successfully been used for recombinant protein production. YPT1, a GTPase involved in secretion, has a weak constitutive promoter that is rarely used in production-based applications due to low expression levels (Segev 1988, Sears 1998). The constitutive promoter for the highly expressed glycolytic enzyme phosphoglycerate

kinase (pPGK1) has also been used in *P. pastoris*. Like pGAP systems, pPGK1-regulated gene expression is effective when grown on glucose, methanol or glycerol, with the highest yields achieved on glucose (de Almeida 2005). pTEF1, the promoter for translation elongation factor 1-α (TEF1), has proven to be a constitutive promoter of comparable strength to pGAP, with tighter growth-associated characteristics, and exhibiting a 2-fold stronger promoter activity in carbon-limited fed-batch cultivation. pTEF1 may provide a good alternative to pGAP systems offering a wider choice of operating conditions and easier methods for expression (Ahn 2007).

3.4 Process Monitoring

Efficient and accurate monitoring of key process variables is essential for effective process control and maintenance of the conditions optimal for recombinant protein production. Factors having particular importance on the process include residual methanol concentrations, protein production rates and cell density. Recent progress has been made in on- and off-line monitoring, as well as (near) real-time quantification of these process variables. No method is superior in all respects, and selection of monitoring strategies for specific systems must be carried out on a case-by-case basis.

3.4.1 Methanol Monitoring

Sufficient levels of methanol are required during the induction phase of pAOX systems to ensure maximal protein production. Excessive methanol concentrations, typically ranging from 3.7 g/l to 20 g/l, are, however, cytotoxic and lead to growth inhibition, while as concentrations between 2 g/l and 3.5 g/l are optimal for protein production (Cuhna 2004, Schenk 2007). The methanol concentration must be maintained within a relatively narrow range, and efficient monitoring methods are thus required (Surribas 2003, Guarna 1997, Cuhna 2004, Zhang 2000a, Curvers 2002).

Chromatographic methods such as GC and HPLC are the most common off-line methanol monitoring methods, despite being expensive and time-consuming. The inherent time lapse between sample collection and analysis may reduce the effectiveness of control heuristics, and the required processing prior to analysis increases the risk of sample contamination. Due to its volatility, the evaporation of methanol at all stages of analysis may skew results. GC has also

been used for on-line monitoring by some researchers (Minning 2001, Curver 2002); however this method has a low sampling frequency (Schenk 2007).

On-line methods are generally based on liquid-gas equilibrium and monitor methanol in the broth by analyzing the fermenter exhaust gas (Schenk 2007). Such methods are typically less expensive than off-line techniques but have long response times and require non-linear calibration (Schenk 2007, Guarna 1997, Cuhna 2004, Katakura 1998, Crowley 2005). One such on-line methanol sensor developed by Guarna and co-workers was successfully used for methanol monitoring in shake flask cultures of a strain of *P. pastoris* producing a recombinant mouse endostatin (Guarna 1997). Surribas and coworkers developed an automated sequential injection analysis (SIA) method as a means of on-line methanol monitoring which was found to be more practical than flow injection analysis (FIA) systems (Surribas 2003). Flame ionization detectors (FIDs) have been shown to accurately measure the methanol content in samples, and can be effectively used for the on-line monitoring of methanol in *P. pastoris* cultures (Gurramkonda 2009).

Fourier Transform mid-Infrared spectroscopy (FTIR), as a near real-time method capable of detecting low protein concentrations, is apposite for *Pichia* cultivation. At protein levels below 1 g/l, methanol concentrations need to be strictly controlled to properly induce and maintain recombinant protein production (Crowley 2005). The calibration of FTIR systems is generally complex and time-consuming, although a simplified calibration method was recently developed for a *P. pastoris* fed-batch fermentation system (Schenk 2007).

Combinations of on- and off-line methanol monitoring strategies are often implemented as a means to increase monitoring, and thus process control, efficiency.

3.4.2 Biomass Monitoring

Accurate monitoring of the biomass concentration, a direct indicator of cell growth, is critical for the optimization of heterologous protein expression systems. The most widespread methods include dielectric spectroscopy, optical probing, infrared spectroscopy and fluorescence microscopy. An excellent review by Kiviharju provides a detailed comparison of these methods (Kiviharju 2008).

Multi-wavelength fluorescence (MWF) is an optical method which can be used in conjunction with multivariate modeling for on-line monitoring of biomass in yeasts. The biomass

concentration is determined based on signals from various natural fluorophores including NAD(P)H, flavins and aromatic amino acids which exhibit characteristic peaks in a well-resolved MWF spectrum (Odman 2009). Surribas *et al*. used MWF combined with chemometric (PLS-1) models to monitor biomass in a fed-batch cultivation of *P. pastoris* producing a *Rhizopus oryzae* lipase. It was found that MWF can satisfactorily (with errors less than 7%) be used for quantitative prediction of biomass for both induced and non-induced batch cultures of *P. pastoris* (Surribas 2006). However, biomass concentration cannot accurately be measured on-line during the production phase in which proteolytic degradation becomes important (Jenzsch 2006).

Where direct biomass quantification methods are not available, indirect estimation using model-based predictions may be applied. Classical mechanistic models based on mass balances are not always the best choice for this purpose, as they require detailed knowledge about process dynamics and biochemical reaction kinetics, which are not always readily available. In these cases, data-driven models may be used, in which easily accessible process data is used for the estimation of non-measurable variables. Jenzsch *et al*. have compared different biomass estimation models based on the carbon dioxide evolution rate (CER) and oxygen uptake rate (OUR) during production of a recombinant protein with *Escherichia coli*, using artificial neural networks (ANN) or principal component analysis (PCA). The authors have reported good model fits and low prediction errors (< 1 g/kg) (Jenzsch 2006, Odman 2009). The best estimates are obtained with artificial neural networks which are commonly used to express non-linear relationships between process variables. These models, however, require specialized software tools and training. Jenzsch suggested two approaches for biomass estimation using ANNs: the direct mapping of the cumulative values of the online measurable OUR and CER quantities, from which biomass concentrations are determined by means of a feed-forward ANN, and a technique based on auto-associative artificial neural networks (AANN). In the first approach, the on-line measured variables are directly mapped onto the biomass concentration using a simple feed forward ANN. The second approach is based on non-linear PCA using auto AANN (Jenzsch 2006).

3.4.3 Protein Production Monitoring

Process optimization can only be achieved with effective measurement, either qualitative or quantitative, of heterologous protein production. The most common protein monitoring methods

include bioactivity assays, ELISA, SDS-PAGE and Western blots coupled with band density scanning (Zhang 2006). Bioactivity assays are applicable only when the enzymatic or interaction mechanisms of the expressed protein are well characterized. Western blots require protein-specific antibodies which are difficult to produce and purify, and thus expensive if not readily available. The enumerated methods are time-consuming, labour-intensive, and destroy samples, making them poor techniques for optimizing protein production in most cases.

Advances in analytical technology have yielded improved protein monitoring methods, including perfusion chromatography (Gadowski 1995, Zhang 1998), electrochemiluminescence (Grimshaw 1997), special biosensors (Gill 1998, Tsoka 1998, Jung 2004) and immunonephlometric assays (Ledue 1998, Cabarello 1999). Baker and coworkers provide a detailed comparison of these methods (Baker 2002). These techniques, however, are limited to monitoring proteins secreted in the extracellular environment and do not measure intracellular protein production (Zhang 2006).

An alternative method, capable of monitoring both intracellular and extracellular recombinant protein production consists of fusing a GFP signal marker to the heterologous protein of interest and detecting the product by fluorescence microscopy (Cha 2005). This method has been successfully used on-line in conjunction with in-situ methods to monitor recombinant protein production in *P. pastoris* (Hisiger 2005, Reischer 2004, Jones 2004). Although only limited work is reported regarding the establishment of such quantitative correlations, Chaa and coworkers report a linear relationship between recombinant protein levels and GFP fluorescence. They compared several expression systems based on this correlation and successfully calculated production yield and productivity from GFP fluorescence measurements (Chaa 2005). Interference with fluorescence signals may limit the usefulness of GFP fusion systems for on-line monitoring of extracellular recombinant protein production. Surribas and coworkers report that *P. pastoris* secretes significant amounts of riboflavin to the extracellular medium. When using multi-wavelength spectrofluorometric techniques, the signals obtained from riboflavin interfere with GFP signals as their fluorescence peaks are similar (520nm and 510nm respectively). Ultra filtration can be employed to effectively separate the compounds (Surribas 2007a) or alternative fluorescent markers can be used. Presently there are no reported cases of using GFP-protein fusions for monitoring protein production in large-scale high cell density cultivations of *P. pastoris*.

3.5 Fed-Batch Cultivation of *P. pastoris* and Process Control Strategies

3.5.1 Fed-Batch Cultivation of P. pastoris

Fed-batch is generally preferred over continuous cultivation as such systems achieve high cell densities while being easier to control. The pAOX1-regulated fed-batch cultivation of *P. pastoris* is generally divided into three phases: the glycerol batch and fed-batch phases, and the methanol induction phase.

The first two phases are collectively referred to as the biomass production phase, as their main purpose is to produce biomass prior to methanol induction. Glycerol is used as the growth substrate at this stage as cells grown on glycerol have a higher specific growth rates than those grown on methanol (0.18L/h (Cos 2005) and 0.14 L/h (Brierley 1990) respectively). The glycerol batch phase lasts approximately 24 hours and is typically initiated with a glycerol concentration around 40 g/l, as higher glycerol concentrations are inhibitory to growth (Cos 2006).

Once the initial glycerol is consumed, as indicated by a spike in measured DO, the glycerol fed-batch phase is initiated. The Invitrogen *Pichia* fermentation protocol recommends using a constant glycerol feeding rate (Invitrogen 2002), although exponential feeding profiles can also be used to avoid glycerol accumulation. During the first two phases, pAOX1 is completely inactive and no recombinant proteins are produced. The length of the glycerol fed-batch phase depends on the desired biomass concentration prior to methanol induction.

During the induction phase, the feeding rate of methanol, which is both the carbon source and pAOX1 inducer, directly impacts the residual methanol concentration, specific growth rate of the culture and heterologous protein expression levels. The recombinant protein yield not only varies with the specific growth rate of the culture, but also with the induction time. It was found, for example, that the production of the rOvIFN-τ protein started to decrease after 31-45 hours, a production shift referred to as decoupling, which was independent of the initial cell density before induction and is associated with a drop in the cells' energy state following protein accumulation (Plantz 2006).

An optional transition phase can be added between the biomass production and methanol induction phases to aid cellular adaptation. During the transition phase, glycerol and methanol are co-fed for a given length of time during which the concentration of glycerol in the feed gradually decreases while the concentration of methanol increases (Zhang 2000a, Minning

2001). Compared to a direct switch to feed solutions with methanol as the sole carbon source, transition phases accelerate the adaptation of cellular metabolism and induction of pAOX1 leading to higher productivities (Jungo 2007a). A recent study by Qureshi delayed the transition phase, inserting it during the induction phase, as opposed to immediately following biomass accumulation. This led to a significant increase in recombinant polygalacturonate lyase production and cell viability by temporarily reducing the methanol-induced stress on the cells, allowing their energy stores to be replenished, and eliminating the weaker ones (Qureshi 2010).

Replacing the culture medium immediately prior to the methanol induction phase significantly improves protein expression due to the removal of residual glycerol and metabolic waste both of which both repress induction. Chen and coworkers report an increase of 260% in recombinant phytase activity in high-density *P. pastoris* cultivations when the culture medium is replaced (Chen 2004).

The fed-batch cultivation of *P. pastoris* under the control of pGAP is divided into two phases: a batch phase for initiation of biomass growth and a fed-batch phase. Either glucose or glycerol is conventionally used as the sole carbon source. No definite conclusions have been drawn as to which substrate is superior for recombinant protein production, with some studies reporting glucose as the better substrate (Döring 1998, Pal 2005) while others report opposite results (Zhang 2007, Tang 2009). In the fed-batch phase, the feed is generally a high concentration substrate solution supplemented with trace salts.

3.5.2 Fed-batch control strategies

The methanol feed rate in pAOX1-regulated systems is one of the most important factors to control, as the residual methanol concentration directly influences the rates of production and proteolytic degradation of heterologous proteins, cell growth, cell lysis and oxygen transfer. When the objective of the process is to produce small quantities of heterologous protein, for instance in structural studies or for the production of cytotoxic compounds, the standard methanol-feeding strategy suggested in Invitrogen's fermentation guidelines may prove adequate. However, when the objective is to maximize product yields, optimization of the methanol-feeding strategy is necessary (Cos 2006).

Many strategies have been proposed and implemented to control methanol feeding with the objective of maximizing protein production and enhancing process reproducibility. The most

common methanol feeding strategies include constant DO feeding (DO-stat), constant specific growth rate feeding (μ-stat), constant methanol concentration feeding, oxygen limited fed-batch (OLFB), and temperature limited fed-batch (TLFB).

3.5.3 DO-stat control

Dissolved oxygen (DO) refers to the relative percentage of oxygen in the medium. Oxygen is required in the first step of the methanol catabolism pathway, and DO must therefore be kept above a minimal level, generally about 20% (Singh 2008, Invitrogen 2002). Excessively high DO levels, however, are cytotoxic and significantly reduce cell viability (Chung 2000), and given the cost of oxygen, maintaining high levels may be economically unfeasible. The measured DO profiles provide valuable information about a culture's growth phase and health. Proper monitoring and control of DO concentrations is vital to the viability of a given process.

Pichia cells utilize methanol through the oxidative pathway only when oxygen is non-limiting. The oxygen concentration in the culture and the optimal methanol feeding rate (MFR) are thus interrelated. In order to prevent methanol accumulation, an efficient methanol-induction strategy with proper DO control is essential for optimum protein production with *Pichia* expression systems (Lim 2003, Woo 1997).

Many simple systems that independently control methanol induction and DO concentrations have been developed (Siegel 1989, Brierley 1990, Barr 1992, d'Anjou 2001, Goodrick 2001). Such systems involve manual adjustments of methanol induction using DO readings as indicators of the culture condition. The DO levels are maintained at certain set points, typically between 15-20% of saturation, through agitation feedback control and by varying the oxygen content in the inlet air stream. In such systems, establishing an optimal methanol feeding strategy is tedious and affected by fluctuations inherent to DO profiles (Lim 2003). Katakura developed a methanol control strategy based on monitoring with a methanol sensor (Katakura 1998), and Zhang proposed a control strategy based on a growth model of *P. pastoris* (Zhang 2002).

DO-stat processes control the substrate feed rate to maintain DO concentrations at a constant optimal level in the culture medium (Lee 2003, Chung 2000, Xiao-Qing Hu 2008). This control strategy allows a culture to reach high cell densities, maintains the culture in a prolonged highly productive state during induction, and prevents substrate accumulation (Chung 2000, Lee 2003). The DO-stat strategy proposed by Lim controlled the ratio of partial pressure of pure O_2 in the

inlet stream (O_2^{PR}) and the MFR to provide an improved maintenance of DO during induction. The DO levels were maintained between 40-45% of saturation through a feedback control loop with measurements taken every 30 seconds. The MFR was decreased by 0.5% when the DO level went below the lower set limit, and was increased by 1% when above DO the upper limit. This linking of DO and MFR prevented cell fatality resulting from both methanol accumulation and oxygen deficiency (Lim 2003).

Oliveira *et al.* developed an automatic feeding strategy based on a closed-loop controller which relied on the manipulation of DO. This strategy is particularly useful in high cell density cultures where oxygen transfer limitations are expected to occur. However, this strategy relies on the assumption that the carbon source does not accumulate in the system. Controlling the DO by manipulating the feed rate is not possible when there is accumulation of the carbon source, as DO control becomes insensitive in such situations (Oliveira 2004).

High density cultures of *P. pastoris* exhibit oscillatory behavior when methanol is fed under DO-stat operation. This, when left unattended, leads to irreversible loss of culture productivity and thus lower yields, and proper controller tuning is therefore required (Chung 2000). Chung *et al.* developed and analyzed a simple mathematical model of a closed-loop DO-stat process. They identified stable and unstable regions in relation to controller tuning and metabolic parameters. Based on this, they derived theoretical stability criteria for the design of metabolic feed controllers during high cell density fermentations. A general linear system of design equations was also developed for application to the design of feedback controllers for fermentation processes (Chung 2000).

DO-stat systems, although simple to operate, are not entirely reliable (Yamawaki 2007). During rapid cell growth, in which oxygen consumption is high, the DO levels are low and as a result methanol concentrations may become limiting if DO-stat control is solely relied upon. This may lead to cell starvation and decrease protein yields. On the other hand, if excessively high (inhibitory) methanol concentrations are reached, the DO concentration rises sharply as a result of cell decline. Under DO-stat control, the rise in DO would trigger an even higher methanol concentration which would further inhibit cell growth and thus decrease protein production.

Although DO-stat control has successfully been implemented in several studies as a means to increase process robustness and performance, the methanol concentration and specific growth

rate were not held constant, making the influence of these parameters on production difficult to determine accurately (Chung, 2000, Inan 2001, Lee 2003, Oliveira 2005, Cos 2006).

3.5.4 µ-stat Control

Another common methanol-feeding strategy is µ-stat control. This control method adjusts the MFR based on mass balance equations to theoretically maintain a constant specific growth rate (µ) (Cos 2006). This simple approach can be implemented based solely on simple cell growth models, and does not require any on-line monitoring of system parameters. Specific growth rate control is regarded as an effective strategy for process optimization as most biochemical processes, including protein production, are either directly or indirectly associated with cell growth (Ren 2005). Maintaining a constant µ enhances process reproducibility and facilitates the systematic study of growth rate-related effects on heterologous protein production.

A common method to maintain a constant µ consists of devising a MFR profile based on kinetic models. Zhang *et al.* established a model describing the relationships between the specific growth rate, methanol concentration and specific methanol consumption rate. The maximum specific growth rate (μ_{max}) calculated from this model was $0.08h^{-1}$ at a methanol concentration of 3.65 g/l while the actual maximum rate was $0.0709h^{-1}$. The maximum recombinant protein yield was obtained with a constant µ of $0.0267h^{-1}$, or approximately one third of μ_{max} (Zhang 2000b). Ren *et al.* developed a macrokinetic model for *P. pastoris* expressing a recombinant human serum albumin (rHSA) based on stoichiometric balances, which described protein production and cell growth (Ren 2003). Based on this model the relationship between the set-point of µ and the production of rHSA was studied. It was found that by using a combination of linear and exponential feeding profiles derived from the model, µ was accurately maintained at the set point. The study also suggested that a µ set-point control strategy is more efficient than maintaining a constant feed rate in maximizing productivity (Ren 2005).

Sinha and coworkers modeled cell growth on methanol with a substrate-feed equation, and used the model to effectively control the process. It was found that the methanol feed strategy is the key to controlling the recombinant protein induction and protease production in *P. pastoris* fermentation. As in other published studies, an optimal zone of operation in terms of methanol feed profile was found which could be controlled by automated substrate growth models. The optimal µ in these experiments was found to be $0.025\ h^{-1}$(Sinha 2003).

Trinh and coworkers compared three methanol feeding strategies for heterologous production of mouse endostatin by *P. pastoris*. The first method was based on methanol consumption, the second on DO concentrations, and the third strategy consisted of μ-stat control with limited methanol feeding, based on a predetermined exponential feeding rate, maintaining the growth rate at $0.02h^{-1}$. It was found that the production of endostatin per unit biomass per unit methanol was 2 times higher in the μ-stat culture compared to others (Trinh 2003).

μ-stat is an open loop control strategy and it presents difficulties in terms of robustness and process stability. Changes in initial conditions or disturbances in process variables may lead to over-accumulation of methanol. To avoid methanol accumulation, μ is usually maintained at a set-point considerably lower than μ_{max} which lowers the productivity of the process (Zhang 2000b, Ren 2003).

Although the listed results may serve as a reference for the design of fermentation systems, each study was based on the production of different heterologous proteins which makes comparison difficult as optimal specific growth rates are highly protein specific.

3.5.5 Constant Methanol Concentration

In DO-stat and μ-stat processes, the methanol concentration is neither measured online nor directly controlled. As a result, deviations from the optimal methanol concentration, methanol depletion or methanol accumulation in the medium may occur. Accurate methanol monitoring, as discussed in section 3.3.1, and subsequent effective control are requisites for robust and reproducible bioprocesses.

A number of methanol concentration control strategies have been proposed. The "on-off" control mode is the simplest feedback control strategy and has been used in numerous studies (Guarna 1997, Katakura 1998, Wagner 1997, and Zhang 2000a). On-off control, however, is only suitable for linear systems while heterologous protein production in *Pichia* is generally a more complex and highly non-linear process. The on-off control strategy may consequently result in fluctuations around the set-point (Zhang 2002, Cos 2006).

Proportional-integral (PI) or proportional-integral-derivative (PID) control heuristics are effective for the maintenance of methanol concentrations. In a PID control system, the system performance depends on the controller gain K_C, the integral time constant, τ_I, and derivative time constant, τ_D (Zhang 2002). Optimal controller settings depend on the specified desired biomass

concentration and the broth volume. Due to the nonlinearity and complexity of fermentation process dynamics, which are subject to inherent and externally imposed variability, the optimal settings of the PID controller are ascertained by trial and error tuning or by other empirical methods (Zhang 2002, Cos 2006). Several methods have been developed to design PID controller settings among which frequency response methods are considered particularly useful. Several studies have used Bode stability criteria with frequency response analysis as a means of controller tuning (Zhang 2002, Chung 2000) while others have employed empirical methods. Due to the complex dynamics in such systems, optimal control parameters may vary considerably which makes this strategy difficult to implement.

Cos *et al.* employed a predictive control algorithm combined with a PI feedback controller in order to optimize the production of a *Rhizopus oryzae* lipase in *P. pastoris*. Their simple model-based system controlled the methanol concentration on-line in fed-batch fermentations. This predictive model requires the first time-derivative of methanol concentration as an input. The proportional gain and integral time constant of the PI controller were fixed for the duration of the cultivation (Cos 2006b).

Yamawaki *et al.* examined the effects of DO-stat and constant methanol control strategies in fed-batch and continuous cultures of *P. pastoris* on cell growth and recombinant scFv expression. In both fed-batch and continuous cultures, it was found that by maintaining the methanol concentration at 3.9 g/l, a higher specific productivity of scFv was obtained in comparison with DO-stat control (Yamawaki 2007). Minning and coworkers found that controlling the methanol feed rate based on the measured methanol concentration was a better control strategy than using DO readings-based heuristics for the cultivation of a *P. pastoris* producing a *Rhizopus oryzae* lipase (Minning 2001, Ren 2005).

3.5.6 Oxygen Limited Fed Batch (OLFB)

In oxygen limited fed-batch (OLFB) cultivations, oxygen is the limiting nutrient instead of methanol. Although oxygen limitation should generally be avoided during the induction phase, as it adversely affects foreign protein yields (Cregg 2000, Khatri 2005b, Bushell 2003, Lee 2003), successful protein production has been achieved under oxygen-depleted conditions (Trentmann 2004, Khatri 2005b).. As low oxygen levels interfere with the production of recombinant proteins (Bushell et al. 2003, Lee et al. 2003, Cregg 2000), the oxygen demand is

reduced by operating at low cell densities or low cultivation temperature, or by using strains with slow methanol utilization (Chiruvolu et al. 1997, Cos et al. 2005, Curvers et al. 2001; Jahic et al. 2003). However, some reports suggest that certain recombinant proteins are produced with higher yields in oxygen-limited processes than in methanol-limited ones (Trentmann 2004; Charoenrat et al. 2005; Hellwig et al. 2001). In OLFB, the residual methanol concentration is kept constant, but the DO concentration is allowed to vary. In these processes, the DO always drops to 0% due to oxygen limitation, which increases the driving force for oxygen transfer.

OLFB cultivations, when compared to methanol-limited processes, have been reported to reduce post-translational product modifications (Trentmann et al., 2004). They may also increase the purity of product secreted to the medium (Charoenrat 2005). Moreover, pure oxygen requirements increase the production costs and may cause difficulties in scale-up when heat exchange and oxygen transfer capacities are low (Curvers 2001). OLFB systems are thus economically attractive as they reduce the oxygen requirements to a minimum (Khatri, 2005b). Trentmann reported that using a limited oxygen supply contributed to the robustness of the process and resulted in the highest yield of purified scFv observed during the study (Trentmann 2004). Oxygen limitation also reduces cell lysis (Charoenrat 2006)

Under oxygen limitation, the methanol consumption is determined by the rate of oxygen transfer in the reactor (Khatri 2005a). In comparative studies, due to the higher oxygen driving forces, the oxygen and methanol consumption rates were approximately 40% higher in OLFB processes compared to those under DO-stat control (Charoenrat, 2005). This results in increased cell density and higher protein production, which increases with methanol consumption (Khatri 2005a, Khatri 2005b).

3.5.7 Temperature Limited Fed-Batch (TLFB)

In high density *Pichia* methanol-limited fed-batch cultures, proteolytic degradation is a significant concern (Cregg 2000). Numerous adjustments to the standard process have been proposed to minimize proteolysis, including the manipulation of culture pH and temperature and adding casamino acids and peptone to the cultivation broth (Clare 1991, Hong 2002, Li 2001, Zhou 2002, Jahic 2003b, Siren 2006).

In temperature-limited fed-batch (TLFB) systems, methanol limitation is replaced by temperature limitation mainly to avoid oxygen deficiency at high cell densities (Jahic 2003b). In

contrast to DO-stat and μ-stat processes, TLFB systems allow for the control of both oxygen and methanol concentrations. The methanol concentration in the broth is maintained constant using previously described methods, while the culture temperature is lowered in order to maintain DO at a particular set point. Cell growth is therefore limited by temperature, and not directly by methanol concentration or DO levels in the medium. TLFB is especially useful for Mut^+ strains, as in these processes non-limiting methanol concentrations have been reported to lead to cell death and oxygen limitation (Surribas 2007a).

Jahic *et al.* used the TLFB method to cultivate a *P. pastoris* producing a fusion protein. It was found that using TLFB resulted in a higher cell density, higher product concentration, lower cell death, as well as drastically lower proteolytic degradation when compared to methanol limited fed-batch. The higher accumulation of product in TLFB processes is explained by decreased protease activity at lower temperatures, which lead to a higher product recovery. In TLFB processes pAOX activity is increased 3.5-fold compared to methanol limited processes, which also contributes to the high yields observed (Jahic 2003b). In a similar set of experiments conducted by Surribas, it was found that cell death is significantly reduced in TLFB processes, leading to more sustainable cultures and a 1.3-fold increase in final product purity (Surribas 2007b). Other researchers have also reported successful use of TLFB cultivations with low temperatures between 10^0C-30^0 C, with significantly lower product degradation (Sirén 2006, Ruiz 2009).

3.5.8 Fed-Batch Shake Flask Cultures

Due to logistical constraints, shake flask experiments are typically batch cultivations, and fed-batch and continuous processes require more complex equipment. However, two novel systems have been developed that allow quasi-fed-batch cultivations to be performed in shake flasks to increase product yields.

In shake flask experiments, standard pulse feeding of methanol may lead to long periods of methanol exhaustion and low pAOX1 activity. To alleviate this problem, Ruottinen and coworkers proposed a new strategy for improved production of heterologous protein in *Pichia* using a quasi-continuous feeding profile. This strategy uses a wireless feeding unit, together with an on-line monitoring system which allows the following of a simple and inexpensive computer-

controlled feeding profile. This led to a higher recombinant protein yield as well as better proteolytic stability of the product in shake flask experiments (Ruottinen 2008).

Panula-Perälä *et al.* developed a novel glucose-limited fed-batch cultivation system with enzyme-based substrate delivery (EnBase^{TM}) that does not require an external glucose feed. Glucose is gradually released into the medium by biomass-proportional enzymatic degradation of starch which is supplied to cells via continuous diffusion from a storage gel. EnBase-controlled cultivations reach cell densities 5 to 20 times greater than standard cultivations as it effectively eliminates substrate over-accumulation in microwells and shake flasks (Panula-Perala 2008). Krause *et al.* applied the EnBase method to the shake flask cultivation of *E.coli* and obtained high cell densities without impairing the productivity of standard cultures. The yield of soluble and correctly folded proteins was significantly improved (Krause 2010). Although there are currently no reports on the use of the EnBase system in *P. pastoris* cultures, it may prove an effective substrate delivery method for such applications.

3.6 Continuous Cultivation of *P. pastoris*

3.6.1 Continuous Cultivation of pAOX1-regulated Systems

Continuous cultivations are generally divided into three phases: preliminary glycerol batch and fed-batch phases, identical to those of fed-batch cultivations, and the continuous feeding phase. Due to longer production times and reduced culture down times, continuous cultivation strategies often result in higher volumetric productivities than fed-batch cultivations.

The dilution rate (D), equal to the specific cell growth rate at steady state, is the most important parameter in continuous cultivations. Many studies have examined the relationship between protein production and the dilution rate. For Mut^{S} (d'Anjou 2000) and Mut^{+} (Jungo 2006, Yamawaki 2007) strains, the steady-state recombinant protein concentration decreases and the volumetric productivity increases with increasing dilution rates. Two exceptions to these observations have been reported. The maximum volumetric productivity was obtained at a moderate D = 0.0333 h^{-1} during the continuous cultivation of recombinant *P. pastoris* for the production of interferon-τ (INF-τ) (Zhang 2004). The volumetric productivity of α-amylase decreased with increasing dilution rate when the continuous feeding was controlled by DO-stat (Nakano 2006). Productivities in these studies could be affected by a number of factors including the concentration of media components, protease accumulation and product-specific effects.

The production of heterologous proteins is also strongly affected by cell density, and the conditions maximizing the volumetric productivity do not necessarily lead to optimal specific production rates. As determined by surface response methodology, during the production of interferon-τ the maximum volumetric productivity (mg INF-τ $L^{-1}h^{-1}$) was achieved at a cell density of 328.9 g/l wet cell weight (WCW) and a dilution rate of 0.0333 h^{-1} while as the maximum specific production rate (mg INF-τ g $cells^{-1}$ h^{-1}) was obtained at the lower cell density of 287.7 g/l WCW and D = 0.0361 (Zhang 2004).

When methanol is used as the sole carbon source, cell growth is slow, especially for Mut^S strains. One way to improve growth and productivity is to use mixed substrate co-feeding strategies, as discussed for fed-batch cultivations, to increase the carbon source concentration and energy supply to recombinant *P. pastoris* cells (Celik 2009b, Katakura 1998, Zhang 2000, Zhang 2003) and reduce induction times. Glycerol is the most common carbon source added to methanol, although due to its repression of pAOX1 may lead to lower protein specific productivities (Xie 2005).

In the continuous cultivation of a Mut^S *P. pastoris* strain expressing a recombinant porcine follicle-stimulating hormone (rFSH), the productivity and extracellular product concentration were respectively 7 and 3.7 times higher than the same values measured in a fed-batch process when glycerol and methanol were co-fed (Boze 2001). The continuous cultivation of a Mut^S strain producing a sea raven antifreeze protein on a mixed methanol-glycerol growth substrate led to a higher productivity than the fed-batch cultivation of the same strain (d'Anjou 2001). This strategy is not always effective. Co-feeding sorbitol and methanol for example does not significantly increase protein yield or productivity compared to methanol feeding, in part due to the low affinity of the strain to sorbitol (Boze 2001). A more detailed review of mixed feeding strategies is provided by Cos (2006).

3.6.2 Continuous pGAP cultivation

pGAP-regulated *P. pastoris* continuous cultivations lead to longer production periods, and production rates of recombinant proteins approximately five- to six-fold higher than fed-batch processes under the control of the same promoter. Proteolytic degradation of the products is also minimized, likely due to the continuous removal of sensitive proteins from the cultivation broth (Goodrick 2001). Economically viable large-scale continuous pGAP-regulated systems for the

development of proteins of industrial or therapeutic importance have been developed (Schilling 2001). Continuous cultivation is also useful for the study of cell growth kinetics, protein production, and metabolism of *P. pastoris* (Solà 2007, Jungo 2007b, Jungo 2007c).

Studies on pGAP-regulated continuous *P. pastoris* systems expressing human granulocyte-macrophage colony stimulating factor (hGM-CSF) (Khasa 2007) and a Fab fragment of anti-HIV antibody 2F5 (Maurer 2006) found that the specific rate of protein production (q_p) increased with the dilution rate. Cultures should therefore be maintained close to the maximum specific growth rate to achieve maximum system productivity.

Foreign protein production under pGAP is affected by culture conditions, notably pH, temperature and DO. The optimal pH and temperature for protein production are 6.5 and 26°C respectively for the production of recombinant *Candida rugosa* lipase (Zhao 2008) and 5 and 30°C respectively for the production of hGM-CSF (Pal 2006) in continuous cultivations. Optimal conditions for yield maximization may be different from those for growth maximization, due to their effect on several factors including protein-specific characteristics, kinetics and sensitivity of products to proteolytic degradation. Culture conditions can be optimized through response surface methodologies (Chang 2006).

In a study by Baumann and coworkers, a 2.5-fold increase in specific productivity was achieved by applying hypoxic conditions to the continuous cultivation of a strain of *P. pastoris* expressing an antibody Fab fragment, under pGAP. Under hypoxic conditions, biomass production decreased and ethanol was produced, indicating a shift from oxidative to oxido-fermentative metabolism. Based on these results, a feeding strategy was designed that used a fed-batch controller to control the feeding of glucose by maintaining the ethanol concentration at approximately 1.0% (v/v) leading to a 230% increase in the volumetric production rate (Baumann 2008).

3.7 Proteolytic Degradation and Stability of Recombinant Proteins

3.7.1 Protease Activity in P. pastoris Cultures

One of the major drawbacks of *P. pastoris* expression systems is the post-secretory proteolytic degradation of recombinant products (Idiris 2010). Proteolysis is largely protein-dependent; in some instances, expressed products remain intact, while in others no functional proteins can be recovered due to proteolytic activity. Differences in sensitivity to proteases depend on amino-

acid sequence and three-dimensional conformation that may expose or conceal recognizable cleavage sequences. Aspartic, cysteine and serine-type proteases, which may be extracellular, intracellular or cell-bound, are found in *P. pastoris* cultures.

Intracellular and extracellular protease levels and cell mortality are higher in cultivations on methanol compared to those on glycerol. Growth lag-phases, following stress caused by methanol itself or by the transition from a given carbon source to methanol during the induction phase of pAOX-regulated cultures, induces autophagic pathways in *P. pastoris* that may lead to over-expression of proteases and their lysis-mediated release to the fermentation medium, resulting in the increased degradation of secreted proteins (Yamashita 2009). Four vacuolar proteases, PrA, PrB, CpY and aminopeptidase, are usually detected in the supernatant of *P. pastoris* cultures (Sinha 2005). The high cell mortality rate during the induction phase is correlated with the accumulation of intracellular reactive oxygen species (ROS), notably formaldehyde and H_2O_2, which are by-products of the methanol metabolism (Xiao 2006). The oxidative stress caused by high intracellular ROS levels damages or lyses cells, leading to the further release of vacuolar proteases.

In addition to lowering functional recombinant protein yields, proteolysis may lead to contamination of the products by degradation intermediates, leading to added downstream purification costs. One of the most effective methods to reduce degradation is the use of protease-deficient strains, such as SMD1165 (his4, pep4, and prb1 deletions), SMD1163 (his4 and prb1 deletions) and SMD1168 (his4 and pep4 deletions). Use of these strains is not always applicable as proteases may be required for correct processing and activation of certain proteins. Proteolytic degradation can also be reduced by modifying the amino-acid sequence of the expressed protein, through modifications to its recombinant DNA sequence, to remove recognizable protease cleavage sites if this is possible without affecting protein function (Zhang 2008). Optimizing the culture pH, temperature and medium composition has also proven to be effective.

Supernatant protease activity can be measured with an activity assay using casein conjugated to a fluorescent marker as a substrate (Jones 1997, Kobayashi 2000, Wang 2005). Protease activity profiles over the course of fed-batch *P. Pastoris* cultivations have been established, although results are highly dependent on the amino-acid sequence and conformation of the

expressed proteins and on cultivation conditions (Kobayashi 2000, Noronha 2002, Sinha 2005, Wang 2005, Wu 2008, Wu 2010).

In the profile reported by Wang for a system expressing a recombinant merozoite surface protein 3 (MSP3), no protease activity was observed during initial growth on glycerol, but following methanol induction protease activity spiked and gradually returned to baseline values (Wang 2005), likely due to stress-related protease release. Wu and coworkers have reported that the total protease activity in a recombinant *P. pastoris* expressing a human consensus interferon-α increased with time throughout the entire fermentation process (Wu 2008). In the activity profile reported by Kobayashi for the expression of a recombinant human serum albumin (rHSA), no protease activity was detected in the broth until approximately 100 hours, after which a dramatic increase in activity and substantial degradation of target proteins was observed, a phenomenon likely caused by nitrogen starvation (Kobayashi 2000a). Given this large variation in protein-dependent protease activity profiles, cultivation conditions must be optimized to minimize degradation on a process-by-process basis.

3.7.2 Effects of Temperature on Proteolytic Activity

The culture temperature affects the stability and functionality of recombinant proteins, as well as the efficiency of the induction phase. A decrease in the induction phase temperature of a *P. pastoris* culture from 30ºC to 23ºC lead to a 3-fold improvement in the production of a recombinant herring antifreeze protein. This result can be explained by increased product stability and cell viability at lower temperatures, the latter of which reduces lysis-associated protease release (Li 2001). In a study by Jahic and coworkers, although the lower temperature used in a temperature limited fed-batch system led to a significant decrease in the cell growth rate, a 100% increase in the yield of a recombinant fusion protein was observed due to minimized protease activity in the supernatant and increased AOX activity (Jahic 2003a). The reduction in proteolysis at lower temperatures is due to a decrease in protease activity and not a decrease in protease production (Sirén 2006). As the effect of temperature on stability again depends on the recombinant proteins expressed, lower temperature strategies are not always effective, but should be taken into consideration while devising fermentation strategies (Wang 2005).

3.7.3 Effects of pH on Proteolytic Activity

The pH of the culture broth can affect both the protease activity and the stability of expressed proteins and it must therefore be monitored and maintained within an optimal range. In Kobayashi's study of rHSA production, protease activity, while undetectable at a pH above 5.6, halved the recombinant protein yields at pH 4.3 (Kobayashi 2000a). Similar effects of pH on protein yield were reported for Sinha's study on the production of interferon-τ (Sinha 2005), that of Wang on the production of MSP3 (Wang 2005) and that of Tojo on the study of human fibrinogen (Tojo 2008). Based on reported results, the optimal operating pH for most recombinant proteins ranges from 5.5-8 to minimize protease activity while maintaining protein stability. Higher pH values decrease cell viability and may lower recombinant product stability or activity. Degradation due to pH-mediated destabilization is, however, again highly dependent on the stability of the recombinant protein itself, which is usually increased by glycosylation and disulfide bond formation. Serine and aspartic proteases, both secreted by *P. pastoris,* are activated at low pH values, which may explain the pH-dependence of proteolytic activity (Kobayashi 2000a, Ohya 2002, Jahic 2003b, Idiris 2010).

3.7.4 Protease Inhibitors

Shi and coworkers report reductions in total protease activity of 53% and 30% in cultures to which serine- and aspartic-type protease inhibitors were added respectively (Shi 2003). In a study by Sinha and coworkers, the total protease activity was reduced by 78% following the addition of phenylmethylsulfonylfluoride (PMSF) (1mM), a serine protease inhibitor, by 45% when EDTA, a metalloprotease inhibitor, was added (1mM) and by 94.2% when a combination of EDTA (1mM) and PMSF (1mM) was added to the growth media, with no further change at higher inhibitor concentrations (Sinha 2005). Protease inhibitors thus provide an alternative method to decrease proteolytic activity in the supernatant of yeast cultures.

The co-expression of protein-based protease inhibitors with the recombinant protein of interest may also prove to be an effective method to increase yields by decreasing proteolytic activity, provided the expression regulation mechanisms do not compete for cellular resources. Recently a recombinant secretory leukocyte protease inhibitor (SLPI) was successfully expressed in *P. pastoris* (Li 2009). Although this strategy has proven effective in plant systems (Rivard

2006, Goulet 2010), additional work remains to be done to confirm the effectiveness of this method in *P. pastoris*.

3.7.5 Effects of Medium Composition

The cultivation of *P. pastoris* on defined media is associated with a high rate of protease expression, especially in high cell density cultures. In defined media, nitrogen depletion results in an increase in protease activity which can be avoided by increasing the initial nitrogen concentration (Kobayashi 2000b). Nutrient starvation leads to autophagic cell degradation and lysis, leading to the release of vacuolar proteases (Todde 2009). Protein yields were shown to improve in several cultures grown on complex or amino acid-enriched media (Werten 1999, Shi 2003, Choi 2006). Complex or enriched media not only prevent nutrient limitation, but may inhibit protease activity by providing competing enzymatic substrates.

As previously discussed in section 3.7.1, the accumulation of ROS, by-products of the methanol metabolism, reduce cell viability and lead to the increased release of intracellular proteases. Adding ascorbic acid, an antioxidant, to the cultivation broth during the induction phase increases cell viability and significantly decreases proteolytic degradation of recombinant proteins (Xiao 2006).

3.8 Modeling of *P. pastoris* Cultivation Systems

Growth kinetics of yeast cultures are commonly described by models based on the Monod equation. The equation parameters vary considerably, with, for example, the maximum specific growth rate (μ_{max}) ranging from 0.046 L/h to 0.16 L/h for *P. pastoris* cultures (Cos 2007).

Growth inhibition occurs at high methanol concentrations. The specific growth rate (μ) of *P. pastoris* increases with the methanol concentration (S), until a critical level is reached above which any increase in methanol leads to a decline of the specific growth rate (Zhang 2000b). This behaviour is fitted to the uncompetitive inhibition growth model shown as equation 1:

$$\mu = \frac{\mu_{\max} S}{k_s + S + S^2 / K_I} \qquad (1)$$

Based on the critical methanol concentration, cell growth is divided into two phases: a growth limited region, in which μ increases with S, and a growth inhibited region during which μ decreases with S.

A kinetic model was developed describing both cell growth and oxygen consumption during the high density fed-batch cultivation of *P. pastoris*. In this model the specific methanol uptake rate is obtained directly from Monod's equation using the measured methanol concentration, and the specific growth rate is derived from the difference between the specific methanol uptake rate and maintenance energy requirements. To relate oxygen consumption to growth kinetics, the methanol flux was divided into two branches, one for cell growth, and the other for energy production. Different oxygen consumption rates were assigned to each branch (Jahic 2002).

For simplicity, unstructured models that assume constant cell concentrations are typically used. An unstructured model based on the Monod equation accurately describes the production of porcine insulin precursor (PIP) in *P. pastoris* (Hang 2008).

For more accurate descriptions of cells' non-linear behaviour, structured models can be used. A simple structured model describing the fed-batch cultivation of a strain of *P. pastoris* expressing a recombinant iduronate 2-sulphate sulfatase (IDShr) protein was proposed by Muñoz and coworkers (Muñoz 2008). In this model, the total biomass is divided into a viable fraction and a dead fraction. The former is further divided into the intracellular substrate fraction, the peroxisome fraction and the remaining components. The main characteristic of this model is the assumption that loss of cell viability and protease generation are caused by oxidative stress, which is proportional to the methanol uptake rate. Another structured kinetic model was developed by Celik and coworkers, describing the production of recombinant human erythropoietin (rHuEPO). *P. pastoris* was divided into rHuEPO, enzymatic and other cell component fractions. This model successfully simulated changes in rHuEPO production caused by variations in methanol feeding rates in all growth stages, and predicted the optimal feeding strategies to maximize yields (Çelik 2009).

The methanol and glycerol metabolic pathways, including phosphorylation, glycolysis, the TCA cycle and the respiratory chain, were incorporated into a macrokinetic model describing the production of rHSA in *P. pastoris*. The model is based on stoichiometric balances of the carbon source and the metabolites ATP and NADH. The specific uptake rates of glycerol and methanol were determined from the Monod model. The growth lag phase was described using a metabolic

regulator model. Similar regulator models were also used to describe the initial lag phase of protein production (Ren 2003).

The morphological features associated with budding during cellular growth of *P. pastoris* are similar to those of the yeast *Saccharomyces cerevisiae*. Experimental data show that heterologous proteins produced by *S. cerevisiae* are mainly secreted during the late G2 and M phases of cell growth (Frykman 2001). Although the relationship between foreign protein production and cell cycle phase in *P. pastoris* remains to be characterized, a cell cycle model describing fed-batch cultivation on glycerol was proposed, based on budding cell distribution data from *S. cerevisiae*. Experimental results show that the fraction of cells that are budding increases with the specific growth rate, and that these observations are well described by the proposed cell cycle model coupled with macrokinetic and bioreactor models. This model, however, is unable to describe the cell cycle distribution observed during the methanol induction phase, when multi-bud phenomena are observed (Jia 2007).

Taking into consideration the effects of operating conditions, Kupcsulik and Sevella formulated an empirical parabolic statistical model to describe the relation between the specific rHSA production rate and the pH and temperature of the cultivation process, which is shown as equation 2.

$$\mu_P = \mu_{P_0} \frac{K_1[H^+]}{K_1K_2 + K_1[H^+] + [H^+]^2} \left\{ a \cdot \exp\left(\frac{-\Delta G_1}{RT}\right) - b \cdot \exp\left(\frac{-\Delta G_2}{RT}\right) \right\} \qquad (2)$$

Parameters in this model were evaluated by fitting experimental results. The optimal conditions for *P. pastoris* cultivation predicted by the model were a pH of 5.64 and a temperature of 20.24°C (Kupcsulik 2005). Holmes and coworkers developed a predictive model of recombinant protein yield based on temperature, pH and DO concentration in the culture medium. When the protein yield is normalized to the volume and density of the culture, this model is scalable from ml to L working volumes (Holmes 2009).

Liang and Yuan established an oxygen transfer model for rHSA production by *P. pastoris* grown on glycerol and methanol. The model relates the oxygen transfer rate to the agitation speed, aeration rate, DO concentration, medium volume and medium temperature (Liang 2007). The oxygen uptake rate is determined with Ren's macrokinetic model (Ren 2003). Assuming that the oxygen transfer rate is in equilibrium with the oxygen uptake rate, this model is able to

estimate the biomass concentration given the system variables. Excellent predictions are achieved by coupling the model with a rolling updating strategy for the parameters.

In certain cases, data-driven models such as ANNs are more applicable than classic mechanical models based on mass balances. Data-driven models are capable of predicting process variables from global measurement variables such as the oxygen or carbon dioxide concentrations in the off-gas. Jin and coworkers developed an artificial neural network pattern recognition (ANNPR) model for the online adaptive control of methanol feeding based on DO and pH measurements to describe phytase production by a recombinant *P. pastoris*. The model predicts the state of the process as either "substrate starved" or "substrate in excess" and a coupled controller then adjusts the feed rate accordingly. This model-based control lead to a three-fold increase in phytase production, compared to systems using traditional DO-stat or online methanol electrode-based on–off control strategies (Jin 2007). The same strategy was used for the production of recombinant porcine interferon-α (Yu 2010).

The described models are only applicable to pAOX1-regulated systems. Far fewer models have been developed to describe pGAP-regulated ones. Tang and coworkers developed a simple model describing cell growth and recombinant phytase production in *P. pastoris* grown on glucose. Although based on data collected in continuous cultivations, the model provides good predictions of fed-batch behaviour under glucose-limited conditions (Tang 2010).

3.9 Economic Considerations

Although *P. pastoris* expression systems are more expensive and time-consuming to develop than their *E. coli* counterparts, once they are established the cost of recombinant protein production is comparable. The specific productivity of *P. pastoris* is generally low, but is compensated by the high cell densities achieved. For the production of therapeutic proteins, many of which require post-translational modifications, *P. pastoris* provides a more economically attractive alternative to mammalian or even insect cells (Vermasvuori 2009). Special economic considerations associated with *P. pastoris* include lower purification costs for secreted proteins and the license and royalty costs associated with common strains or vectors when developing commercial applications.

As previously discussed in this review, the cultivation method significantly affects the efficiency, and consequently the economic viability, of fermentation processes. Recombinant

protein production kinetics are highly dependent on the operating conditions and product-specific factors, and careful economic analyses must be performed during system design. It is important to note that the production rate and titre or volume of a process must be considered in conjunction with the cost of manufacture when comparing different processes, as the operational scale affects operating costs (Farid 2007).

An analysis of a projected pGAP-regulated *P. pastoris* fed-batch cultivation system expressing phytase was performed by our research group as part of an industrial report submitted to Zell Technologies Inc. (unpublished data). System performance and economics for cultivations on glucose and glycerol were compared for a process consisting of three 60-tonne fermenters. The phytase premix cost of manufacture for plants in the United States and China was also compared.

The total manufacturing costs for the 40% phytase premix produced in a US plant, including labour, materials, utilities and plant overhead, were estimated for fed-batch cultivation on glucose, glycerol and repeated fed-batch on glycerol. Repeated fed-batch refers to a hybrid of fed-batch and continuous processes, involving the removal of up to 75% of the culture volume once the exponential growth phase is completed (as determined by optical density analyses), and replacing it with fresh growth media. This strategy leads to increased yields compared to fed-batch cultivation as cellular waste and debris are removed and cells are maintained in the exponential growth phase. A separate analysis of repeated fed-batch cultivation on glycerol was performed for a China-based plant. The design basis and results are summarized in Table 3.1.

Due to the lower cost of glycerol compared to glucose, and given the comparable growth profiles on the two substrates, cultivations on glycerol are generally more economically attractive. Inexpensive sources of glycerol are increasingly becoming available. Tang and coworkers, for instance, demonstrated that the raw glycerol produced as waste during biodiesel production can be effectively used as the growth substrate in *P. pastoris* cultivations without the need for costly refinement (Tang 2009). The differences in manufacturing costs are not surprising, as non-US-based facilities have been shown to cost approximately 28% less than US-based ones (Farid 2007). It is interesting that total utilities, which include cooling and process water, steam and electricity, account for 20-30% of the total direct operating costs in each process.

Table 3.1. Comparative cost estimates for the production of 40% phytase premix by Pichia pastoris in fed-batch cultivations. All data taken from a report submitted to Zell Technologies Inc. (unpublished data)

Operation Mode	Substrate	Total Broth per Year (L)	Yearly production (lbs premix/yr)	Location of Plant	Manufacturing Cost (USD/lb premix)
Fed-batch	Glucose	16748000	5939000	US	1.84
Fed-batch	Glycerol	16748000	5939000	US	1.56
Repeated Fed-batch	Glycerol	31752000	11260000	US	1.06
Repeated Fed-batch	Glycerol (Local)	31752000	11260000	China	0.77
Repeated Fed-batch	Glycerol (imported)[a]	31752000	11260000	China	0.89

[a] Figures based on glycerol imported from the US

As evidenced by these results, changing the operating mode can significantly affect the viability of a process. Such optimization studies contribute to the development of low cost recombinant protein production. Additional measures to improve efficiency include genetic manipulation of the expression system to boost productivity or limit proteolysis, or minimize reagent requirements by conducting medium optimization studies.

Process costs are generally classified as either capital investment, based mostly on equipment costs, or costs of manufacturing which include raw materials, utilities and indirect costs such as insurance and depreciation. The former are typically estimated by multiplying the cost of a comparable existing plant by a corrective "Lang factor" generally between 3 and 5. The latter costs are generally estimated from process flow sheets, current market value of reagents, and empirically determined correlations based on total costs from comparable plants. An excellent review of cost estimation models and available software packages is provided by Farid (2007).

3.10 Update and Recent Developments

Although the basic aspects of *Pichia pastoris* bioprocesses discussed in this chapter remain unchanged beyond incremental improvements and the development of new monitoring or control systems or culture models, since this review paper was published (online) in 2010 some interesting progress has been achieved in regards to the development of novel expression

systems, mainly in terms of expanding the repertoire of strains, promoters and expression regulatory elements, which greatly increases the available options when designing or selecting a strain for a new bioprocess. A few of these developments are described here. The other aspects described in this chapter fall outside the direct scope of the research described in Chapters 6 and 7 of this thesis, and a comprehensive updated review of these topics is therefore not included here.

As discussed in section 3.3, one of the factors having the greatest impact on recombinant protein expression profiles is the promoter used to modulate production. Inducible promoters are used to uncouple growth and recombinant protein expression to minimize the metabolic load during exponential growth and which allow the controlled expression of cytotoxic or difficult to process products. Novel promoters that have proved effective and indeed provide viable alternatives to the pAOX systems include those for a Na+/phosphate symporter (pPHO89) induced by phosphate starvation (Ahn 2009), a gene in the thiamine biosynthesis pathway (pTHI11), which is repressed by thiamin (Stadlmayr 2010), and tightly regulated versatile promoters for alcohol dehydrogenase (pADH1), repressed by glucose and methanol, induced by glycerol and ethanol, enolase (pENO1), repressed by glucose, methanol and ethanol, induced by glycerol, and glycerol kinase (pGUT1) repressed by methanol, and induced by glucose, glycerol and ethanol (Cregg 2012). The characterization of these inducible promoters allow for a more versatile control of heterologous protein production.

Novel constitutive promoters, generally used to simplify process operation, include those of a glycosyl phophatidyl inositol-anchored protein (pGCW14) which produced yields 5- and 10-times greater than those obtained with the pGAP promoter when grown on glucose and glycerol respectively (Liang 2013a), of a glucose transporter (pG1) and an aldehyde dehydrogenase (pG6) which are repressed by glycerol, and induced by glucose limitation (Prielhofer 2013). In addition to identifying new promoter candidates, the continued modification and optimization of existing promoter sequences can greatly affect the product yields obtained, with, for example, a version of pGAP modified to produce 17-fold greater yields than the wild-type promoter (Qin 2011). A full list of the promoters used for recombinant protein production in *P. pastoris* and a discussion of the underlying metabolic mechanisms was published by Vogl et al. (2013).

Another approach to yield improvement involves optimizing recombinant protein secretory mechanisms, most commonly through modifications of signal peptide sequences through codon

optimization, directed evolution or deletion mutagenesis (Liang 2013b, Lin-Cereghino 2013). Considerable work has also been done on the development and modification of novel *P. pastoris* strains, and the last few years have seen a large increase in the number of available auxotrophic mutants, and considerable metabolic engineering has been performed to produce protease-deficient lines, as well as strains with a variety of glycosylation patterns and post-translational modifications. Although a detailed review of these subjects is again outside the scope of this work, the interest reader may refer to the excellent review by Ahmad et al. (2014).

3.11 Concluding Remarks

Ongoing research in *P. pastoris* cultivation is fuelling its continued development as one of the most versatile and powerful recombinant protein production systems available. Given its high yields, low cultivation costs and ability to produce functional proteins suitable for therapeutic use, it is not surprising that this system is of considerable industrial interest.

There is no single optimal cultivation method for recombinant protein production, as it is affected by a multitude of product-specific factors. Given this variability in product properties, the direct comparison of cultivation systems producing different proteins is difficult and of limited usefulness. Although several studies compare different cultivation methods for strains expressing the same heterologous products (Surribas 2007b, Yamawaki 2007, Trinh 2003, Minning 2001, Jahic 2003a, Trentmann 2004, Charoenrat 2005), there are currently no overarching studies comparing all common systems in depth, and a case-by-case approach to individual process design or optimization is required.

Despite its promise, several difficulties remain to be addressed before *P. pastoris* can be touted as a standard production system. As previously discussed, proteolytic degradation of certain recombinant proteins, exacerbated by protease over-expression and secretion when yeast cells are subjected to methanol-induced stress or nutrient deficiency, remains an obstacle to the development of economically viable large-scale operations. Another important limiting economic factor is the cost of downstream purification of recombinant proteins, even in product-secreting systems. Bioproducts destined for therapeutic use or human consumption have very strict purity requirements which may make the use of yeast expression systems economically onerous compared to systems generally regarded as safe (GRAS) such as green microalgae or certain species of plants. Although pAOX1 is a powerful promoter, the storage and handling of

large quantities of methanol is hazardous, and due to cytotoxic effects stringent control must be maintained over methanol concentrations in cultivation operations, which becomes increasingly difficult as the operational scale increases.

Although the present review focuses on operational methods to maximize recombinant product yields, several genetic-level strategies may prove useful in furthering the use of *P. pastoris* systems. As progress is made in the design of synthetic genetic regulation networks, for instance, novel 'artificial' promoters may be developed whose activity would be comparable to or surpass that of pAOX1 without the downsides related to methanol. As our understanding of the effect of amino-acid sequence on protein folding increases, computer models will be able to better predict the sensitivity of recombinant proteins to proteases, which will allow the design of genes whose products will be optimized to minimize degradation. Although much work remains to be accomplished, *P. pastoris* holds much promise as an expression system, further evidenced by the continuing interest and current dynamicity of research on the subject.

3.12 References

Ahmad M, Hirz M, Pichler H, Schwab H. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. Appl. Microbiol. Biotechnol. 2014; 98, 5301-17.

Ahn J, Hong J, Lee H, Park M, Lee E, Kim C, Choi E, Jung J, Lee H, Translation elongation factor 1-α gene from *Pichia pastoris*: molecular cloning, sequence, and use of its promoter, Appl. Genet. Mol. Biotechnol. 2007; 74, 601-8.

Ahn J, Hong J, Park M, Lee H, Lee E, Kim C, Lee J, Choi E, Jung J, Lee H. Phosphate-responsive promoter of a *Pichia pastoris* sodium phosphate symporter. Appl. Environ. Microbiol. 2009; 75, 3528-34.

Baker KN, Rendall MH, Patel A, Boyd P, Hoare M, Freedman RB, James DC, Rapid monitoring of recombinant protein products: a comparison of current technologies, Trends Biotechnol. 2002; 20.

Barr KA, Hopkins SA, Sreekrishna K. Protocol for efficient secretion of HAS developed form *Pichia pastoris*, Pharm. Eng. 1992; 12, 48–51.

Baumann K, Maurer M, Dragosits M, Cos O, Ferrer P, Mattanovich D, Hypoxic Fed-Batch Cultivation of *Pichia pastoris* Increases Specific and Volumetric Productivity of Recombinant Proteins, Biotechnol. Bioeng. 2008; 100 177-83.

Boer H, Teeri TT, Koivula A, Characterization of *Trichoderma reesei* Cellobiohydrolase Cel7A Secreted from *Pichia pastoris* Using Two Different Promoters, Biotechnol. Bioeng. 2000; 69, 486-94.

Boze H, Laborde C, Chemardin P, Richard F, Venturin C, Combarnous Y, Moulin G, High-level secretory production of recombinant porcine follicle-stimulating hormone by *Pichia pastoris*, Process Biochem. 2001; 36, 907-13.

Brierley RA, Bussineau C, Kosson R, Melton A, Siegel RS, Fermentation development of recombinant *Pichia pastoris* expressing the heterologous gene: bovine lysozyme, Ann. N.Y. Acad. Sci. 1990; 589, 350-62.

Bushell ME, Rowe M, Avignone-Rossa CA, Wardell JN, Cyclic fedbatch culture for production of human serum albumin in *Pichia pastoris*. Biotechnol. Bioeng. 2003; 82, 678–83.

Caballero M, Ruiz R, de Marquez PM, Seco M, Borque L, Escanero JF, Development of a microparticle-enhanced nephelometric immunoassay for quantitation of human lysozyme in pleural eVusion and plasma, J. Clin. Lab. Anal. 1999; 13, 301–7.

Cardello RJ, San KY, The design of controllers for batch bioreactors. Biotechnol. Bioeng. 1988; 32, 519-26.

Çelik E, Çalik P, Oliver SG, A structured kinetic model for recombinant protein production by Mut^+ strain of *Pichia pastoris*, Chem. Eng. Sci. 2009; 64, 5028-35.

Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM, Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*, Curr. Opin. Biotechnol. 2002; 13, 329-32.

Chaa HJ, Shin HS, Lim HJ, Cho HS, Dalal NN, Pham MQ, Bentley WE, Comparative production of human interleukin-2 fused with green fluorescent protein in several recombinant expression systems. Biochem. Eng. J. 2005; 24, 225-33.

Chang SW, Shieh CJ, Lee GC, Akoh CC, Shaw JF, Optimized Growth Kinetics of *Pichia pastoris* and Recombinant *Candida rugosa* LIP1 Production by RSM, J. Mol. Microbiol. Biotechnol. 2006; 11, 28-40.

Charoenrat T, Cairns MK, Andersen HS, Jahic M, Enfors SO, Oxygen-limited fed batch process: An alternative control for *Pichia pastoris* recombinant protein processes, Bioprocess. Biosyst. Eng. 2005; 27, 399-406.

Chen CC, Wu PH, Huang CT, Cheng KJ, *Pichia pastoris* fermentation strategy for enhancing the heterologous expression of an *Escherichia coli* phytase, Enzyme Microb. Technol.2004; 35, 315-20.

Chiruvolu V, Gregg JM, Meagher MM, Recombinant protein production in an alcohol oxidase-defective strain of *Pichia pastoris* in fedbatch fermentations, Enzyme Microb. Technol. 1997; 21, 277-83.

Choi DB, Park EY, Enhanced production of mouse α-amylase by feeding combined nitrogen and carbon sources in fed-batch culture of recombinant *Pichia pastoris*, Proc. Biochem. 2006; 41, 390-7.

Chung JD, Design of metabolic feed controllers: Application to high-density fermentations of *Pichia pastoris*, Biotechnol. Bioeng. 2000; 68, 298-307.

Clare JJ, Romanos MA, Rayment F, Rowedder JE, Smith MA, Payne MM, Sreekrishna K, Henwood CA, Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies, Gene. 1991; 105, 205-12.

Cos O, Resina D, Ferrer P, Montesinos JL, Valero F, Heterologous production of *Rhizopus oryzae* lipase in *Pichia pastoris* using the alcohol oxidase and formaldehyde dehydrogenase promoters in batch and fed-batch cultures, Biochem. Eng. J.2005; 26, 86-94.

Cos O, Ramón R, Montesinos JL, Valero F, Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: A review, Microb. Cell Fact. 2006; 5, 17.

Cos O, Ramon R, Montesinos JL, Valero V, A simple model-based control for *Pichia pastoris* allows a more efficient heterologous protein production bioprocess, Biotechnol. Bioeng.2006b; 95, 145-54.

Cregg JM, Vedvick TS, Raschke WC, Recent Advances in the Expression of Foreign Genes in *Pichia pastoris*, Nat. Biotechnol. 1993; 11, 905-10.

Cregg JM, Cereghino JL, Shi J, Higgins DR, Recombinant Protein Expression in *Pichia pastoris*, Mol. Biotechnol. 2000; 16, 23-52.

Cregg J, Tolstorukov I. *P. pastoris* ADH promoter and use thereof to direct expression of proteins. United States patent US 8222386. 2012.

Crowley J, Arnold SA, Wood N, Harvey LM, McNeil B, Monitoring a high cell density recombinant *Pichia pastoris* fed-batch bioprocess using transmission and reflectance near infrared spectroscopy, Enzyme Microb. Technol. 2005; 36, 621-8.

Cunha AE, Clemente JJ, Gomes R, Pinto F, Thomaz M, Miranda S, Pinto R, Moosmayer D, Donner P, Carrondo MJT, Methanol induction optimization for scFv antibody fragment production in *Pichia pastoris*, Biotechnol. Bioeng. 2004; 86, 458–67.

Curvers S, Brixius P, Klauser T, Thömmes J, Botz DW, Takors DR, Wandrey C, Human chymotrypsinogen B production with *Pichia pastoris* by integrated development of fermentation and downstream processing. Part 1. Fermentation, Biotechnol. Prog. 2001; 17, 495–502.

Curvers S, Linnemann J, Klauser T, Wandrey C, Takors R, Recombinant protein production with *Pichia pastoris* in continuous fermentation - kinetic analysis of growth and product formation, Eng Life Sci 2002; 2, 229-35.

Daly R, Hearn MTW, Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production, J. Mol. Recognit. 2005; 18, 119-38.

d'Anjou MC, Daugulis AJ, A Rational Approach to Improving Productivity in Recombinant *Pichias pastoris* Fermentation, Biotechnol. Bioeng. 2000; 72, 1-11.

de Almeida JRM, de Moraes LMP, Torres FAG, Molecular characterization of the 3-phosphoglycerate kinase gene (*PGK1*) from the methylotrophic yeast *Pichia pastoris*, Yeast. 2005; 22, 725-37.

Delroisse JM, Dannau M, Gilsoul JJ, El Mejdoub T, Destain J, Portetelle D, Thonart P, Haubruge E, Vandenbol M, Expression of a synthetic gene encoding a *Tribolium castaneum* carboxylesterase in *Pichia pastoris*, Protein Expression Purif.2005; 42, 286-94.

Döring F, Klapper M, Theis S, Daniel H, Use of the Glyceraldehyde-3-phosphate Dehydrogenase Promoter for Production of Functional Mammalian Membrane Transport Proteins in the Yeast *Pichia pastoris*, Biochem. Biophys. Res. Commun.1998; 250, 531-5.

Duan HM, Umar S, Hu YP, Chen JC, Both the AOX1 promoter and the FLD1 promoter work together in a *Pichia pastoris* expression vector, World J. Microbiol. Biotechnol. 2009; 25, 1779-83.

Farid SS, Process economics of industrial monoclonal antibody manufacture, J. Chromatogr. B. 2007; 848, 8-18.

Frykman S, Srienc F, Cell Cycle-Dependent Protein Secretion by *Saccharomyces cerevisiae*, Biotechnol. Bioeng. 2001; 76, 259-68.

Gadowski L, Abdul-Wajid A, Quantitation of monoclonal antibodies by perfusion chromatography–immunodetection, J. Chromatogr. A 1995; 715, 241–45.

Gao MJ, Shi ZP. Process control and optimization for heterologous protein production by methylotrophic *Pichia pastoris*. Chinese J. Chem. Eng. 2013; 21, 216-26.

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

Gellissen G, Heterologous protein production in methylotrophic yeasts, Appl. Microbiol. Biotechnol. 2000; 54, 741-50.

Gill A, Bracewell DG, Maule CH, Lowe PA, Hoare M, Bioprocess monitoring: an optical biosensor for rapid bioproduct analysis, J. Biotechnol. 1998; 65, 69–80.

Goodrick JC, Xu M, Finnegan R, Schilling BB, Schiavi S, Hoppe H, Wan NC, High-Level Expression and Stabilization of Recombinant Human Chitinase Produced in a Continuous Constitutive *Pichia pastoris* Expression System, Biotechnol. Bioeng. 2001; 74, 492-7.

Goulet C, Benchabane M, Anguenot R, Brunelle F, Khalf M, Michaud D, A companion protease inhibitor for the protection of cytosol-targeted recombinant proteins in plants, Plant Biotechnol. J. 2010; 8, 142-54.

Graf A, Dragosits M, Gasser B, Mattanovich D, Yeast systems biotechnology for the production of heterologous proteins, FEMS Yeast Res, 2009; 9, 335-48.

Grimshaw C, Gleason C, Chojnicki E, Young J, Development of an equilibrium immunoassay using electrochemiluminescent detection for a novel recombinant protein product and its application to pre-clinical product development, J. Pharm. Biomed. Anal. 1997; 16, 605–12.

Guarna MM, Lesnicki GJ, Tam BM, Robinson J, Radziminski CZ, Hasenwinkle D, Boraston A, Jervis E, MacGillivray RTA, Turner RFB, Kilburn DG, On-line monitoring and control of methanol concentration in shake-flasks cultures of *Pichia pastoris*, Biotechnol Bioeng. 1997; 56, 279-86.

Gurkan C, Ellar DJ, Recombinant production of bacterial toxins and their derivatives in the methylotrophic yeast *Pichia pastoris*, Microb. Cell Fact. 2005; 4, 33.

Gurramkonda C, Adnan A, Gäbel T, Lünsdorf H, Ross A, Nemani SK, Swaminathan S, Khanna N, Rinas U, Simple high-cell density fed-batch technique for high-level recombinant protein production with *Pichia pastoris*: Application to intracellular production of Hepatitis B surface antigen, Microb. Cell Fact. 2009; 8, 13.

Gustafsson C, Govindarajan S, Minshull J, Codon bias and heterologous protein expression, Trends Biotechnol. 2004; 22, 346-53.

Hang HF, Chen W, Guo MJ, Chu J, Zhuang YP, Zhang SL, A simple unstructured model-based control for efficient expression of recombinant porcine insulin precursor by *Pichia pastoris*, Korean J. Chem Eng. 2008; 25, 1065-9.

Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A, Promoter library designed for fine-tuned gene expression in *Pichia pastoris*, Nucleic Acids Res. 2008; 36, e76.

He X, Liu N, Li W, Zhang Z, Zhang B, Ma Y, Inducible and constitutive expression of a novel thermostable alkaline β-mannanase from alkaliphilic *Bacillus* sp. N16-5 in *Pichia pastoris* and characterization of the recombinant enzyme, Enzyme Microb. Technol. 2008; 43, 13-8.

Hellwig S, Emde F, Raven NPG, Henke M, van der Logt P, Fischer R, Analysis of single-chain antibody production in *Pichia pastoris* using on-line methanol control in fed-batch and mixed-feed fermentations, Biotechnol. Bioeng. 2001; 74, 344–52.

Hisiger S, Jolicoeur M, A multiwavelength fluorescence probe: Is one probe capable for on-line monitoring of recombinant protein production and biomass activity?, J. Biotech. 2005; 117, 325-32.

Holmes WJ, Darby RAJ, Wilks MDB, Smith R, Bill RM, Developing a scalable model of recombinant protein yield from *Pichia pastoris*: the influence of culture conditions, biomass and induction regime, Microb. Cell Fact. 2009; 8, 35.

Hong F, Meinander NQ, Jönsson LJ, Fermentation strategies for improved heterologous expression of laccase in Pichia pastoris. Biotechnol Bioeng 2002; 79, 438-49.

Hu XQ, Chu J, Zhang Z, Zhang SL, Zhuang YP, Wang YH, Guo MJ, Chen HX, Yuan ZY, Effects of different glycerol feeding strategies on S-adenosyl-l-methionine biosynthesis by PGAP-driven Pichia pastoris overexpressing methionine adenosyltransferase, J. Biotechnol. 2008; 137, 44–9.

Idiris A, Tohda H, Kumagai H, Takegawa K, Engineering of protein secretion in yeast : strategies and impact on protein production, Appl. Microbiol. Biotechnol. 2010; 86, 403-17.

Inan M, Chiruvolu V, Eskridge KM, Vlasuk GP, Dickerson K, Brown S, Meagher MM, Optimization of temperature-glycerol-pH conditions for a fed-batch fermentation process for recombinant hookworm (*Ancylostoma caninum*) anticoagulant peptide (AcAP-5) production by *Pichia pastoris*. Enzyme Microb. Technol. 1999; 24, 438-45.

Inan M, Meagher MM, The effect of ethanol and acetate on protein expression in *Pichia pastoris*, J. Biosci. Bioeng. 2001; 92, 337-41.

Invitrogen, *Pichia* Fermentation Process Guidelines Version B 05/30/02. Available online: http://tools.invitrogen.com/content/sfs/manuals/pichiaferm_prot.pdf, last accessed 10/05/10.

Jahic M, Rotticci-Mulder JC, Martinelle M, Hult K, Enfors SO, Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein, Bioprocess. Biosyst. Eng. 2002; 24, 385-93.

Jahic M, Gustavsson M, Jansen AK, Martinelle M, Enfors SO, Analysis and control of proteolysis of a fusion protein in *Pichia pastoris* fed-batch processes, J. Biotechnol. 2003a; 102, 45-53.

Jahic M, Wallberg F, Bollok M, Garcia P, Enfors SO, Temperature limited fed-batch technique for control of proteolysis in *Pichia pastoris* bioreactor cultures, Microb. Cell Fact. 2003b; 2, 6.

Jenzsch M, Simutis R, Eisbrenner G, Stückrath I, Lübbert A, Estimation of biomass concentrations in fermentation processes for recombinant protein production, Bioprocess Biosyst. Eng. 2006; 29, 19-27.

Jia L, Yuan JQ, Cell cycle model for recombinant *Pichia pastoris* during glycerol fed-batch cultivation, Process Biochem. 2007; 42, 828-33.

Jin H, Zheng ZY, Gao MJ, Duan ZY, Shi ZP, Wang ZX, Jin J, Effective induction of phytase in *Pichia pastoris* fed-batch culture using an ANN pattern recognition model-based on-line adaptive control strategy, Biochem. Eng. J. 2007; 37, 26-33.

Jones LJ, Upson RH, Haugland RP, Panchuk-Voloshina N, Zhou M, Haugland RP, Quenched BODIPY Dye-Labeled Casein Substrates for the Assay of Protease Activity by Direct Fluorescence Measurement, Anal. Biochem. 1997; 251, 144-52.

Jones JJ, Bridges AM, Fosberry AP, Gardner S, Lowers RR, Newby RR, James PJ, Hall RM, Jenkins O, Potential of real-time measurement of GFP-fusion proteins, J. Biotech. 2004; 109, 201-11.

Jung JM, Shin YB, Kim MG, Ro HS, Jung HT, Chung BH, A fusion protein expression analysis using surface plasmon resonance imaging, Anal. Biochem. 2004; 330, 251–6.

Jungo C, Rérat C, Marison IW, von Stockar U, Quantitative characterization of the regulation of the synthesis of alcohol oxidase and of the expression of recombinant avidin in a *Pichia pastoris* Mut^+ strain, Enzyme Microb. Technol. 2006; 39, 936-44.

Jungo C, Marison I, von Stockar U, Regulation of alcohol oxidase of a recombinant *Pichia pastoris* Mut+ strain in transient continuous cultures, J. Biotechnol. 2007a; 130, 236-46.

Jungo C, Urfer J, Zocchi A, Marison I, von Stockar U, Optimisation of culture conditions with respect to biotin requirement for the production of recombinant avidin in *Pichia pastoris*, J. Biotechnol. 2007b; 127, 703-15.

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

Katakura Y, Zhang WH, Zhuang GQ, Omasa T, Kishimoto M, Goto W, Suga KI, Effect of methanol concentration on the production of human beta(2)-glycoprotein I domain V by a recombinant *Pichia pastoris*: A simple system for the control of methanol concentration using a semiconductor Gas Sensor, J. Ferment. Bioeng. 1998; 86, 482-7.

Khasa YP, Khushoo A, Srivastava L, Mukherjee KJ, Kinetic studies of constitutive human granulocyte-macrophage colony stimulating factor (hGM-CSF) expression in continuous culture of *Pichia pastoris*, Biotechnol. Lett. 2007; 29, 1903-8.

Khatri NK, Hoffmann F, Oxygen-limited control of methanol uptake for improved production of a single-chain antibody fragment with recombinant *Pichia pastoris*, Appl. Microbiol. Biotechnol. 2006a; 72, 492-8.

Khatri NK, Hoffmann F, Impact of methanol concentration on secreted protein production in oxygen-limited cultures of recombinant *Pichia pastoris*, Biotechnol. Bioeng. 2006b; 93, 871-9.

Kiviharju K, Salonen K, Moilanen U, Eerikäinen T, Biomass measurement online: The performance of in-situ measurements and software sensors, J. Ind. Microbiol. Biotechnol. 2008; 35, 657-65.

Kim SJ, Lee JA, Kim YH, Song BK, Optimization of the Functional Expression of Coprinus cinereus Peroxidase in *Pichia pastoris* by Varying the Host and Promoter, J. Microbiol. Biotechnol. 2009; 19, 966-71.

Kobayashi K, Kuwae S, Ohya T, Ohda T, Ohyama M, Ohi H, Tomomitsu K, Ohmura T, High-level Expression of Recombinant human Serum Albumin from the Methylotrophic yeast *Pichia pastoris* with Minimal protease Production and Activation, J. Biosci. Bioeng. 2000a; 89, 55-61.

Kobayashi K, Kuwae S, Ohya T, Ohda T, Ohyama M, Tomomitsu K, High Level Secretion of Recombinant Human Serum Albumin by Fed-Batch Fermentation of the Methylotrophic Yeast, *Pichia pastoris*, Based on Optimal Methanol Feeding Strategy, J. Biosci. Bioeng. 2000b; 90, 280-8.

Krause M, Ukkonen K, Haataja T, Ruottinen M, Glumoff T, Neubauer A, Neubauer P, Vasala A, A novel fed-batch based cultivation method provides high cell-density and improves yield of soluble recombinant proteins in shaken cultures, Microb. Cell Fact. 2010; 9, 11.

Kupcsulik B, Sevella B, Optimization of Specific Product Formation Rate by Statistical and Formal Kinetic Model Descriptions of an HSA producing *Pichia pastoris* Mut^s strain, Chem. Biochem. Eng. 2005; 19, 99-108.

Kuwae S, Ohyama M, Ohya T, Ohi H, Kobayashi K, Production of Recombinant Human Antithrombin by *Pichia pastoris*, J. Biosci. Bioeng. 2005; 99, 264-71.

Ledue TB, Weiner DL, Sipe JD, Poulin SE, Collins MF, Rifai N, Analytical evaluation of particle-enhanced immunonephelometric assays for C-reactive protein, serum amyloid A, and mannose-binding protein in human serum, Ann. Clin. Biochem. 1998; 35, 745–53.

Lee CY, Nakano A, Shiomi N, Lee EK, Katoh S, Effects of substrate feed rate on heterologous protein expression by Pichia pastoris in DO-stat fed-batch fermentation, Enzyme Microb. Technol. 2003; 33, 358–65.

Li Z, Xiong F, Lin Q, d'Anjou M, Daugulis AJ, Yang DSC, Hew CL, Low-Temperature Increases the Yield of Biologically Active Herring Antifreeze Protein in *Pichia pastoris*, Protein Expression Purif. 2001; 21, 438-45.

Li P, Anumanthan A, Gao XG, Ilangovan K, Suzara VV, Düzgünes N, Renugopalakrishnan V, Expression of Recombinant Proteins in *Pichia pastoris*, Appl. Biochem. Biotechnol. 2007; 142, 105-24.

Li ZG, Moy A, Sohal K, Dam C, Kuo P, Whittaker J, Whittaker M, Düzgünes N, Konopka K, Franz AH, Lin-Cereghino JL, Lin-Cereghino GP, Expression and characterization of recombinant human secretory leukocyte protease inhibitor (SLPI) protein from *Pichia pastoris*, Protein Expression Purif. 2009; 67, 175-81.

Liang J, Yuan J, Oxygen transfer model in recombinant *Pichia pastoris* and its application in biomass estimation, Biotechnol. Lett. 2007; 29, 27-35.

Liang S, Zou C, Lin Y, Zhang X, Ye Y. Identification and characterization of P GCW14 : a novel, strong constitutive promoter of *Pichia pastoris*. Biotechnol. Lett. 2013a; 35, 1865-71.

Liang S, Li C, Ye Y, Lin Y. Endogenous signal peptides efficiently mediate the secretion of recombinant proteins in *Pichia pastoris*. Biotechnol. Lett. 2013b; 35, 97-105.

Lim HK, Choi SJ, Kim KY, Jung KH, Dissolved-oxygen-stat controlling two variables for methanol induction of rGuamerin in *Pichia pastoris* and its application to repeated fed-batch, Appl. Microbiol. Biotechnol. 2003; 62, 342-8.

Lin-Cereghino GP, Stark CM, Kim D, Chang JWJ, Shaheen N, Poerwanto H, Agari K, Moua P, Low LK, Tran N, Huang AD, Nattestad M, Oshiro KT, Chavan A, Tsai JW, Lin-Cereghino J. The effect of α-matin factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. Gene. 2013; 519, 311-7.

Liu H, Tan X, Russel KA, Veenhuis M, Cregg JM, *PER3*, a Gene Required for Peroxisome Biogenesis in *Pichia pastoris*, Encodes a Peroxisomal Membrane Protein Involved in Protein Import, J. Biol. Chem. 1995; 270, 10940-51.

Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM, Heterologous protein production using the *Pichia pastoris* expression system, Yeast. 2005; 22, 249-70.

Maurer M, Kühleitner M, Gasser B, Mattanovich D, Versatile modeling and optimization of fed-batch processes for the production of secreted heterologous proteins with *Pichia pastoris*, Microb. Cell Fact. 2006; 5, 37.

Menendez J, Valdes I, Cabrera N, The *ICL1* gene of *Pichia pastoris*, transcriptional regulation and use of its promoter, Yeast. 2003; 20, 1097-1108.

Minning S, Serrano A, Ferrer P, Sola C, Schmid RD, Valero F, Optimization of the high-level production of Rhizopus oryzae lipase in *Pichia pastoris*, J Biotechnol. 2001; 86, 59–70.

Mochizuki S, Hamato N, Hirose M, Miyano K, Ohtani W, Kameyama S, Kuwae S, Tokuyama T, Ohi H, Expression and Characterization of Recombinant Human Antithrombin III in *Pichia pastoris*, Protein Expression Purif. 2001; 23, 55-65.

Muñoz DFR, Enciso NAA, Ruiz HC, Avellaneda LAB, A simple structured model for recombinant IDShr protein, Biotechnol. Lett. 2008; 30, 1727-34.

Nakano A, Lee CY, Yoshida A, Matsumoto T, Shiomi N, Katoh S, Effects of Methanol Feeding Methods on Chimeric α-Amylase Expression Continuous Culture of *Pichia pastoris*, J. Biosci. Bioeng. 2006; 101, 227-31.

Noronha EF, de Lima BD, de Sá CM, Felix CR, Heterologous production of Aspergillus fumigatus keratinase in *Pichia pastoris*, World J. Microbiol. Biotechnol. 2002; 18, 563-8.

Ödman P, Johansen CL, Olsson L, Gernaey KV, Lantz AE, On-line estimation of biomass, glucose and ethanol in Saccharomyces cerevisiae cultivations using in-situ multi-wavelength fluorescence and software sensors, J. Biotechnol. 2009; 144, 102-12.

Ohya T, Morita M, Miura M, Kuwae S, Kobayashi K, High-level production of prourokinase-annexin V chimeras in the methylotrophic yeast *Pichia pastoris*, J. Biosci. Bioeng. 2002; 94, 467-73.

Oliveira R, Clemente JJ, Cunha AE, Carrondo MJT. Adaptive dissolved oxygen control through the glycerol feeding in a recombinant *Pichia pastoris* cultivation in conditions of oxygen transfer limitation, J. Biotechnol. 2005; 116, 35-50.

Pal Y, Khushoo A, Mukherjee KJ, Process optimization of constitutive human granulocyte-macrophage colony-stimulating factor (hGM-CSF) expression in *Pichia pastoris* fed-batch culture, Appl. Microbiol. Biotechnol. 2006; 69, 650-7.

Panula-Perälä J, Šiurkus J, Vasala A, Wilmanowski R, Casteleijn MG, Neubauer P, Enzyme controlled glucose auto-delivery for high cell density cultivations in microplates and shake flasks, Microb. Cell Fact. 2008; 7, 31.

Pavlou AK, Reichert JM, Recombinant protein therapeutics : success rates, market trends and values to 2010, Nat. Biotechnol. 2004; 22, 1513-9.

Plantz BA, Sinha J, Villarete L, Nickerson KW, Schlegel VL, *Pichia pastoris* fermentation optimization: Energy state and testing a growth-associated model, Appl. Microbiol. Biotechnol. 2006; 72, 297-305.

Prielhofer P, Maurer M, Klein J, Wenger J, Kiziak C, Gasser B, Mattanovich D. Induction without methanol: novel regulated promoters enable high-level expression in *Pichia pastoris*. Microb. Cell Factories. 2013; 12, 5.

Qin X, Qian J, Yao G, Zhuang Y, Zhang S, Chu J. GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. Appl. Environ. Microbiol. 2011; 77, 3600-8.

Qureshi MS, Zhang DX, Du GC, Chen J, Improved production of polygalacturonate lyase by combining a pH and online methanol control strategy in a two-stage induction phase with a shift in the transition phase, J. Ind. Microbiol. Biotechnol. 2010; 37, 323-33.

Reischer H, Schotola I, Striedner G, Potschacher F, Bayer K, Evaluation of the GFP signal and its aptitude for novel on-line monitoring strategies of recombinant fermentation processes, J. Biotech. 2004; 108, 115-25.

Ren HT, Yuan JQ, Bellgardt KH, Macrokinetic model for methylotrophic *Pichia pastoris* based on stoichiometric balance, J. Biotechnol. 2003; 106, 53-68.

Ren HT, Yuan J, Model-based specific growth rate control for *Pichia pastoris* to improve recombinant protein production,. J Chem. Technol. Biotechnol. 2005; 80, 1268-72.

Resina D, Cos O, Ferrer P, Valero F, Developing High Cell Density Fed-Batch Cultivation Strategies for Heterologous Protein Production in *Pichia pastoris* Using the Nitrogen Source-Regulated FLD1 promoter, Biotechnol. Bioeng. 2005; 91, 760-7.

Resina D, Maurer M, Cos O, Arnau C, Carnicer M, Marx H, Gasser B, Valero F, Mattanovich D, Ferrer P, Engineering of bottlenecks in *Rhizopus oryzae* lipase production in *Pichia pastoris* using the nitrogen source-regulated FLD1 promoter, New Biotechnol. 2009; 25, 396-403

Rivard D, Anguenot R, Brunelle F, Le VQ, Vézina LP, Trépanier S, Michaud D, An in-built proteinase inhibitor system for the protection of recombinant proteins recovered from transgenic plants, Plant Biotechnol. J. 2006; 4 359-68.

Ruiz J, Pinsach J, Álvaro G, Gonzaĺez G, de Mas C, Resina D, López-Santín J, Alternative production process strategies in E. coli improving protein quality and downstream yields, Process Biochem. 2009; 44, 1039–45.

Ruottinen M, Bollok M, Kögler M, Neubauer A, Krause M, Hämäläinen ER, Myllyharju J, Vasala A, Neubauer P, Improved production of human type II procollagen in the yeast *Pichia pastoris* in shake flasks by a wireless-controlled fed-batch system, BMC Biotech. 2008; 8, 1-12.

Sanchez S, Demain AL. Enzymes and bioconversions of industrial, pharmaceutical and biotechnological significance. Org. Process. Res. Dev. 2011; 15, 224-30.

Schenk J, Marison IW, von Stockar U, A simple method to monitor and control methanol feeding of *Pichia pastoris* fermentations using mid-IR spectroscopy, J. Biotechnol. 2007; 128, 344-53.

Schilling BM, Goodrick JC, Wan NC, Scale-Up of a High Cell Density Continuous Culture with *Pichia pastoris* X-33 for the Constitutive Expression of rh-Chitinase, Biotechnol. Prog. 2001; 17, 629-33.

Sears IB, O'Connor J, Rossanese OW, Glick BS, A Versatile Set of Vectors for Constitutive and Regulated Gene Expression in *Pichia pastoris*, Yeast. 1998; 14, 783-90.

Segev N, Mulholland J, Botstein D, The Yeast GTP-Binding YPT1 Protein and a Mammalian Counterpart Are Associated with the Secretion Machinery, Cell. 1988; 52, 915-24.

Seidler NW, Basic biology of GAPDH, Adv. Exp. Med. Biol. 2013; 985, 1-36.

Shen S, Sulter G, Jeffries TW, Cregg JM, A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*, Gene, 1998; 216, 93-102.

Shi XZ, Karkut T, Chamankhah M, Alting-Mees M, Hemmingsen SM, Hegedus D, Optimal conditions for the expression of a single-chain antibody (scFv) gene in *Pichia pastoris*, Protein Expression Purif. 2003; 28, 321-30.

Siegel RS, Brierley RA, Methylotrophic yeast *Pichia pastoris* produced in high-cell density fermentations with high cell yields as vehicle for recombinant protein production, Biotechnol Bioeng 1989; 34, 403–4.

Silva VC, Peres MFS, Gattas EAL, Application of methylotrophic yeast *Pichia pastoris* in the field of food industry – A review, J. Food Agric. Environ. 2009; 7, 268-73.

Sinclair G, Choy FYM, Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*, Protein Expression Purif. 2002; 26, 96-105.

Singh S, Gras A, Vandal CF, Ruprecht J, Rana R, Martinez M, Strange PG, Wagner R, Byrne B, Large-scale functional expression of WT and truncated human adenosine A2A receptor in *Pichia pastoris* bioreactor cultures, Microb. Cell Fact.2008; 7, 28.

Sinha J, Plantz BA, Zhang W, Gouthro M, Schlegel VL, Liu CP, Meagher MM, Improved production of recombinant ovine interferon-t by Mut+ strain of *Pichia pastoris* using an optimized methanol feed profile, Biotechnol. Prog. 2003; 19, 794-802.

Sinha J, Plantz BA, Inan M, Meagher MM, Causes of proteolytic Degradation of Secreted Recombinant Proteins Produced in Methylotrophic Yeast *Pichia pastoris*: Case Study With Recombinant Ovine Interferon-T, Biotechnol. Bioeng. 2005; 89, 102-12.

Sirén N, Weegar J, Dahlbacka J, Kalkkinen N, Fagervik K, Leisola M, von Weymarn N, Production of recombinant HIV-1 Nef (negative factor) protein using *Pichia pastoris* and a low-temperature fed-batch strategy, Biotechnol. Appl. Biochem. 2006; 44, 151-8.

Solà A, Jouhten P, Maaheima H, Sánchez-Ferrando F, Szyperski T, Ferrer P, Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates, Microbiol. 2007; 153, 281-90.

Sreekrishna K, Brankamp RB, Kropp KE, Blankenship DT, Tsay JT, Smith PL, Wierschke JD, Subramaniam A, Birkenberger LA, Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*, Gene. 1997; 190, 55-62.

Stadlmayr G, Mecklenbrauker A, Rothmuller M, Maurer M, Sauer M, Mattanovich D, Gasser B. Identification and characterisation of novel *Pichias pastoris* promoters for heterologous protein production. J. Biotechnol. 2010; 150, 519-29.

Surribas A, Cos O, Montesinos JL, Valero F, On-line monitoring of the methanol concentration in *Pichia pastoris* cultures producing an heterologous lipase by sequential injection analysis. Biotechnol. Lett. 2003; 25, 1795-800.

Surribas A, Geissler D, Gierse A, Scheper T, Hitzmann B, Montesinos JL, Valero F, State variables monitoring by in-situ multi-wavelength fluorescence spectroscopy in heterologous protein production by *Pichia pastoris*, J. Biotechnol. 2006; 124, 412-9.

Surribas A, Resina D, Ferrer P, Valero F, Rivoflavin may interfere with on-line monitoring of secreted green fluorescence protein fusion proteins in *Pichia pastoris*, Microbial Cell Factories 2007a; 6, 1-7.

Surribas A, Stahn R, Montesinos JL, Enfors SO, Valero F, Jahic M, Production of a Rhizopus oryzae lipase from *Pichia pastoris* using alternative operational strategies, J. Biotechnol. 2007b; 130, 291-9.

Tang SQ, Boehme L, Lam H, Zhang Z, *Pichia pastoris* fermentation for phytase production using crude glycerol from biodiesel production as the sole carbon source, Biochem. Eng. J. 2009; 43, 157-162.

Tang SQ, Potvin G, Reiche A, Zhang Z, Modeling of Phytase Production by Cultivation of *Pichia pastoris* Under the Control of the GAP Promoter, Int. J. Chem. Reactor Eng. 2010; 8, A9.

Todde V, Veenhuis M, van der Klei IJ, Autophagy: Principles and significance in health and disease, Biochim. Biophys. Acta. 2009; 1792, 3-13.

Tojo N, Miyagi I, Miura M, Ohi H, Recombinant human fibrinogen expressed in the yeast *Pichia pastoris* was assembled and biologically active, Protein Expression Purif. 2008; 59, 289-96.

Trentmann O, Khatri NK, Hoffmann F, Reduced oxygen supply increases process stability and product yield with recombinant *Pichia pastoris*. Biotechnol. Prog. 2004; 20, 1766–75.

Trinh LB, Phue JN, Shiloah J, Effect of methanol feeding strategies on production and yield of recombinant mouse endostatin from *Pichia pastoris*, Biotechnol. Bioeng. 2003; 82, 438-44.

Tsoka S, Gill A, Brookman JL, Hoare M, Rapid monitoring of virus-like particles using an optical biosensor: a feasibility study, J. Biotechnol. 1998; 63, 147–53.

Vassileva A, Chugh DA, Swaminathan S, Khanna N, Expression of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris* using the *GAP* promoter, J. Biotechnol. 2001; 88, 21-35.

Vermasvuori R, Koskinen J, Salonen K, Sirén N, Weegar J, Dahlbacka J, Kalkkinen N, von Weymarn N, Production of Recombinant HIV-1 Nef Protein Using Different Expression Host Systems: A Techno-Economical Comparison, Biotechnol. Progr. 2009; 25, 95-102.

Vogl T, Glieder A. Regulation of *Pichia pastoris* promoters and its consequences for protein production. Nat. Biotechnol. 2013; 30, 385-404.

Wagner LW, Matheson NH, Heisey RF, Schneider K, Use of a silicone tubing sensor to control methanol concentration during fed batch fermentation of *Pichia pastoris*, Biotechnol. Tech. 1997; 11, 791-5.

Walsh G. Biopharmacetucial benchmarks 2010. Nat. Biotechnol.2010; 28, 917-24.

Wang J, Nguyen V, Glen J, Henderson B, Saul A, Miller LH, Improved Yield of Recombinant Merozoite Surface protein 3 (MSP3) from *Pichia pastoris* Using Chemically Defined Media, Biotechnol. Bioeng. 2005; 90, 838-47.

Waterham HR, Digan ME, Koutz PE, Lair SV, Cregg JM, Isolation of the *Pichia pastoris* glyceraldehydes-3-phosphate dehydrogenase gene and regulation and use of its promoter, Gene. 1997; 186, 37-44.

Werten MWT, van den Bosch TJ, Wind RD, Mooibroek H, de Wolf FA, High-yield Secretion of Recombinant Gelatins by *Pichia pastoris*, Yeast. 1999; 15, 1087-96.

Welch M, Villalobos A, Gustafsson C, Minshull J, You're one in a googol: optimizaing genes for protein expression, J. R. Soc. Interface. 2009; 6, S467-76.

Woo SH, Park SH, Lim HK, Jung KH, Extended operation of a pressurized 75-L bioreactor for shLkn-1 production by *Pichia pastoris* using dissolved oxygen profile control, J. Ind. Microbiol. Biotechnol. 2005; 32, 474-80.

Wu JM, Lin JC, Chieng LL, Lee CK, Hsu TA, Combined use of *GAP* and *AOX1* promoter to enhance the expression of human granulocyte-macrophage colony-stimulating factor in *Pichia pastoris*, Enzyme Microb. Technol. 2003a; 33, 453-9.

Wu JM, Chieng LL, Hsu TA, Lee CK, Sequential expression of recombinant proteins and their separate recovery from a *Pichia pastoris* cultivation, Biochem. Eng. J. 2003b; 16, 9-16.

Wu D, Ma D, Hao YY, Chu J, Wang YH, Zhuang YP, Zhang SL, Incomplete formation of intramolecular disulfide bond triggers degradation and aggregation of human consensus interferon-α mutant by *Pichia pastoris*, Appl. Microbiol. Biotechnol. 2010; 85, 1759-67.

Wu D, Hao YY, Chu J, Zhuang YP, Zhang SL, Inhibition of degradation and aggregation of recombinant human consensus interferon-α mutant expressed in *Pichia pastoris* with complex medium in bioreactor, Appl. Microb. Cell Physiol. 2008; 80, 1063-71.

Xiao A, Zhou X, Zhou L, Zhang Y, Improvement of cell viability and hirudin production by ascorbic acid in *Pichia pastoris* fermentation, Appl. Microbiol. Biotechnol. 2006; 72, 837-44.

Xie JL, Zhou QW, Pen D, Gan RB, Qin Y, Use of different carbon sources in cultivation of recombinant *Pichia pastoris* for angiostatin production, Enzyme Microb. Technol. 2005; 36, 210-6.

Yamashita S, Yurimoto H, Murakami D, Yoshikawa M, Oku M, Sakai Y, Lag-phase autophagy in the methylotrophic yeast *Pichia pastoris*, Genes Cells. 2009; 14, 861-70.

Yamawaki S, Matsumoto T, Ohnishi Y, Kumada Y, Shiomi N, Katsuda T, Lee EK, Katoh S, Production of Single-Chain Variable Fragment Antibody (scFv) in Fed-Batch and Continuous Culture of *Pichia pastoris* by Two Different Methanol Feeding Methods, J. Biosci. Bioeng. 2007; 104, 403-7.

Yu RS, Dong SJ, Zhu YM, Jin H, Gao MJ, Duan ZY, Zheng ZY, Shi ZP, Li Z, Effective and stable porcine interferon-α production by *Pichia pastoris* fed-batch cultivation with multi-variables clustering and analysis, Bioprocess Biosyst. Eng. 2010; 33, 473-83.

Zhang J, Zhou H, Ji Z, Regnier F, Monoclonal antibody production with on-line harvesting and process monitoring, J. Chromatogr. B Biomed. Sci. Appl. 1998; 707, 257–65.

Zhang W, Inan M, Meagher MM, Fermentation strategies for recombinant protein expression in the methylotrophic yeast *Pichia pastoris*, Biotechnol. Bioprocess Eng. 2000a; 5, 275-87.

Zhang W, Bevins MA, Plantz BA, Smith LA, Meagher MM, Modeling *Pichia pastoris* Growth on Methanol and Optimizing the Production of a Recombinant Protein, the Heavy-Chain Fragment C of *Botulinum* Neurotoxin, Serotype A, Biotechnol. Bioeng. 2000b; 70, 1-8.

Zhang W, Smith LA, Plantz BA, Schlegel VL, Meagher MM, Design of methanol feed control in *Pichia pastoris* fermentations based upon a growth model, Biotechnol Prog 2002; 18, 1392–99.

Zhang WH, Potter KJ, Plantz BA, Schlegel VL, Smith LA, Meagher MM, *Pichia pastoris* fermentation with mixed-feeds of glycerol and methanol: growth kinetics and production improvement, J. Ind. Microbiol. Biotechnol. 2003; 30, 210-5.

Zhang W, Liu CP, Inan M, Meagher MM, Optimization of cell density and dilution rate in *Pichia pastoris* continuous fermentations for production of recombinant proteins, J. Ind. Microbiol. Biotechnol. 2004; 31, 330-4.

Zhang Y, Yang B, In vivo optimizing of intracellular production of heterologous protein in *Pichia pastoris* by fluorescent scanning, Anal. Biochem. 2006; 357, 232-9.

Zhang AL, Zhang TY, Luo JX, Chen SC, Guan WJ, Fu CY, Peng SQ, Li HL, Constitutive expression of human angiostatin in *Pichia pastoris*, J. Ind. Microbiol. Biotechnol. 2007; 34, 117-22.

Zhang Q, Ding F, Xue X, Xu X, Pan W, Changing the N-terminal sequence protects recombinant *Plasmodium falciparum* circumsporozoite protein from degradation in *Pichia pastoris*, Appl. Microbiol. Biotechnol. 2008; 78, 139-45.

Zhang AL, Luo JX, Zhang TY, Pan YW, Tan YH, Fu CY, Tu FZ, Recent advances on the GAP promoter derived expression system of *Pichia pastoris*, Mol. Biol. Rep. 2009; 36, 1611-9.

Zhao W, Wang J, Deng R, Wang X, Scale-up fermentation of recombinant *Candida rugosa* lipase expressed in *Pichia pastoris* using the GAP promoter, J. Ind. Microbiol. Biotechnol. 2008; 35, 189-95.

Zhou XS, Zhang YX, Decrease of proteolytic degradation of recombinant hirudin produced by *Pichia pastoris* by controlling the specific growth rate, Biotechnol. Lett. 2002; 24, 1449–53.

Chapter 4 – Development of recombinant microalgae strains for the production of phytase and xylanase

Authors: Potvin G and Zhang Z

This chapter presents the work performed towards the development of recombinant phytase-producing microalgae to serve as animal feed additives. Given the limitations of a strict scientific paper format, which is not amenable to a full description and discussion of key design considerations and approaches or experimental details involved in this work, some liberties are taken in its presentation. Section 4.2 consists of a detailed description of the approach to gene design and transformation vector assembly; Section 4.3 describes the transformation and strain screening methodology; and Section 4.4 presents the experimental results. For the sake of clarity and organization, relevant materials and methods are presented where they become relevant to the discussion, and not in their own section. Full sequences of all of the gene constructs and oligonucleotides used in this work are found in appendices A.1 and A.2, respectively.

4.1 Abstract

Microalgae offer many advantages as systems for recombinant protein production, as they combine the high growth rate and ease of cultivation of microorganisms with the ability to perform post-transcriptional and post-translational modifications of plant systems. Given that many microalgal species are deemed safe for consumption, and are in fact of considerable nutritional value, they present an interesting potential platform for the production of animal feed additives that would avoid the extensive purification costs associated with conventional yeast or bacterial systems.

This work focused on the development and cultivation of novel recombinant microalgal strains that express and store phytase and xylanase, enzymes that improve nutrient utilization and degradation of non-digestible matter by animals. When fed to crop animals, such microalgal additives would not only provide the nutritional benefits of the enzymes, already recognized and exploited in animal feeds through purified additives, but also provide the value-added nutrition of microalgal biomass, while conferring environmental benefits through decreased nutrient waste. The expression of different gene constructs, under the control of different 5'UTR/promoter

sequences at different genetic loci, and designed using different codon optimization algorithms, was investigated, in conjunction with the enzymatic activities of cell culture lysates.

4.2 Introduction

See Sections 2.2 and 2.3 of this document for a detailed review of expression platforms and their characteristics, as well as an introduction to the properties that make microalgae particularly valuable in such applications.

Although a surge of interest for microalgal biotechnology has led, over the last decade, to the successful transformation of several microalgal species, recombinant protein production in these systems is usually hindered by low expression levels. Although the continued development of genetic engineering tools for microalgae has allowed for the expression of fully functional antibodies (Franklin and Mayfield 2005, Tran et al. 2009), therapeutics (Boehm 2007, Weathers et al. 2010), and bactericides (Li and Tsai 2009) at economically viable levels, success remains, to a large extent, anecdotal, and no wide-ranging system or protocol leading to reliable high-level expression has been established.

The work described here sought to develop novel strains of microalgae that express recombinant enzymes in their chloroplasts, to be used as animal feed additives. Although the main proteins of interest in the current work were phytase and xylanase, easily assayable enzymes commonly used as feed additives, the development of successful expression cassettes would provide opportunities to produce microalgal strains expressing other enzymes of industrial interest, as well as therapeutics or other bioproducts.

Although significant progress has been achieved over the last decade in expressing recombinant proteins in microalgal hosts, yields generally remain quite low. It should also be noted that compared to the other proteins expressed at commercially viable levels (a threshold estimated at approximately 2-3% of total soluble protein (TSP)) in microalgae, the phytase and xylanase genes of interest in the present project are large and relatively complex. Although the production of these recombinant proteins in microalgal chloroplast should be feasible, it was impossible to predict at the outset of the work, given the current limits to our understanding of the biochemical factors and processes that affect these systems, whether the yields achieved would be sufficient to lead to a viable process for the commercial production of the proteins of

interest. Given the targeted end-product, however, a certain leeway is available regarding acceptable expression levels.

4.3 Experimental Approach and Gene Design

4.3.1 Background Information and Rationale

The work presented here sought to modify the chloroplasts of *Chlamydomonas reinhardtii* to express recombinant phytase or xylanase, which could subsequently be used as animal feed additives. Phytase catalyzes the hydrolysis of phytic acid, the main form of phosphorus storage in plants (accounting for 50-80% of the phosphorus), which is indigestible by monogastric crop animals such as swine and poultry (Hussin et al. 2007). It is commonly used as an animal feed additive as it increases the bioavailability of the phosphorus found in plant matter-rich feeds, while avoiding the environmental complications associated with feed supplementation with inorganic phosphorus, which is poorly absorbed by animals, is mostly excreted, and subsequently leaches into soil and surrounding bodies of water causing a slew of environmental problems. Xylanase catalyzes the hydrolysis of xylan, and is thus widely used to bleach textiles and cellulose pulp, in large-scale baking processes, and in bioethanol production (Polizeli et al. 2005). When used as an additive, xylanase increases the nutritional value of feed by releasing sugars that can be metabolized by crop animals from otherwise indigestible xylan. Green microalgae are Generally Regarded as Safe (GRAS), as they produce no endotoxins or other harmful compounds. These microalgae-based feed supplements could therefore be fed directly to animals, would provide the nutritional benefits of the expressed enzymes, as well as the value-added nutritional benefits of the microalgae themselves, which are rich in vitamins, proteins and oils, and would have a positive environmental impact due to reduced phosphorus contamination, all at a lower cost than the traditional yeast- or bacteria-based expression platforms, primarily due to reduced nutrient and purification requirements.

Although, since the beginning of this work, phytases have previously been expressed in *C. reinhardtii* (Yoon et al. 2011) and *Dunaliella tertiolecta* (Georgianna et al. 2013) chloroplasts, our approach is novel, and since our level of understanding of chloroplasts systems as recombinant protein production platforms remains imperfect, the work described here, including the novel sequences, localizations, and codon optimization strategies, will provide insight into

the factors affecting protein yields in these systems, especially when analyzed in conjunction with other reported studies in this field.

For the sake of brevity, the remainder of this document outlines the design considerations and experimental approaches used to develop these microalgal strains using only phytase as an example. Everything discussed, however, also applies to the development of xylanase-producing strains, and relevant details are indicated where necessary.

4.3.2 Host Strain Selection

Despite the recent surge of interest in microalgal technology, to date only a handful of species, out of the thousands identified, are routinely transformed, namely those belonging to the genera *Chlamydomonas, Chlorella, Volvox, Haematococcus* and *Dunaliella* (Griesbeck et al. 2006, Raja et al. 2008, Rosenberg et al. 2008). Of these, *C. reinhardtii* is by far the best characterized and to date has proven to be the most successful as a platform for recombinant protein expression (Lumbreras et al. 1998, Manuell et al. 2007, Mayfield et al. 2007, Eichler-Stahlberg et al. 2009, Surzycki et al. 2009). Since genetic data, transformation protocols and expression strategies are most readily available for *C. reinhardtii*, it was used as the expression host in this work. The wild-type *C. reinhardtii* strain CC-1690 from the Chlamydomonas Culture Center (CCC - Duke University, North Carolina) was used as the host organism for all subsequent genetic engineering work, and wild-type strains CC-620 (mt+) and CC-621 (mt-) were mated to produce autolysin, an enzyme used to remove cell walls from the cultures to be transformed.

4.3.3 Gene Base Sequence Selection

A gene sequence was required as the basis for the design of the customized genes used in this work. Given that phytase was to be expressed in microalgal chloroplasts, in which post-translational modifications including glycosylation are not performed, only prokaryotic phytases were considered. Desirable characteristics for the phytase gene products are high specific activity (for phytate degradation), thermostability, and high activity at low pH (mammalian digestive tracts ~ pH 2). An extensive list of isolated phytases is provided by Rao et al. (2009) .

Most phytases are relatively thermostable and have a wide, if mesophilic, pH operating range. Thermostability is important if the additive is to be included in the feed pellets themselves, as the pelleting process commonly includes steps at temperatures between 65 and 95ºC (Wyss et al.

1999). If the enzyme is not thermostable, it can be added after the pelleting step, but the added processing requirements lead to increased production costs.

Based on these considerations, four bacterial phytases, listed below, were of particular interest for this application. One unit of phytase (U) is defined as the quantity of enzyme required to liberate one μmol of orthophosphate from phytic acid per minute at pH 5.5 and 37°C.

i. Phytase from *Citrobacter braaki* has a high specific activity (1122 U/mg) although only limited information is available regarding its physicochemical properties (Kim et al. 2006).
ii. Phytase isolated from *Citrobacter amalonaticus* CGMCC 1696 has a very high specific activity (3548 U/mg), a pH optimum of 4.5, is resistant to proteases, and has a temperature optimum of 55˚C (Luo et al. 2007).
iii. Phytase isolated from *Yersinia intermedia* has a very high specific activity (3960 U/mg), is highly active at pH 2-6, is resistant to proteases and has a temperature optimum of 55˚C (Huang et al. 2006).
iv. A phytase isolated from *Erwinia carotovora* (var. *carotovora* ATCC10276) is highly active at low temperatures, but very thermolabile, has a pH optimum of 5.5 and a temperature optimum of 40˚C (Huang et al. 2009).

Thermostability, although important in terms of process economics, is of secondary concern if microalgal additives are to be processed separately to the main feed pellets, and therefore not necessarily be subjected to high temperatures. Processing consists of drying and washing of algal aggregates. Although high temperatures would allow for faster and thus less expensive drying, all processing can be performed at lower temperatures. It is also hypothesized that intracellular storage of the enzymes, as opposed to extracellular secretion, may protect products from degradation, thermal or otherwise.

Based on their favourable characteristics, the phytases from *Y. intermedia* and *C. amalonaticus* were used as the base sequences for gene design (GenBank accession numbers DQ986462 and DQ975370 respectively). These phytases have higher activities at lower pH values (pH 1-6), higher specific activities and slightly higher thermostability than their *E. coli* and *Aspergillus niger* counterparts, as well as higher resistance to proteases. The lack of

glycosylation has also been confirmed, making these strains amenable to chloroplast expression. The remainder of this document uses the phytase from *Y. intermedia* to describe all techniques used, which were also used for the *C. amalonaticus* phytase and the selected xylanase genes. Based on similar considerations, thermoacidophilic xylanases from *Alicyclobacillus* (Bai et al. 2010) and *Actinomadura* (Sriyapai et al. 2011) strains (GenBank accession numbers GQ240233 and AB562500 respectively) were selected.

4.3.4 Gene Design – Codon Optimization

It is well established that the genomes of different organisms, and the different genomes of single organisms, employ codon biases as mechanisms for optimizing and regulating protein expression (Gustafsson et al. 2004). As is the case for most heterologous genes, optimizing the codon usage of microalgae-destined transgenes to reflect this bias increases the expression efficiency by increasing the translation rates, and may decrease the transgene's susceptibility to silencing (Heitzer et al. 2007). In prokaryotic genomes, such as those of microalgal chloroplasts, codon bias is the single most important determinant of protein expression (Lithwick and Margalit 2003, Surzycki et al. 2009,), and adjustment of codons in transgenes is necessary for high level (i.e. commercially viable) expression (Franklin et al. 2002, Mayfield et al. 2003).

The Codon Adaptation Index (CAI) is used as a quantitative tool to predict heterologous gene expression levels based on their codon usage. It is defined as the ratio of a given codon to the most expressed codon for a given amino acid. As the chloroplast, mitochondrial and nuclear genomes of microalgae may exhibit different codon biases, as is the case for *C. reinhardtii*, genome-specific CAI objectives should be used for optimal translation effectiveness. A codon bias database for most sequenced species is available online at http://www.kazusa.or.jp/codon/ (Nakamura et al. 2000). Two optimization strategies were investigated.

4.3.4.1 – Optimization Method 1

This method replaces “un-optimized” codons in the base gene sequence with “optimized” ones, defined as codons occurring less than, and more than, 10 times per 1000 codons in the *C. reinhardtii* chloroplast genome, respectively (Franklin et al. 2002, Manuell et al. 2007). To avoid translation rate bottlenecks, when a codon encoding a given amino-acid is found again within 3 amino acids of itself, and if a codon of similar frequency (other codon with frequency over

10/1000) is available, a substitution is made, as shown in green in the sequence below. This small refinement of the general optimization strategy aims to avoid lags in translation by reducing competition for the same tRNAs. The full sequence with corresponding amino acids is shown in Figure 4.1.

```
atg aca ata aca gta gat agt ctg cga tta tcc gta ctg acc ttg ata ctc aat agt
atg aca att act gta gat agt tta cgt tta tca gta tta aca tta att tta aat agt
 M   T   I   T   V   D   S   L   R   L   S   V   L   T   L   I   L   N   S

tat gcg att agt gcc gcg ccg gtt gcc ata caa ccc acg ggc tat aca ttg gag cga
tat gct att agt gca gct cca gtt gca att caa cca aca ggt tat act tta gaa cgt
 Y   A   I   S   A   A   P   V   A   I   Q   P   T   G   Y   T   L   E   R

gtg gtt att ttg agc cgc cat ggt gtt cgc tcg cca acc aaa caa aca cag tta atg
gtt gta att tta tca cgt cat ggt gtt cgt tca cca aca aaa caa act caa tta atg
 V   V   I   L   S   R   H   G   V   R   S   P   T   K   Q   T   Q   L   M

aat gat gtt acc cct gac acg tgg ccg caa tgg ccg gtc gcc gca gga tac tta acc
aat gat gtt aca cct gat act tgg cca caa tgg cct gtt gct gca ggt tac tta aca
 N   D   V   T   P   D   T   W   P   Q   W   P   V   A   A   G   Y   L   T

ccc cga ggt gca caa tta gtg aca ttg atg ggc gga ttc tat ggt gat tac ttc cgt
cca cgt ggt gca caa tta gtt aca tta atg ggt ggt ttc tat ggt gat tat ttc cgt
 P   R   G   A   Q   L   V   T   L   M   G   G   F   Y   G   D   Y   F   R

agc caa ggg tta ctc gca gca ggg tgc cca act gac gcg gtt att tat gct cag gcc
tca caa ggt tta tta gct gca ggt tgt cca act gat gct gtt att tat gct caa gca
 S   Q   G   L   L   A   A   G   C   P   T   D   A   V   I   Y   A   Q   A

gat gtt gat caa cga acg cgt tta acg ggg cag gca ttc ctt gat gga ata gca ccg
gat gtt gat caa cgt aca cgt tta act ggt caa gca ttc ctt gat ggt att gca cca
 D   V   D   Q   R   T   R   L   T   G   Q   A   F   L   D   G   I   A   P

ggg tgt gga ctg aaa gta cat tat cag gct gat ttg aaa aaa gtg gat ccg ctg ttt
ggt tgt ggt tta aaa gta cat tat caa gct gat tta aaa aaa gtt gat cca tta ttt
 G   C   G   L   K   V   H   Y   Q   A   D   L   K   K   V   D   P   L   F

cat ccc gtc gac gcg ggg gtg tgt aag tta gat tcg aca caa acc cat aag gct gtt
cat cca gtt gat gct ggt gtt tgt aaa tta gat tca aca caa act cat aaa gct gtt
 H   P   V   D   A   G   V   C   K   L   D   S   T   Q   T   H   K   A   V

gag gag cga cta ggt ggg cca tta agt gaa ctg agc aaa cgc tat gct aag ccc ttt
gaa gaa cgt tta ggt ggt cca tta agt gaa tta tca aaa cgt tat gct aaa cca ttt
 E   E   R   L   G   G   P   L   S   E   L   S   K   R   Y   A   K   P   F

gcc cag atg ggt gag att ctg aat ttt gcg gca tct cct tac tgt aaa tca ctg caa
gct caa atg ggt gaa att tta aat ttt gct gca tct cct tac tgt aaa tca tta caa
 A   Q   M   G   E   I   L   N   F   A   A   S   P   Y   C   K   S   L   Q

cag caa ggg aaa acc tgt gat ttt gcc aac ttt gca gcg aat aag atc acg gtg aac
caa caa ggt aaa aca tgt gat ttt gct aac ttc gca gct aat aaa att aca gtt aac
 Q   Q   G   K   T   C   D   F   A   N   F   A   A   N   K   I   T   V   N

aag ccg ggg aca aaa gtc tcg ctc agc gga cca ctg gca ctg tca tca acc tta ggt
aaa cca ggt aca aaa gtt tca tta tct ggt cca tta gca tta tca tct aca tta ggt
 K   P   G   T   K   V   S   L   S   G   P   L   A   L   S   S   T   L   G

gag atc ttt ttg cta caa aat tca caa gcg atg cct gat gtt gcc tgg cat cgg tta
gaa att ttt tta tta caa aat tca caa gct atg cct gat gtt gct tgg cat cgt tta
 E   I   F   L   L   Q   N   S   Q   A   M   P   D   V   A   W   H   R   L

acg gga gaa gat aat tgg atc tcg tta tta tcg ttg cac aat gcg caa ttt gat tta
aca ggt gaa gat aat tgg att tca tta tta tca tta cat aat gct caa ttt gat tta
 T   G   E   D   N   W   I   S   L   L   S   L   H   N   A   Q   F   D   L
```

```
atg gca aaa aca cct tat atc gct cgt cat aag ggc aca ccg ttg ctg caa cag atc
atg gca aaa aca cct tat att gct cgt cat aaa ggt aca cca tta tta caa caa att
 M   A   K   T   P   Y   I   A   R   H   K   G   T   P   L   L   Q   Q   I

gag act gcc ctc gtc ctt cag cgt gat gct cag ggg caa aca ttg cca tta tca cct
gaa act gct tta gtt ctt caa cgt gat gct caa ggt caa aca tta cca tta tca cct
 E   T   A   L   V   L   Q   R   D   A   Q   G   Q   T   L   P   L   S   P

caa acc aaa att ctg ttc ctc ggg gga cat gat aca aac atc gcc aat att gct gga
caa aca aaa att tta ttc tta ggt ggt cat gat aca aac att gct aat att gca ggt
 Q   T   K   I   L   F   L   G   G   H   D   T   N   I   A   N   I   A   G

atg ttg ggg gct aac tgg caa tta cca cag cag ccc gat aat acc cca cct ggg ggg
atg tta ggt gct aac tgg caa tta cca caa caa cct gat aat aca cca cct ggt ggt
 M   L   G   A   N   W   Q   L   P   Q   Q   P   D   N   T   P   P   G   G

gga ttg gtc ttc gag cta tgg caa aac cca gat aat cat caa cgt tat gtc gcg gtg
ggt tta gtt ttc gaa tta tgg caa aac cca gat aat cat caa cgt tat gta gct gtt
 G   L   V   F   E   L   W   Q   N   P   D   N   H   Q   R   Y   V   A   V

aaa atg ttc tat caa aca atg ggc caa ttg cga aat gct gag aaa cta gac ctg aaa
aaa atg ttc tat caa aca atg ggt caa tta cgt aat gct gaa aaa tta gat tta aaa
 K   M   F   Y   Q   T   M   G   Q   L   R   N   A   E   K   L   D   L   K

aac aat ccg gct ggt agg gtc cct gtt gca ata gac ggt tgt gaa aat agt ggt gat
aac aat cca gct ggt cgt gtt cct gta gca att gat ggt tgt gaa aat agt ggt gat
 N   N   P   A   G   R   V   P   V   A   I   D   G   C   E   N   S   G   D

gac aaa ctt tgt cag ctt gat acc ttc caa aag aaa gta gct cag gcg att gaa cct
gat aaa ctt tgt caa ctt gat aca ttc caa aaa aaa gta gca caa gct att gaa cct
 D   K   L   C   Q   L   D   T   F   Q   K   K   V   A   Q   A   I   E   P

gct tgc cat att taa
gct tgt cat att taa
 A   C   H   I   *
```

Figure 4.1. *Sequence of Y. intermedia phytase optimized with the first optimization strategy. Basic DNA coding sequence (5'-3') of Y. intermedia phytase in black, 'optimized' codons in red, codons remaining unchanged from the base sequence in blue, and substituted abundant codons in green. An analysis of sequence properties is found in Table 4.1.*

4.3.4.2 – Optimization Method 2

In prokaryotic genomes, without the assistance of chaperone molecules or post-translational modifications, proper folding and assembly of proteins is highly affected by translational rates (Rasala et al. 2010). It is well established that preferred codons generally correspond to the most abundant tRNAs (Ikemura 1985). Ran and Higgs (2010) propose that codon biases coevolve with tRNA abundance to maximize protein translation rates, as opposed to accuracy, in highly expressed genes. It is possible, however, that coevolution of codon usage in genes necessary for cellular function, whether or not they are considered to have a high expression, also occurs to ensure maximal *functional* protein yields. Translation rates too high or too low, which may be caused by an overabundance or deficiency in common codons (directly related to the abundance of tRNAs), may derail proper folding or formation of disulfide bonds, or prematurely terminate translation. Given the target application, only yields of functional proteins were of interest. As

such, an alternative coding algorithm was considered, in which not only the codon bias of the host organism is considered, but also the relative abundance of codons used by the source organism. The rationale for this approach is based on the hypothesis that variations in translation rates of the phytase gene in the source organism (*Y. intermedia*) are affected by selection, and that rate reductions may play a role in *functional* protein yields. Given that tRNAs with a given anti-codon can bind to multiple codons (and that codons can be translated by multiple tRNAs), it is conceivable that this approach may also lead to lower translational accuracy by allowing incorrect binding, which must be determined experimentally.

The modified algorithm considers the codon bias of both *C. reinhardtii* chloroplasts and *Y. intermedia*, both compiled by Nakamura (Nakamura et al. 2000). 'Optimized' codons in *C. reinhardtii* using this method are those found at the frequencies closest to the frequency of that same codon in *Y. intermedia* (per 1000 codons). The full optimized sequence is shown in Figure 4.2. The algorithm applied to the base sequence is as follows:

i. For each codon in the base sequence, the corresponding amino acid is identified
ii. The frequency of occurrence in *Y. intermedia* (source organism) is verified
iii. The codon coding for same amino acid in *C. reinhardtii* with frequency closest to the one in *Y. intermedia* is identified
iv. Closest frequency codons for highly abundant ones (over 10/1000 codons) is the closest numerical frequency over 10/1000 codons. Closest frequency codons for low abundance ones (lower than 10/1000 codons) is the closest numerical frequency lower than 10/1000. Assumes abundant codons (above 10/1000) are not limited by the tRNA pool, but similar frequencies are selected to replicate translational rates assuming a tRNA pool mirroring codon use in both organisms.

```
atg aca ata aca gta gat agt ctg cga tta tcc gta ctg acc ttg ata ctc aat agt
atg act atc act gta gat agt tta cgc ctt agc gta tta aca ctt atc cta aat agt
 M   T   I   T   V   D   S   L   R   L   S   V   L   T   L   I   L   N   S

tat gcg att agt gcc gcg ccg gtt gcc ata caa ccc acg ggc tat aca ttg gag cga
tat gca att agt gct gca cct gtt gct atc caa ccc acc ggt tat act ctt gaa cgc
 Y   A   I   S   A   A   P   V   A   I   Q   P   T   G   Y   T   L   E   R

gtg gtt att ttg agc cgc cat ggt gtt cgc tcg cca acc aaa caa aca cag tta atg
gtt gtt att ctt agt cgt cat ggt gtt cgt agc cct aca aaa caa act caa ctt atg
 V   V   I   L   S   R   H   G   V   R   S   P   T   K   Q   T   Q   L   M

aat gat gtt acc cct gac acg tgg ccg caa tgg ccg gtc gcc gca gga tac tta acc
aat gat gtt aca cct gat acc tgg cct caa tgg cct gta gct gct gga tac ctt aca
 N   D   V   T   P   D   T   W   P   Q   W   P   V   A   A   G   Y   L   T
```

```
ccc cga ggt gca caa tta gtg aca ttg atg ggc gga ttc tat ggt gat tac ttc cgt
ccc cgc ggt gct caa ctt gtt act ctt atg ggt gga ttc tat ggt gat tac ttc cgt
 P   R   G   A   Q   L   V   T   L   M   G   G   F   Y   G   D   Y   F   R

agc caa ggg tta ctc gca gca ggg tgc cca act gac gcg gtt att tat gct cag gcc
agt caa ggt ctt cta gct gct ggt tgc cct act gat gca gtt att tat gca caa gct
 S   Q   G   L   L   A   A   G   C   P   T   D   A   V   I   Y   A   Q   A

gat gtt gat caa cga acg cgt tta acg ggg cag gca ttc ctt gat gga ata gca ccg
gat gtt gat caa cgc acc cgt ctt acc ggt caa gct ttc cta gat gga atc gct cct
 D   V   D   Q   R   T   R   L   T   G   Q   A   F   L   D   G   I   A   P

ggg tgt gga ctg aaa gta cat tat cag gct gat ttg aaa aaa gtg gat ccg ctg ttt
ggt tgt gga tta aaa gta cat tat caa gca gat ctt aaa aaa gtt gat cct tta ttt
 G   C   G   L   K   V   H   Y   Q   A   D   L   K   K   V   D   P   L   F

cat ccc gtc gac gcg ggg gtg tgt aag tta gat tcg aca caa acc cat aag gct gtt
cat ccc gta gat gca ggt gtt tgt aag ctt gat agc act caa aca cat aag gca gtt
 H   P   V   D   A   G   V   C   K   L   D   S   T   Q   T   H   K   A   V

gag gag cga cta ggt ggg cca tta agt gaa ctg agc aaa cgc tat gct aag ccc ttt
gaa gaa cgc ttg ggt ggt cct ctt agt gaa tta agt aaa cgt tat gca aag ccc ttt
 E   E   R   L   G   G   P   L   S   E   L   S   K   R   Y   A   K   P   F

gcc cag atg ggt gag att ctg aat ttt gcg gca tct cct tac tgt aaa tca ctg caa
gct caa atg ggt gaa att tta aat ttt gca gct tct cct tac tgt aaa tca tta caa
 A   Q   M   G   E   I   L   N   F   A   A   S   P   Y   C   K   S   L   Q

cag caa ggg aaa acc tgt gat ttt gcc aac ttt gca gcg aat aag atc acg gtg aac
caa caa ggt aaa aca tgt gat ttt gct aac ttt gct gca aat aag att acc gtt aac
 Q   Q   G   K   T   C   D   F   A   N   F   A   A   N   K   I   T   V   N

aag ccg ggg aca aaa gtc tcg ctc agc gga cca ctg gca ctg tca tca acc tta ggt
aag cct ggt act aaa gta agc cta agt gga cct tta gct tta tca tca aca ctt ggt
 K   P   G   T   K   V   S   L   S   G   P   L   A   L   S   S   T   L   G

gag atc ttt ttg cta caa aat tca caa gcg atg cct gat gtt gcc tgg cat cgg tta
gaa att ttt ctt ttg caa aat tca caa gca atg cct gat gtt gct tgg cat aga ctt
 E   I   F   L   L   Q   N   S   Q   A   M   P   D   V   A   W   H   R   L

acg gga gaa gat aat tgg atc tcg tta tta tcg ttg cac aat gcg caa ttt gat tta
acc gga gaa gat aat tgg att agc ctt ctt agc ctt cac aat gca caa ttt gat ctt
 T   G   E   D   N   W   I   S   L   L   S   L   H   N   A   Q   F   D   L

atg gca aaa aca cct tat atc gct cgt cat aag ggc aca ccg ttg ctg caa cag atc
atg gct aaa act cct tat att gca cgt cat aag ggt act cct ctt tta caa caa att
 M   A   K   T   P   Y   I   A   R   H   K   G   T   P   L   L   Q   Q   I

gag act gcc ctc gtc ctt cag cgt gat gct cag ggg caa aca ttg cca tta tca cct
gaa act gct cta gta cta caa cgt gat gca caa ggt caa act ctt cct ctt tca cct
 E   T   A   L   V   L   Q   R   D   A   Q   G   Q   T   L   P   L   S   P

caa acc aaa att ctg ttc ctc ggg gga cat gat aca aac atc gcc aat att gct gga
caa aca aaa att tta ttc cta ggt gga cat gat act aac att gct aat att gca gga
 Q   T   K   I   L   F   L   G   G   H   D   T   N   I   A   N   I   A   G

atg ttg ggg gct aac tgg caa tta cca cag cag ccc gat aat acc cca cct ggg ggg
atg ctt ggt gca aac tgg caa ctt cct caa caa ccc gat aat aca cct cct ggt ggt
 M   L   G   A   N   W   Q   L   P   Q   Q   P   D   N   T   P   P   G   G

gga ttg gtc ttc gag cta tgg caa aac cca gat aat cat caa cgt tat gtc gcg gtg
gga ctt gta ttc gaa ttg tgg caa aac cct gat aat cat caa cgt tat gta gca gtt
 G   L   V   F   E   L   W   Q   N   P   D   N   H   Q   R   Y   V   A   V

aaa atg ttc tat caa aca atg ggc caa ttg cga aat gct gag aaa cta gac ctg aaa
aaa atg ttc tat caa act atg ggt caa ctt cgc aat gca gaa aaa ttg gat tta aaa
 K   M   F   Y   Q   T   M   G   Q   L   R   N   A   E   K   L   D   L   K

aac aat ccg gct ggt agg gtc cct gtt gca ata gac ggt tgt gaa aat agt ggt gat
aac aat cct gca ggt agg gta cct gtt gct atc gat ggt tgt gaa aat agt ggt gat
 N   N   P   A   G   R   V   P   V   A   I   D   G   C   E   N   S   G   D
```

```
gac aaa ctt tgt cag ctt gat acc ttc caa aag aaa gta gct cag gcg att gaa cct
gat aaa cta tgt caa cta gat aca ttc caa aag aaa gta gca caa gca att gaa cct
 D   K   L   C   Q   L   D   T   F   Q   K   K   V   A   Q   A   I   E   P

gct tgc cat att taa
gca tgc cat att taa
 A   C   H   I   *
```

Figure 4.2. *Sequence of Y. intermedia phytase optimized with the second optimization algorithm. The base coding sequence (5'-3') of Y. intermedia phytase in black, 'optimized' codons in red, codons remaining unchanged from the base sequence in blue, and related amino acids. An analysis of sequence properties is found in Table 4.1*

This modified algorithm was interesting to investigate as it could either increase the functional protein yield through improving the folding and assembly of recombinant products, if this factor is significant for phytase assembly, or decrease overall protein yields due to lower translational rates (compared to substitution of most abundant codons where discrepancies in frequency exist).

Table 4.1. Analysis of gene composition and sequence homology among the base phytase sequence from *Y. intermedia* and the sequences optimized using both optimization algorithms. The CAI for all sequences is based on the *C. reinhardtii* chloroplast genome codon tables.

	***C. reinhardtii* chloroplast genome**	**Phytase Base Sequence (*Y. intermedia*)**	**Optimized with Method 1**	**Optimized with Method 2**
GC content (%)	33.7	----	34.0	39.5
1st position GC (%)	44.4	----	52.7	59.5
2nd position GC (%)	37.5	----	42.3	42.3
3rd position GC (%)	19.4	----	7.0	16.7
CAI before optimization	----	0.388	----	----
CAI after optimization	----	----	0.937	0.713
Sequence homology with base sequence (%)	----	100	82.96	78.81
Sequence homology with first optimization (%)	----	82.96	100	86.1

4.3.5 Promoter/5'UTR/3'UTR Selection

Several chloroplast-specific promoters and 5'UTR regions have been tested for high-yield recombinant protein expression (*rbcL* (Goldschmidt-Clermont 1991, Kato et al. 2007), *atpA* (Goldschmidt-Clermont 1991, Surzycki et al. 2009), *and psbD* (Surzycki et al. 2009)). The large majority of resulting yields are <1% TSP (Mayfield et al. 2007). Higher (i.e. potentially economically viable) levels have been achieved, however, for heterologous proteins under the control of the *psbA* promoter in *psbA*-deficient strains. (10% TSP (Manuell et al. 2007, 20.9% TCP (Surzycki et al. 2009), variable (Rasala et al. 2010)). Good expression has also been

achieved with *psbD* and *atpA* promoters. The *psbA* and *psbD* promoters were selected for this work, along with their corresponding 5' UTR. The 3' UTR sequence selected, as long as one is present, has no significant effect on mRNA or protein yields (Barnes et al. 2005), and was thus selected based on available vectors to minimize the required manipulations or synthesis of DNA constructs. The *psbA* promoter and 5'/3' UTRs were obtained from plasmid p-546 from the CCC, in which sequences could be inserted in-frame between NcoI and XbaI restriction sites for expression. The expression vector containing the *psbD* promoter and corresponding UTRs was designed in-house, as described in detail in section 4.3.8.

4.3.6 Selectable Markers

As the objective of the present work is the development of products destined for animal consumption, selectable markers had to be carefully chosen. For instance, markers conferring a resistance to antibiotics by expressing recombinant products are undesirable in the final product to avoid any risk, however small, of horizontal gene transfer or environmental contamination. Given the current social and political climate in regards to genetically modified organisms, the product should contain the lowest number of deviations possible from its "natural" state to minimize political resistance and the public's opprobrium.

Two general selection strategies are available, namely the rescue of a knock-out strain by complementation with an active gene, or co-transformation with a gene cassette allowing for chemical (e.g. antibiotic resistance), or visual selection of successful transformants.

The first alternative involves the transformation of algal strains in which a given selectable gene is knocked-out, with a DNA construct, generally a wild-type gene, which restores its proper function. The most common genes targeted for knock-out/rescue consists of genes for metabolic enzymes involved in the degradation of nutrients (e.g. *NIT1* which allows the use of nitrate as a nitrogen source, selectable by media composition (Galban et al. 1996)) or genes involved in photosynthesis that if knocked out do not impair non-photosynthetic growth (e.g. photosystem I (*psaX*) or II (*psbX*) genes).

The second alternative involves the expression of a foreign DNA construct coding for a protein that allows for the selection of transformed strains. The most common genes used for this selection strategy provide antibiotic resistance, but visual markers such as GFP or luciferase have also been useful. In chloroplasts, the most commonly used markers of this type include the *aphA*-

6 gene, which encodes an aminoglycoside phophotransferase enzyme from *Acinetobacter baumannii* which confers resistance to kanamycin and related antibiotics (Bateman and Purton 2000), or the *aada* gene, which encodes an aminoglycoside adenyl transferase, conferring resistance to streptomycin and spectinomycin (Goldschmidt-Clermont 1991). The expression of heterologous products is not necessary, and inducing an A1123G point mutation in the endogenous 16S rRNA gene for example confers high level spectinomycin resistance without interfering with normal cellular machinery or expressing foreign proteins (Harris et al. 1989), which has the noted advantage of avoiding increases in the metabolic load caused by the expression of selection markers.

Either strategy can achieve selection by cotransformation of a target gene and the selection marker on the same construct (one vector) or with simultaneous cotransformation on distinct vectors. A high proportion of target integration by homologous recombination in the chloroplast genome is observed when the selectable marker is integrated, even if on a separate vector. The frequency of integration of both constructs is, however, lower if the latter strategy is employed, and a given proportion of successful transformants will not integrate the target gene. For larger constructs integrated at different locations, such as is the case here, distinct vectors were used. Plasmid p-228 from the CCC can induce the aforementioned point mutation in the 16S rRNA gene, and was thus used as the selectable marker for cotransformation with the vectors containing the expression cassettes.

4.3.7 Restriction Sites/Primer Annealing Sites/Gene Sequence Refinement

The following refinements were made to the optimized gene sequences to facilitate their manipulation and detection. The complete gene sequences are found in appendix A.1.

4.3.7.1 - Restriction Sites

Restriction sites for NcoI and XbaI (CCATGG, TCTAGA respectively) were added to 5’ and 3’ ends of the synthetic genes respectively to allow for their insertion into expression vectors while maintaining the correct reading frame. The designed genes do not contain any internal cleavage sites for these enzymes, as confirmed by the NEB cutter tool (http://tools.neb.com/NEBcutter2/).

4.3.7.2 - PCR primer binding sites for cloning and screening

'Artificial' PCR primer annealing sites (for Fphy and Rphy primers) were added at both ends of synthetic genes to allow for the amplification of the entire sequences to facilitate their manipulation. Given that primers designed directly from the phytase sequences have low Tm values, high self-complementarity and high end self-complementarity, and may lead to the mutational removal of restriction sites, which were critical for subsequent manipulations, through improper annealing, sequences allowing for the use of optimal primers were inserted at the 5' and 3' ends of the genes directly upstream and downstream of restriction sites. These primer annealing sites were randomly generated, have similar high melting temperatures, and low self-complementarity. Using these 'artificial' annealing sites in all constructs had the added benefit of allowing for the use of a single set of primers to amplify all sequences of interest in full. These primer sequences are not found in the *C. reinhardtii* plastid genome (based on NCBI database). Internal primers amplifying partial gene fragments were also designed for all constructs (using Primer Premier 5 software) and used to confirm successful sub-cloning and gene integration. The full list of primers is included in appendix A.2.

4.3.7.3 - Protein detection and quantification

Western blotting was used to quantify protein expression. As no antibodies specific to the proteins of interest were commercially available, a synthetic polyhistidine tag (his-tag) was added at the C-terminal end of all gene constructs. The common his-tag epitope generally has negligible effect on protein structure (Carson et al. 2007), but may in some cases affect protein activity (Freydank et al. 2008). The associated DNA sequence optimized for six histidine residues in *C. reinhardtii* chloroplasts became: 5'-CATCATCATCATCATCAT-3'. A protease cleavage site was included between the coding region of the genes and the C-terminal his-tag for subsequent tag removal and to determine whether it had any effects on protein activity or structure. Commercial kits are available for a wide range of cleavage/purification protocols. Factor Xa, a commonly used serine endopeptidase was used as a Factor Xa/his-tag removal kit was available from Qiagen. Factor Xa cleaves the C-terminal site of the peptide sequence IEGR (Ile-Glu-Gly-Arg) (Kim et al. 2007). The codon optimized sequence for its expression in *C. reinhardtii* chloroplast is 5'- ATTGAAGGTCGT -3' which was inserted at the N-terminal end of

the his-tag. See the full list of constructs, including the modifications outlined here, in appendix A.1.

4.3.8 Design of the ppsbD vector

Transforming algal cells with the *psbA* promoter-regulated constructs in plasmid p-546 knocks out the endogenous *psbA* gene by homologous recombination. Protein D1, the product of the *psbA* gene, plays an integral role in the photosystem II chain, and its removal therefore causes the loss of photosynthesis (Stern and Harris 2009). As photosynthetic growth is one of the main economic advantages of producing recombinant proteins in microalgae, its restoration is desirable for subsequent cultivation, which is achieved by restoring the expression of D1. D1 acts as a negative feedback inhibitor of the light-activated *psbA* promoter, likely to prevent its over-accumulation during temporary light exposure. Reinsertion of a cassette expressing D1 under the control of another copy of the *psbA* promoter leads to significantly lower expression of our recombinant protein of interest, but its expression under the control of another highly-active promoter, such as the *psbD* promoter, leads to the restoration of photosynthesis with only minor reductions in product levels associated with D1-mediated repression of the *psbA* promoter (Manuell et al. 2007). Since the *psbD* promoter is highly active and has shown to be effective at recombinant protein expression itself, the same construct can be used to express recombinant phytase without the photosynthetic-eliminating effects of the p-546 constructs. The full sequence for the *psbD*-bearing expression vector, designed in-house, is found in Figure 4.3.

The targeted insertion site selected was the region between the *psbA* and *S2* (Ser tRNA) genes within the inverted repeat. Given that the flanking genes are very well expressed, as they produce elements necessary for photosynthesis and protein translation, this region of the genome should not have any epigenetic effects leading to the repression of expression. A 750bp sequence corresponding to base pairs 56625-57375 (GenBank number for full chloroplast genome: BK000554.2) was used as evenly-divided regions flanking the inserted gene constructs.

For insertion in a vector and subsequent manipulations, restriction sites for SacI (GAGCTC) and KpnI (GGTACC) were placed at the 5’ and 3’ ends of the flanking sequences. These restriction enzymes were chosen as they were compatible with the pBSSK+ plasmid used as the vector backbone, and did not cleave any of the gene constructs of interest. The same primer

binding sites for Fphy and Rphy primers were included for the amplification of the entire sequence.

gaccgacatcatcagtttgaccgagctcataaagctttgtaaatactaaatgaataaactcacttctttttttatta
ccattagataataaaagatacaagcccgaagggcttgggggaattcgcacaaaaagcttacatcgaagtcacttaaa
cacatatcacatatacgctcctaaatctaagggtttataagcgtatatgttatggctaaacggtgtgtagctagcat
cagtgttgtataagttctacttaaatatccacactaaggaatgcgtgtcggtgcagaaagcatcgctactggtaaat
tattggatttcaccacagcttgggagggagctacaaatccaacaacaaagacaaaaatcccaataaacaacactgtc
acttgtgttacaacaaaacattttaaattttaaatttttttcctaattatatatttttacttgcaaaatttataaaaat
tttatgcatttttatatcataataataaaaccttttattcatggtttataatataataattgtgatgactatgcacaa
agcagttctagtcccatatatataactatatataacccgtttaaagatttatttaaaaatatgtgtgtaaaaaatgc
ttatttttaatttttatttttatataagttataatattaaatacacaatgattaaaattaaataataataaatttaacg
taacgatgagttgtttttttattttggagatacacgcaccatggtctagaattttttatttttcatgatgtttatgtg
aatagcataaacatcgtttttattttttatggtgtttaggttaaatacctaaacatcattttacatttttaaaatta
agttctaaagttatcttttgtttaaatttgcctgtgctttataaattacgatgtgccagaaaaataaaatcttagct
ttttattatagaatttatctttatgtattatatttttataagtaataaaagaaatagtaacatactaaagattaaagc
tgttcgaagaaaagcttgatctagcattaaattaattcgatctataagacaagcaaaactttcttcttcgttgacta
aaaccaaccattgtgtgaccactttattgtcaacaaaacaatttttttcatgaaggatttcttgtaacaaacaaaag
aaatcctttaacaaaataggatttaaatcagccatacttaacaccaaaagatctaagtcttggaaaagcattaaaag
tgctcctttagtattattaattggcaatttttgaaatgctttttttatagtaccataaagtattttcttggcttctt
ctgataaactataagataagcttaaattaaccaagaaagctagtcctagttgtggaggaggtaccgtagttcgtatc
cgttaggctgg

Figure 4.3. *Full sequence (5'-3') of the ppsbD expression cassette. Construct includes flanking regions for homologous recombination (56625-57000 and 57001-57375 – black) restriction sites for sequence manipulation (SacI, NcoI, XbaI and KpnI - blue), the ppsbD promoter and 5'UTR (green) and rbcL 3'UTR (orange). Full sequence is flanked by primer binding sites for Fphy and Rphy primers (red).*

The sequences for the *psbD* promoter and 5'UTR and the 3'UTR for the *rbcL* gene were taken from the full chloroplast genome listed on the NCBI database. The *rbcL* 3'UTR sequence, as opposed to the 3'UTR of *psbD*, was used to avoid homologous recombination at the endogenous *psbD* gene site, which should have no impact on heterologous protein expression (barnes et al. 2005). The 5' and 3' UTRs are shown in Figure 4.3. The same NcoI and XbaI restriction sites are used for the insertion of heterologous sequences in-frame. The full construct, ligated into the pBSSK+ vector, is referred to as the ppsbD vector.

4.3.9 Final Gene Constructs

All transformation vectors, consisting of the p-546 or ppsbD plasmids containing in-frame sequences for the genes of interest, were assembled using standard molecular biology techniques as detailed in Sambrook and Russel (2001) . It should be noted that two types of constructs were assembled based on the p-546 plasmid, the first using the recommended NcoI and XbaI restriction sites for gene insertion, and the second using the NdeI restriction site instead of the

NcoI site (with 5' restriction sites in the genes of interest changed using the appropriate forward overhanging PCR primers, see appendix A.2), in order to remove a duplicate initiation codon which may affect recombinant product yields. The generalized construct structures are depicted in Figure 4.4.

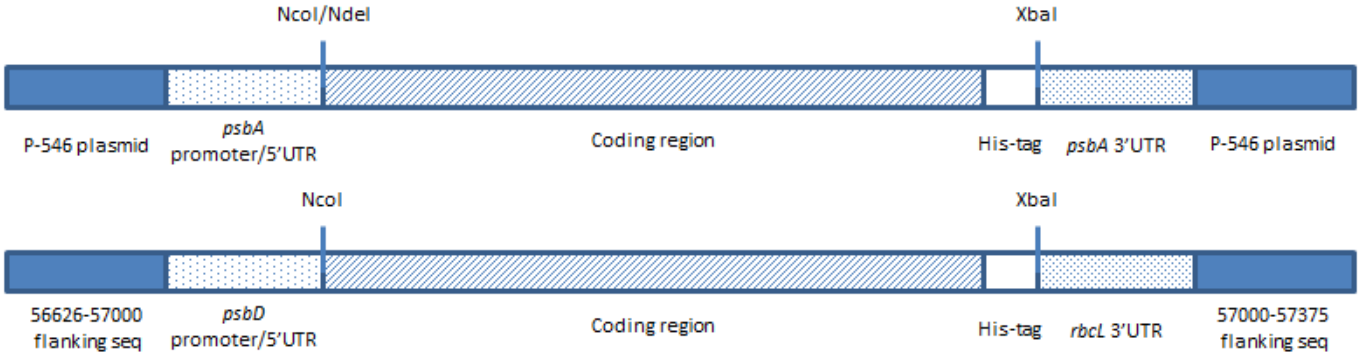

Figure 4.4. *Expression cassettes and flanking regions for the psbA (top) and psbD (bottom) promoter-mediated constructs. Coding regions correspond to any of the optimized gene sequences of interest in this work.*

4.4 Transformation and Transformant Screening

4.4.1 Transformation

Chloroplast transformation occurs by homologous recombination (see Figure 4.5) which allows for the targeted insertion of gene sequences at specific locations in the genome, as determined by flanking homologous sequences. Several methods are available to deliver DNA to *C. reinhardtii* chloroplasts. Successful transformations have been reported using glass bead puncturing, electroporation, microinjection, particle bombardment and *Agrobacterium*-mediated transformation. A more detailed discussion of these methods is found in Chapter 2 of this document.

Due to its high chloroplast transformation efficiency, simple operation and to the reproducibility of results, the particle bombardment method was used for all transformations. A Bio-Rad PDS-1000/He biolistic particle bombardment device was generously provided by Dr. Illimar Altosaar from the department of Biochemistry, Microbiology and Immunology at the University of Ottawa, and S550D gold nanoparticles, optimized for *C. reinhardtii* transformation, have been obtained from Seashell Technologies Ltd (La Jolla, California). Although the removal of cell walls or the use of a cell wall-deficient strain is not strictly

necessary for this method to be effective, it does increase transformation efficiency. Cells were pre-treated with autolysin produced by the mating of CC-620 and CC-621 strains of *C. reinhardtii* for the removal of cell walls prior to transformation.

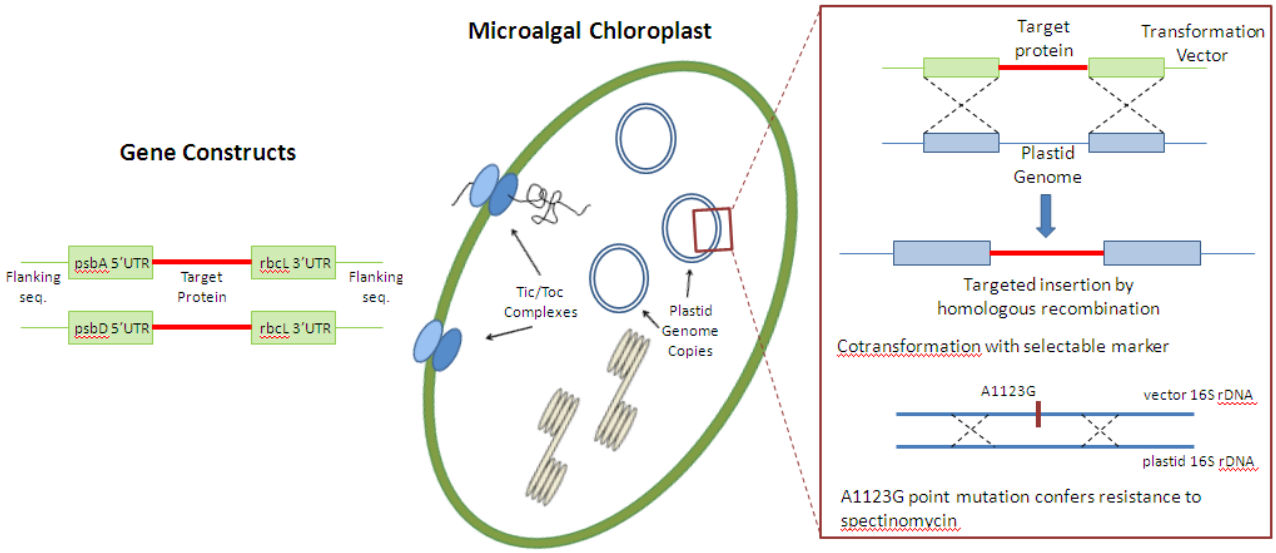

Figure 4.5. *Homologous recombination-based insertion of heterologous gene constructs in microalgal chloroplasts and representation of the co-transformation with p-228 approach used in this work.*

Prior to transformation, microalgal cells were cultivated in Tris-Acetate-Phosphate (TAP) medium (2.42 g/l Tris, 0.375 g/l NH_4Cl, 0.1 g/l $MgSO_4 \cdot 7H_2O$, 0.05 g/l $CaCl_2 \cdot 2H_2O$, 0.11 g/l K_2HPO_4, 0.055 g/l KH_2PO_4, 1 ml/l Hutner trace elements solution, 1 ml/l acetic acid) supplemented with 5-fluorodeoxyuridine, which reduces the copy number of plastid genomes, facilitating the isolation of homoplasmic strains (Wurtz et al. 1977). Cells are grown until the mid-log phase (~3 x 10^6 cells/ml) prior to transformation.

Particle bombardment transformations were performed using 1350 psi rupture disks, a 9cm travel distance, and a 28mmHg vacuum according to manufacturer's instructions. The vectors containing the expression cassettes, and the p-228 selection plasmid were deposited on the nanoparticles in a 1:1 ratio. As previously described, the p-228 plasmid contains a copy of the chloroplast 16S rRNA gene with an A1123G point mutation that confers high level spectinomycin resistance. 100 µl of CC-1690 grown in 5'fluorodeoxyuridine-supplemented TAP was plated on TAP plates (15 g/l agar), were transformed, incubated for 24 hours in low

illumination, and the cells were transferred to TAP plates supplemented with 150 µg/ml spectinomycin and incubated for 10 days in low illumination.

4.4.2 Transformant Screening

Although the transformants are guaranteed to have integrated the p-228 plasmid point mutation conferring spectinomycin resistance, they had to be screened to determine whether the gene construct was also inserted. Transformant colonies were transferred to 2 ml of TAP supplemented with spectinomycin and incubated for 2 days. Cultures were spun down, and total DNA was extracted using a standard chloroform/phenol extraction protocol (Sambrook and Russel 2001). The DNA of each transformant was then screened by PCR using primers with annealing sites contained within the foreign sequences of interest. Transformants having inserted the genes of interest were transferred to 100 ml of fresh TAP medium, incubated for 1 week at low illumination, and repeatedly subcultured until homoplasmicity was achieved. Homoplasmicity was determined by PCR using primers specific to the regions flanking the gene insertion site in the chloroplast genome, and separation by electrophoresis. Homoplasmicity is confirmed by only observing larger bands (corresponding to the wild-type region between the annealing sites and gene construct), and the disappearance of the smaller band (corresponding only to the wild-type region), as compared to an untransformed wild-type sample. See the full list of primers in appendix A.2.

4.4.3 Product Detection

Homoplasmic transformant cultures were lysed by micro-tip sonication on ice, and the cell lysates were run on a 4-20% SDS-PAGE gel. Standard Western blotting protocols were used. Proteins were transferred to a nitrocellulose membrane, which was subsequently blocked with 5% skim milk. Proteins were labeled using a mouse anti-c-terminal his-tag primary antibody, and a goat anti-mouse secondary antibody conjugated to horseradish peroxidase (HRP) (Cedarlane Laboratories). Products were detected using an Immobilon Western Chemiluminescent HRP substrate with a Positope multi-epitope protein (Life Technologies) used as the positive control. Reverse transcriptase PCR was performed using a 5 Prime Masterscript RT-PCR system, and using standard RNA purification protocols.

4.5 Results

36 homoplasmic strains, listed in Table 4.2, having incorporated most of the expression cassettes have been developed. Protein expression was quantified by Western blotting through the detection of the C-terminal his-tag added to each construct. Based on Western blotting, none of the strains express the genes of interest at detectable levels (i.e. only a signal from the positive control could be detected).

To determine the step at which expression was blocked, total RNA was extracted and the presence of recombinant gene-associated mRNA was determined by reverse-transcriptase PCR using standard protocols and primers specific to internal D1 sequences (found in the chloroplast genome) for the positive controls. Results are not explicitely shown here, as they only consist of blank gels or membranes, with a single band corresponding to positive controls. Based on results, no significant presence of recombinant gene mRNA was detected, which suggests pre- or post-transcriptional regulation, likely in the form of nucleolytic degradation of strands in the chloroplast, which is common, and perhaps exacerbated by the absence of stability-promoting elements which have been shown to be effective at increasing protein yields Lumbreras et al. 1998), through the premature end of transcription due to silencing or codon selection-related effects, which remain poorly understood (Purton et al. 2013), or through silencing effects that, although likely playing a role, currently remain incompletely characterized. Several strategies that may improve, or indeed allow, recombinant protein production, are discussed in detail in Chapter 2 of this document.

It should also be noted that several recent studies on the development of recombinant microalgal strains that were deemed successful had protein expression levels as low as 0.0003% TCP (Geng et al. 2004, Eichler-Stahlberg et al. 2009, Feng et al. 2014), which is lower than the detection threshold of the methods used to quantify expression in this work. The use of improved detection methods such as, for example, an ELISA of concentrated protein samples may very well provide comparable yields, but as these yields would not be sufficient for the targeted application, this possibility is not pursued further.

Table 4.2. List of homoplasmic transformants and associated expression cassettes.

Optimization Method	Expression cassette	Promoter		
		psbD	psbA (p546 plasmid)[a]	psbA (no duplicate initiation codon)[b]
1	*Y. intermedia* Phytase	4	3	5
	C. amalonaticus Phytase	1	1	--
	Alicyclobacillus Xylanase	--	--	3
	Actinomadura Xylanase	2	3	1
2	*Y. intermedia* Phytase	6	1	1
	Alicyclobacillus Xylanase	1	3	1

a - Constructs integrated using the recommended NcoI and XbaI restriction sites

b - Constructs integrated using the NdeI restriction site instead of the NcoI, and reading frame adjusted accordingly

4.6 Conclusions

Although no expression of the genes of interest was detected in any of the transformants, the work presented here, in conjunction with the results obtained in other published work, provides insight into the different factors affecting successful expression of recombinant products. The experimental approach described here is sound, and may serve as a useful guide for other researchers interested in undertaking similar work.

Difficulties in developing recombinant strains such as the ones indicated here are, currently, common, despite the development of increasingly reliable approaches and protocols. Although the work is certainly worth continuing, with, for example, the investigation of fused endogenous protein fragments, or the screening of additional promoters or insertion loci, given the uncertainty in terms of the time required to achieve success, continuation along these avenues of research were not deemed appropriate for this doctoral project. The work presented here involved the establishment of an entire research laboratory, including the building of research capacity through equipment acquisition, identification or development of all necessary protocols, and accumulation of expertise, none of which was present or available in our research group when this project was started. These contributions will, however, allow for the continuation of this work by other researchers, as well as the undertaking of a variety of other molecular biology, metabolic engineering, or biochemical engineering research programs.

4.7 References

Bai Y, Wang, J, Zhang Z, Yang P, Shi P, Luo H, Meng K, Huang, H, Yao, B. A new xylanase from thermoacifophilic Alicyclobacillus sp. A4 with broad-range pH activity and pH stability. J. Ind. Microbiol. Biotechnol. 2010; 37, 187-94.

Barnes D, Franklin S, Schultz J, Henry R, Brown E, Coragliotti A, Mayfield SP. Contribution of 5'- and 3'-untranslated regions of plastid mRNAs to the expression of *Chlamydomonas reinhardtii* chloroplast genes. Mol Gen. Genomics. 2005; 274, 625-36.

Bateman JM, Purton S. Tools for chloroplast transformation in *Chlamydomonas*: expression vectors and a new dominant selectable marker. Mol. Gen. Genet. 2000; 263, 404-10.

Boehm R. Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. Ann NY Acad Sci 2007; 1102, 121–34.

Carson M, Johnson DH, McDonald H, Brouillette C, DeLucas LJ. His-tag impact on structure. Acta Crystallogr D. 2007; 63, 295-301.

Eichler-Stahlberg A, Weisheit W, Ruecker O, Heitzer M. Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. Planta. 2009; 229, 873-83.

Feng S, Feng W, Zhao L, Gu H, Li Q, Shi K, Guo S, Zhang N. Preparation of transgenic *Dunaliella salina* for immunization against white spot syndrome virus in crayfish. Arch Virol. 2014; 159, 519-25.

Franklin S, Ngo B, Efuet E, Mayfield SP. Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. Plant J. 2002; 30, 733-44.

Franklin S, Mayfield SP. Recent developments in the production of human therapeutic proteins in eukaryotic algae. Expert Opinion Biol Therap 2005; 5, 225–35.

Freydank AC, Brandt W, Drager B. Protein structure modeling indicates hexahistidine-tag interference with enzyme activity. Proteins. 2008; 72, 173-83.

Galvan A, Quesada A, Fernandez E. The use of mutants to study nitrate assimilation in green microalgae. Scientia Marina. 1996; 60, 191-4.

Geng DG, Han Y, Wang YQ, Wang P, Zhang LM, Li WB, et al. Construction of a system for the stable expression of foreign genes in *Dunaliella salina*. Acta Bot. Sin.2004; 46, 342-6.

Georgianna DR, Hannon, MJ, Marcuschi M, Wu, SQ, Botsch K, Lewis, AJ, Hyun, J, Mendez, M, Mayfield, SP. Production of recombinant enzymes in the marine alga Dunaliella tertiolecta. Algal Res. Biomass. Bioprod. 2013; 2, 2-9.

Goldschmidt-Clermont M. Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker for site-directed transformation of Chlamydomonas. Nucl. Acids Res. 1991; 19, 4083-9.

Griesbeck C, Kobl I, Heitzer M. *Chlamydomonas reinhardtii*. Mol. Biotechnol. 2006; 34, 213-23.

Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol. 2004; 22, 346-53.

Harris EH, Burkhart BD, Gillham NW, Boynton JE. Antibiotic-resistance mutations in the chloroplast 16S and 23S ribosomal-RNA genes of Chlamydomonas reinhardtii – Correlation of genetic and physical maps of the chloroplast genome. Genetics. 1989; 123, 281-92.

Heitzer M, Eckert A, Fuhrmann M, Griesbeck C. Influence of Codon Bias on the Expression of Foreign Genes in Microalgae. In: León R, Gaván A, Fernández E, editors. Transgenic Microalgae as Green Cell Factories. Landes Bioscience/Springer Science + Business Media, LLC, 2007. p. 46-53

Huang HQ, Luo HY, Yang PL, Meng K, Wang YR, yuan TZ, Bai YG, Yao B. A novel phytase with preferable characteristics from *Yersinia intermedia*. Biochem. Biophys. Res. Comm. 2006; 350, 884-9.

Huang HQ, Luo HY, Wang YR, Fu DW, Shao N, Yang PL, Meng K, Yao B. Novel Low-Temperature-Active Phytase from *Erwinia carotovora* var. *carotovota* ACCC 10276. J. Microbiol. Biotechnol. 2009; 19, 1085-91.

Hussin ASM, Farouk AE, Greiner R, Salleh HM, Ismail AF. Phytate-degrading enzyme production by bacteria isolated from Malaysian soil. World J Microbiol. Biotechnol. 2007; 23, 1653-60.

Ikemura T. Codon usage and TRNA content in univellular and multicellular organisms. Mol Biol. Evol. 1985; 2, 13-34.

Kato K, Marui T, Kasai S, Shinmyo A. Artificial control of transgene expression in *Chlamydomonas reinhardtii* chloroplast using the *lac* regulation system from *Escherichia coli*. J. Biosci. Bioeng. 2007; 104, 207-213.

Kim YO, Kim HW, Lee JH, Kim KK, Lee SJ. Molecular cloning of the phytase gene from *Citrobacter braakii* and its expression in *Saccharomyces cerevisiae*. Biotechnol. Lett. 2006; 28, 33-8.

Kim SW, Kim JB, Lee WS, Jung WH, Ryu JM, Jang HW, Jo YB, Jung JK, Kim JH. Enhanced protease cleavage efficiency on the glucagon-fused interleukin-2 by the addition of synthetic oligopeptides. Protein Expres. Purif. 2007; 55, 159-65.

Li SS, Tsai HJ. Transgenic microalgae as a non-antibiotic bactericide producer to defend against bacterial pathogen infection in the fish digestive tract. Fish Shellfish Immunol 2009; 26, 316–25.

Lithwick G, Margalit H. Hierarchy of sequence-dependent features associated with prokaryotic translation. Genome Res 2003; 13, 2665-73

Lumbreras V, Stevens DR, Purton S. Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. Plant J. 1998; 14, 441-7.

Luo HY, Huang HQ, Yang PL, Wang YR, Yuan TZ, Wu NF, Yao B, Fan YL. A Novel Phytase appA from *Citrobacter amalonaticus* CGMCC 1696: Gene Cloning and Overexpression in *Pichia pastoris*. Curr. Microbiol. 2007; 55, 185-92.

Manuell AL, Beligni MV, Elder JH, Siefker DT, Tran M, Weber A, et al. Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast. Plant Biotechnol. J. 2007; 5, 402-12.

Mayfield SP, Franklin SE, Lerner RA. Expression and assembly of a fully active antibody in algae. Proc. Natl. Acad. Sci. USA. 2003; 100, 438-42.

Mayfield SP, Schultz J. Development of a luciferase reporter gene, luxCt, for *Chlamydomonas reinhardtii* chloroplast. Plant J. 2004; 37, 449-58.

Mayfield SP, Manuell AL, Chen S, Wu J, Tran M, Siefker D, et al. *Chlamydomonas reinhardtii* chloroplasts as protein factories. Curr. Opin. Biotechnol. 2007; 18, 126-33.

Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status of the year 2000. Nucleic Acids Res. 2000; 28, 292.

Polizeli MLTM, Rizzatti ACS, Monti R, Terenzi HF, Jorge JA, Amorim DS. Xylanases from fungi: properties and industrial applications. Appl Microbiol Biotechnol.2005; 67, 577-91.

Purton S, Szaub JB, Wannathong T, Young R, Economou CK. Genetic Engineering of Algal Chloroplasts: Progres and Prospects. Russian J. Plant Physiol.2013; 60, 491-9.

Raja R, Hemaiswarya S, Kumar NA, Sridhar S, Rengasamy R. A perspective on the biotechnological potential of microalgae. Crit Rev Microbiol 2008; 34, 77–88.

Ran W, Higgs PG. The Influence of Anticodon-Codon Interactions and Modified Bases on Codon Usage Bias in Bacteria. Mol. Biol. Evol. 2010; 27, 2129-40.

Rao DECS, Rao KV, Reddy TP, Reddy VD. Molecular characterization, physicochemical properties known and potential applications of phytases: An overview. Crit. Rev. Biotechnol. 2009; 29, 182-98.

Rasala BA, Muto M, Lee PH, Jager M, Cardoso RMF, Behnke CA, Kirk P, Hokanson CA, Crea R, Mendez M, Mayfield SP. Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. Plant Biotechnol J. 2010; 8, 719-33.

Rosenberg JN, Oyler GA, Wilkinson L, Betenbaugh MJ. A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution. Curr Opinion Biotechnol 2008; 19, 430–6.

Sambrook J, Russel DW. Molecular Cloning: A laboratory manual 3rd ed. CSHL Press, Cold Spring Harbor, New York. 2001.

Sriyapai, T, Somyoonsap, P, Matsui, K, Kawai, F, Chansiri, K. Cloning of a thermostable xylanase from Actinomadura sp. S14 and its expression in *Escherichia coli* and *Pichia pastoris*. J. Biosci. Bioeng. 2011; 111, 528-36.

Stern DB, Harris EB (editors). The Chlamydomonas Sourcebook, 2nd ed. Vol. 2 – Organellar and Metabolic Processes. Academic Press, Elsevier. 2009.

Surzycki R, Greenham K, Kitayama K, Dibal F, Wagner R, Rochaix JD, et al. Factors effecting expression of vaccines in microalgae. Biol. 2009; 37, 133-38.

Tran M, Zhou B, Pettersson PL, Gonzalez MJ, Mayfield SP. Synthesis and assembly of a full length human monoclonal antibody in algal chloroplasts. Biotechnol Bioeng 2009; 104, 663–73.

Weathers PJ, Towler MJ, Xu J. Bench to batch: advances in plant cell culture for producing useful products. Appl Microbiol Biotechnol 2010; 85, 1339–51.

Wurtz EA, Boynton JE, Gillham NW. Perturbation of chloroplast DNA amounts and chloroplast gene transmission in *Chlamydomonas reinhardtii* by 5-fluorodeoxyuridine. Proc Natl Acad Sci USA 1977; 74, 4552-6.

Wyss M, Pasamontes L, Friedlein A, Rémy R, Tessier M, Kronenberger A, Middendorf A, Lehmann M, Schnoebelen L, Röthlisberger U, Kusznir E, Wahl G, Müller F, Lahm HW, Vogel K, van Loon APGM. Biophysical Characterization of Fungal Phytases (myo-Inositol Hexakisphosphate Phosphohydrolases) : Molecular Size, Glycosylation Pattern, and Engineering of Proteolytic Resistance. Appl Environ Microbiol 1999; 65, 359-66.

Yoon, SM, Kim, SY, Li, KF, Yoon BH, Choe, S, Kuo, MMC. Transgenic microalgae expressing Escherichia coli AppA phytase as feed additive to reduce phytate excretion in the manure of young broiler chicks. Appl. Microbio. Biotechnol.2011; 91, 553-63.

Chapter 5 - Selective screening of the microflora of the digestive tract of *Castor canadensis* for the isolation of novel cellulase genes

Authors: Gabriel Potvin, Carol Pietka, Jason Zhang

5.1 Abstract

The digestive tract of a Canadian beaver (*Castor canadensis*) was screened for cellulase-producing microorganisms, with the objective of isolating novel cellulase sequences that may have some characteristics of interest for the production of cellulosic bioethanol. Microflora samples from the beaver's stomach, small and large intestines and cecum were cultured, producing a wide range of observable phenotypes, and the bulk microbial stock was selectively screened for cellulase production. A strain of cellulase-producing *Bacillus thuringiensis* was identified by 16S rRNA gene sequencing, fragments of cellulase genes were amplified using primers designed based on highly-conserved cellulase domains identified in other microbial species and sequenced, and a full gene sequence for a β-glucosidase gene was obtained. The gene was optimized and expressed in *E. coli*, and its product characterized. The β-glucosidase produced displays an enzymatic activity optimum at pH 6 and 50˚C, which corresponds to a higher optimal pH and temperature than other reported *B. thuringiensis* enzymes.

5.2 Introduction

Over the last few decades, the ongoing depletion of fossil fuel reserves, geopolitical instability and the resulting volatility of the fossil fuel market, and a growing societal acceptance of environmental responsibility have all spurred research into the development of carbon-neutral and renewable energy sources. One area that has particularly garnered interest is the development of an economically viable process for the production of cellulosic bioethanol, which can be used as a primary fuel or as a fuel additive having anti-knock properties (Park et al. 2010). In particular, there is a large focus on the breakdown of lignocellulosic biomass, given its widespread availability and low cost, into sugars that can subsequently be converted to ethanol through fermentation.

Although the development of an economically viable bioethanol production process is hindered by a variety of logistical (e.g. biomass supply and transport), infrastructure (e.g. fuel

delivery and transport), and scale-related challenges, technical difficulties related to biomass processing remain key obstacles to overcome.

The efficient hydrolysis of cellulose is one of the basic requirements, and unfortunately one of the more challenging aspects, of cellulosic ethanol production. Cellulose is traditionally degraded by acid hydrolysis, through the high-temperature treatment of biomass with concentrated sulphuric acid (Zhu et al. 2011), which involves high processing costs, and the handling of hazardous and environmentally harmful reagents. Alternatively, cellulose can be hydrolyzed with cellulases, which would avoid the harsh operating conditions required by acid hydrolysis, thus leading to lower processing costs and presenting considerable environmental benefits. Cellulases, like all enzymes, are easily degraded, and in their inactive form can be completely recycled as amino-acid supplements to yeast fermentation media.

Cellulase-mediated hydrolysis, however, presents its own challenges. It is, for one, a multi-enzyme process, which complicates reactor and process design (Andric et al. 2010). The general term cellulase refers to all members of the multi-enzyme family, notably endocellulases (EC 3.2.1.4), which catalyze the degradation of inter-molecular bonds in crystalline cellulose, exocellulases (EC 3.2.1.91), which convert cellulose into cellobiose or cellotetrose intermediates, and β-glucosidases (EC 3.2.1.21), which degrade these intermediates into glucose monomers (Beguin and Aubert 1994). In the present work, the general term cellulase refers to any and all enzymes included in this family, unless otherwise indicated.

To further complicate matters, lignins, aromatic polymers that can account for up to 40 dry wt% of certain biomass sources, are large and highly stable compounds that reduce cellulase activity by limiting its access to its substrate by forming a structural matrix around crystalline cellulose strands. To maximize the effectiveness of cellulase, lignocellulosic biomass must be thermally and mechanically treated, and fractionated to remove lignin residue prior to enzymatic digestion. Cellulase efficiency is also limited by the inhibitory effects of cellobiose intermediates and glucose products, as well as potential interference by high substrate concentrations. Although progress has been achieved in developing value-added applications of lignin residues that help to offset these processing costs, the price of the bioethanol produced in this process is not competitive with that of traditional fossil fuels (Doherty et al. 2011).

Much of the ongoing research in lignocellulosic bioethanol production focuses on improving the economic viability of the process by reducing the cost of cellulolytic enzyme production

(Oehgren et al. 2007, Maki et al. 2009), improving the fractionation of cellulose and lignin compounds (Moxley et al. 2008, Kim et al. 2009), or developing enzymes capable of operating on a wider range of substrates or able to overcome lignin interference (Bettiga et al. 2009, Yano et al. 2009).

The current work uses a "bio-prospecting" approach and seeks to isolate novel cellulase genes whose products may have characteristics of industrial interest, i.e. allow for the effective degradation of cellulose. The identification of new, more effective cellulase enzymes could improve the economic viability of the bioethanol production process, or provide a new starting point for subsequent protein chemistry studies that could improve the effectiveness of the currently available cellulase. Canadian beavers (*Castor canadensis*), whose diets consist mainly of plant matter including bark and wood, can digest 20-40% of the cellulose contained in these sources (Vispo et al. 1995), which suggests that the beaver's digestive tract contains cellulase-secreting microorganisms. The present study seeks to isolate and identify these microorganisms, isolate and sequence the genes for these potentially novel cellulases, and characterize the genes' products. To our knowledge, no work has previously been completed on the screening of the digestive tract microflora of beavers.

5.3 Materials and Methods

5.3.1 Bulk microbial stock

A wild adult Canadian beaver was legally trapped in the Bolingbroke region of Ontario, Canada, donated to our research group, and kept on ice for approximately three hours prior to processing. The entire stomach and tissue samples from the large and small intestines, and the cecum were excised (see Figure 1 for approximate sample locations), and kept on ice until bacterial samples were taken. Swabs were taken from each sample, spread on solid LB (Tryptone 10 g/l, Yeast extract 5 g/l, NaCl 10 g/l, agar 15 g/l) and YPD (Peptone 20 g/l, Yeast extract 10 g/l, Glucose 10 g/l, agar 15g/l), and incubated at 37ºC for 24h. Bacterial swabs from each plate were then inoculated in separate flasks containing 100 ml sterile liquid LB or YPD, and incubated in a rotary shaker for 36 hours at 37ºC and 200 rpm to produce bulk microbial stocks.

5.3.2 Selective screening for cellulase-producing microorganisms

One hundred microliters of each bulk stock was screened on pine cone plates (25 g/l finely ground cleaned and sterilized pinecone powder, 5 g/l peptone, 5 g/l NaCl, 3 g/l yeast extract, 20 g/l agar). Individual colonies were selected and grown in 5 ml liquid LB in 50 ml falcon tubes at 37ºC and 150 rpm for 24 hours. Liquid cultures were transferred to flasks containing 100 ml liquid LB and incubated for a further 24 hours. Resulting liquid cultures were used as pure stocks of cellulase-producing microorganisms. Aliquots of each culture were stored at -80ºC in 25% glycerol.

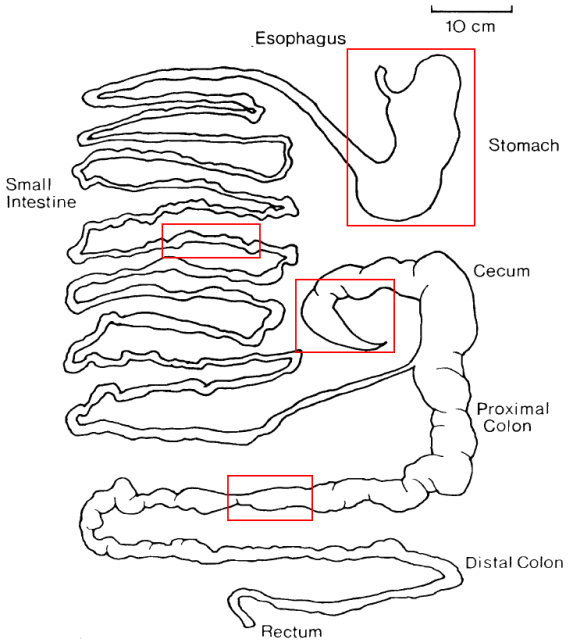

Figure 5.1. *Digestive tract of an adult C. canadensis. The approximate size and location of excised samples are indicated by the boxes. Figure taken and modified from Vispo et al. (1995).*

5.3.3 Identification of Microbial Species

Species identification of cellulase-producing microorganisms was achieved by 16S rRNA gene amplification and sequencing. Total DNA of cultures was isolated using a Promega Wizard DNA Purification Kit, and the 16S rRNA gene was amplified using standard forward (5'-AGAGTTTGATCTGGCTCAG - 3') and reverse (5' – ACGGCTACCTTGTTAC GACTT - 3') primers, both synthesized by Life Technologies Inc. PCR products were purified using a

QIAquick PCR Purification kit (Qiagen Inc.), and sequenced by StemCore Laboratories (Ottawa ON, Canada). Species were identified by BLAST alignment using NCBI databases.

5.3.4 Cellulase gene isolation

The genomes of the species most closely related to the isolated microorganism, as determined by the BLAST analysis of 16S rRNA genes, were analyzed to identify putative sequences for endo- and exo-cellulases, and/or β-glucosidases. The highly conserved regions of these sequences were used to design degenerate forward and reverse primers for PCR amplification, in order to confirm partial sequences. Flanking primers are then designed based on whole genome data to amplify the entire gene, which is then sequenced. Restriction sites for BamHI and XbaI were added to both ends of the isolated sequence to facilitate subsequent manipulation, codons were optimized, and the gene constructs were synthesized by Genscript Inc.

5.3.5 Bacterial Transformations and Transformation Vectors

The synthesized gene constructs were used to transform XL-1 Blue supercompetent *Escherichia coli* cells (Agilent Technologies), which are used for all cloning steps using recommended protocols. Gene sequences were cloned into a pET21a+ plasmid using standard protocols for subsequent expression under the control of the T7 promoter, and this vector was then used to transform BL21(DE3)pLysS competent cells for expression of the construct.

5.3.6 β-glucosidase Assay

Cellulase production was induced in BL21 cells by adding isopropyl β-D-1-thiogalactopyran (IPTG) to cultures to a final concentration of 30 mM, and incubated at 37°C for two hours. 2 ml aliquots of induced cultures were lysed, centrifuged, and the supernatant used for the enzyme assay, which was performed as described previously (Zhang et al. 2009). Briefly, cellobiose was dissolved in 50mM citrate buffer to a final concentration of 15 mM and used as the substrate solution. Equal volumes of substrate solution and culture supernatant were incubated at 50°C unless otherwise specified for 30 minutes, and the reaction was stopped by immersion in boiling water for 5 minutes. The concentration of glucose in samples was measured with an Agilent 1100 series HPLC using an Aminex HPX-87H ion exchange column, and compared to standards.

The assay was repeated at different pH values (adjusted by the addition of NaOH) and temperatures to determine the enzymatic activity of cell lysates under different conditions.

5.4 Results/Discussion

5.4.1 Bulk Stock and Screening

Although the isolation of cellulase-producing microorganisms was successful using the methodology described in the previous section, a few design considerations are worthy of further discussion, namely in terms of the media used for screening, and of the type of screened microorganism.

In terms of the former, a multitude of microbial phenotypes were observed on the initial LB and YPD plates. At the screening step, plates containing finely ground pinecones as the sole carbon (cellulose) source were used, and a single microbial phenotype was observed. The effectiveness of these unconventional pinecone plates, was confirmed by successfully repeating the screening procedure using standard carboxymethylcellulose (CMC) plates ($NaNO_3$ 2 g/l, K_2HPO_4 1 g/l, $MgSO_4$ 0.5 g/l. KCl 5 g/l, CMC 2 g/l, Peptone 0.2 g/l, agar 17 g/l) as described by Kasana et al. (2008). This isolated strain was used for all subsequent work.

In terms of the screening procedure, only aerobes and facultative aerobes were screened for cellulase production, while strict anaerobes were not viable under the experimental conditions. This approach is justified by previous work demonstrating that most cellulase-producing microorganisms of interest are facultative anaerobes (Reguera et al. 2001, Wenzel et al. 2002). It would be interesting, in future work, to expand this work to perform an anaerobic screening of the digestive tract to verify the validity of this assumption.

5.4.2 Species Identification

The total DNA of ten isolated strain was extracted and the 16S rRNA gene was amplified by PCR, sequenced, and identified by BLAST analysis using NCBI databases. The full sequence is shown in Figure 5.2. Based on BLAST results, the isolated microorganism is a strain of *Bacillus thuringiensis* (100% alignment), a soil bacterium, and therefore likely not a part of the beaver's endogenous microflora. Since the stomach and intestines were partially full when excised, it is likely that the bacteria were present as residual biomass on the plant matter.

TCGAGCGAATGGATTGAGAGCTTGCTCTCAAGAAGTTAGCGGCGGACGGGTGAGCAACACGTGG
GTAACCTGCCCATGGGATAACTCCGGGAAACCGGGGCTAATACCGGATAACATTTTGAACTGCA
TGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGGACCCGCGTCGCATTAGCTAG
TTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACA
CTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGAC
GAAAGTCTGACGGAGCAACGCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTT
AGGGAAGAACAAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGG
CTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCG
TAAAGCGCGCGCAGGTGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTC
ATTGGAAACTGGGAGACTTGAGTGCACAAGAGGAAAGTGGAATTCCATGT

Figure 5.2. *Full sequence of the amplified 16S rRNA gene fragment. The sequence corresponds to a strain of B. thuringiensis (full genome of closest match GenBank accession number CP009351.1.1)*

Although endo- and exo-cellulases from some *B. thuringiensis* strains have previously been partially characterized (Lin et al. 2012), and that β-glucosidases are generally highly conserved, differences in selective pressures in specific environments may cause an evolutionary divergence sufficient to substantially change the enzymatic properties of the products of interest.

5.4.3 Primer design for isolation of full sequence

Putative cellulase gene sequences of the species most closely related to *B. thuringiensis* based on the homology of 16S rRNA sequences (i.e. several strains of *B. thuringiensis, Bacillus weistephanensis, Bacillus amyloliquefaciens*， and *Bacillus subtilis* – all soil bacteria) were aligned to identify highly-conserved regions. These high-homology sequences were used to design primer pairs to produce fragments of different cellulase genes. The corresponding sequences were amplified and sequenced, and compared to the full *B. thuringiensis* genome to identify gene products and loci. The list of primer pairs and putative gene products are listed in Table 5.1. The identification of resulting PCR products is done by BLAST alignment of the fragment sequences between 300 and 800 bp in length, which corresponds to the range most likely to contain fragments of cellulase genes based on the designed primers and sequences for known/putative genes.

The fifth pair of primers yielded a fragment of a putative *B. thuringiensis* β-glucosidase gene with a 98% homology with fragments of known sequences, which suggests that the full sequence is also highly conserved. To confirm this, primer pairs flanking and internal to the identified gene fragment locus, listed in Table 5.2, were designed to assemble the full sequence from 4

overlapping contigs. The resulting contigs were aligned, and the final full cellulase sequence is shown in Figure 5.3.

Table 5.1. PCR products obtained with primers designed based on highly conserved regions of putative cellulase sequences from closely related bacterial species

Primer Pairs		Product Size (bp)	(Putative) Source Genes
F	GTTTGCCGACGAACTTTCTACTGA	441	ABC Transporter Gene
R	CATAGACGGAATACCATTTACAG		
F	ATCGCATTCGCAGTAGATGTTGG	309	PEP C
R	CATAGACGGAATACCATTTACAG		
F	CTGATACATACCACAAGGGCGATTC	303	Deblocking Aminopeptidase
R	ATTTCTCGTAATCTTCCGTCCAA		
F	AACGGTGACGTGAACTGGAAGCG	785	Glucanase Aminopeptidase
R	CATTTCCAGACGCATCGCTTGTT		
F	GGGGACTCCTTGGGATGAGATTG	488	β-glucosidase
R	GCGGATTTACGCCATTGTAGGAA		

Table 5.2. Primer pairs used to produce four overlapping contigs that provide the entire sequence of the partially identified β-glucosidase gene

Contig		Primer Pairs
1	**F**	ACCACATTTGTTTAAGTCTCCACCT
	R	CCATTCGTGAAGTAGACCAGGAAAC
2	**F**	ATCCCTGCTTATTCCACATTACA
	R	CTATCAGGGATTAGTAAAGGAGG
3	**F**	CCATTTCTCATAATCTTCCGTCCAA
	R	AATTGTTGCGTTATTAGGCATGATG
4	**F**	TTTTGAATCGCCCTTGTGGTATG
	R	GGCCCTAGTTTGGAGGGAGAATT

The isolated sequence is nearly identical to those found in sequenced *B. thuringiensis* strains, with only two nucleotide differences (G586A and G968T), which is in agreement with the assumptions on the highly conserved nature of this enzyme. A further characterization of the thermochemical properties of this product will allow the determination of whether it could serve as a basis for subsequent protein engineering work towards the development of more effective biochemical catalysts.

AAACTATTTATGTGCTTGTAAATCGTTTTTTATTACTTCCCAGAGTGAATTATCTTGAGGCATC
GGTTTACCCTCTTCATCCCATTTTATTTTTCCGGTTCCCAATTCATAATTCATCCCTGTCCATG
TGTCTTCACGGAATGCATAGAATGATTTATGCCAACCTTTTTGGTTGAAGATAGAAATAAGATC
TTGCATGTATTGGGTAGCTCCTGGAACAGTACGGTTAATTCCAAACTCCTCTGCAATAATTCGA
TTAGATGATACATGATTTTTCTTAGACCATTGTTGGATTGGCTTCAAAAATTTCTCTAATCCCT
GCTTATTCCACATTACAGGTTTCTCTAAATCTCCTACTTTTACTAATCCTGGATATTGATACTC
CTTGTTTTGTTTTTCACCTTGACTCGTTAATTCATATGGTTCATACATATGAAATGCATATAGT
GTTTTTTTATCATTTACTGGTTTTAAATATTTAAAGGCCCATGGAGTAGCATATAAACCTGAAT
CTAAAATAATCGGCGTTTCCTGGTCTACTTCACGAATGGAATTGATCACTTTTTGATACAATCT
GTTTAAATCAGCTGTCGTTCCCTTCACTTTTGAATACCATTTCTCATAATCTTCCGTCCAAAAA
TCATTATATCTATTATTTTTAGCTGTTTCTGGATGTGGTTCATTTATAATATTATAACCAACCA
CCGCAGGGTGATCTTTTAATTCCAGAGCAAGATCCTTCCAAAATTGACTTGCTTGTTCTTGATA
CTTCTCTTCTTCCCATATTCTGTCGTCATTCTTGTTATTATTAAATTGGCGCCATCGATCACCA
GGTAAGGATAACATTGTAAGAACAACTTTCATTCCTTGTGATTGTGCGGCATCTAAATCCGCCT
TTAATTTTTCTAAATCCTCCTTTACTAATCCCTGATAGTTATCTGCATTACCTATCAGAAAATC
CTTTCCAGAAGTATCTGGTTTATCCTCAAATAGAAAACCTTTATCTTTTGCCCATTTATCAGGT
GCTAAACGTACATATTCAATATTAGCTTCTTTTGCACTTTTATAGTTTTCAGGTAATGACGTAC
TATTCATGAAATTAGTACCCTTTCTTTTCGAATCCCAAAAACTAATTTTTGAATCGCCCTTGTG
GTATGTATCAGCTCTTACATTCATTTCCTGTACTCCAAAACTCATCATGCCTAATAACGCAACA
ATTGGCAAAATTTTTTTCACAACTTTTCCTTCTTTCTCTTTATATTTAACACACAT

Figure 5.3. *Full sequence of the isolated cellulase gene. The sequence has a 99% homology with other B. thuringiensis cellulase genes (from sequenced genomes with GenBank accession numbers CP004123.1, CP003889.1, and CP001907.1).*

5.4.4 Enzyme Characterization

The full β-glucosidase sequence shown in Figure 5.3 was codon optimized for expression in *E. coli*, restriction sites for NcoI and XbaI were added at the 5' and 3' ends respectively, and the whole construct was flanked by 'artificial' primer annealing sites for full sequence amplification using Fcell (5'-GACCGACATCATCAGTTTGACC-3') and Rcell (5'-CCAGCCTAACGGATACGAACTAC-3') primers. The β-glucosidase was inserted in-frame into a pET21a+ plasmid under the control of a T7 promoter. Lysate supernatants were used as the enzyme solution for the activity assays.

The β-glucosidase assay was performed at different pH and temperature values to determine the enzymatic characteristics of the isolated product. Results are shown in Figure 5.4. The assay was also performed at pH 4 and 5, but no enzymatic activity was detected at these values, and the data is therefore not presented here.

Based on these results, the β-glucosidase activity peaks at 50˚C and pH 6, indicating that the enzyme in question is more mesophilic than the previously studied *B. thuringiensis* endo- and exo-cellulases (Lin et al. 2012) and β-glucosidase (Panalazaridou et al. 2003), and has a higher optimal temperature. Although the mesophilic pH optimum is not beneficial for cellulosic ethanol production, in particular the initial acid hydrolysis of the biomass, the better thermostability could provide an advantage in this process or further studied to determine the conformational or biochemical factors conferring this improved thermostability.

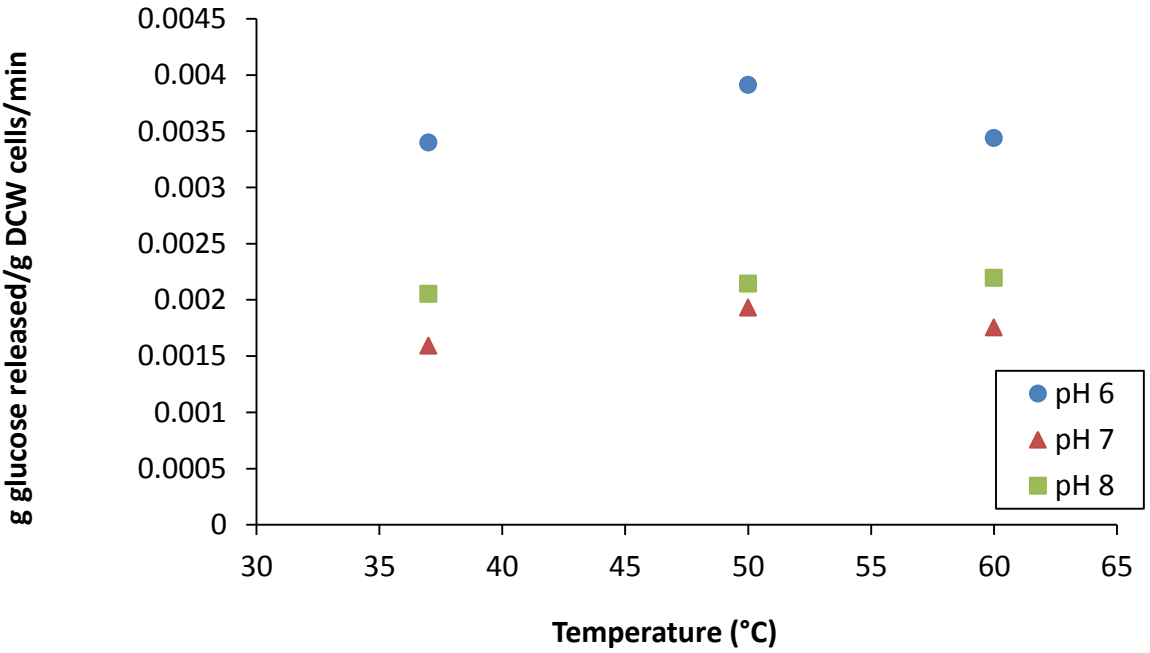

Figure 5.4. *β-glucosidase activity of cleared cell lysates at different pH levels and temperatures. The assay was performed in citrate buffer adjusted to pH 6.0, 7.0, and 8.0, at 37°C, 50°C, and 60°C for 30 minutes each. The amount of glucose liberated from 750 µL of cellobiose solution by each aliquot of lysate was quantified by HPLC. All data is normalized to wild-type E. coli lysate activity.*

5.5 Conclusions

The risk of such 'bio-prospecting' approaches for the isolation of novel genes is that the isolated sequences may yield products with characteristics that are not competitive with commercially available alternatives, or that these sequences will be similar to those already known. Although the cellulase sequence isolated from the strain of *B. thuringiensis* obtained from the beaver sample is very similar to those identified in other strains of the same species, its product has a

more mesophilic pH optimum and a higher thermal optimum. As previously discussed, however, this work is worth repeating using an anaerobic screening process, which would yield results more aligned with the initial objectives this work, namely the identification of candidates specific to the digestive tract of the beaver.

5.6 References

Andric P, Meyer AS, Jensen PA, Dam-Johansen K. Reactor design for minimizing product inhibition during enzymatic lignocelluloses hydrolysis: I. Significance and mechanism of cellobiose and glucose inhibition on cellulolytic enzymes. Biotechnol Adv. 2010; 28, 308-24.

Beguin P, Aubert JP. The biological degradation of cellulose. FEMS Microbiol Rev. 1994; 13, 25-58.

Bettiga M, Bengtsson O, Hahn-Hagerdal B, Gorwa-Grauslund MF. Arabinose and xylose fermentation by recombinant saccharomyces cerevisiae expressing a fungal pentose utilization pathway. Microb Cell Fact 2009; 8.

Doherty WOS, Mousavioun P, Fellows CM. Value-adding to cellulosic ethanol: Lignin polymers. Ind Crops Prod. 2011; 33, 259-76.

Kasana RC, Salwan R, Dhar H, Dutt S, Gulati A. A Rapid and Easy Method for the Detection of Microbial Cellulases on Agar Plates Using Gram's Iodine. Curr Microbiol. 2008; 57, 503-7.

Kim Y, Mosier NS, Ladisch MR. Enzymatic digestion of liquid hot water pretreated hybrid poplar. Biotechnol Prog. 2009; 25, 340-8.

Lin L, Kan X, Yan H, Wang D. Characterization of extracellular cellulose-degrading enzymes from *Bacillus thuringiensis* strains. Electronic J. Biotechnol. 2012; 15, 1-7.

Maki M, Leung KT, Win W. The prospets of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. Int J Biol Sci. 2009; 5, 500-16.

Moxley G, Zhu Z Zhang YHP. Efficient sugar release by the cellulose solvent-based lignocelluloses fractionation technology and enzymatic cellulose hydrolysis. J Agric Food Chem. 2008; 56, 7885-90.

Oehgren K, Vehmaanperae J, Siika-Aho M, Galbe M, Viikari L, Zacchi G. High temperature enzymatic prehydrolysis prior to simultaneous saccharification and fermentation of steam pretreated corn stover for ethanol production. Enzyme Microb Technol. 2007; 40, 607-13.

Papalazaridou A, Charitidou L, Sivropoulou A. beta-glucosidase enzymatic activity of crystal polypeptide of the *Bacillus thuringiensis* strain 1.1. J. Endotoxin. Res. 2003; 9, 215-24.

Park C, Choi Y, Kim C, Oh S, Lim G, Moriyoshi Y. Performance and exhaust emission characteristics of a spark ignition engine using ethanol and ethanol-reformed gas. Fuel. 2010; 89, 2118-25.

Reguera G, Leschine SB. Chitin degradation by cellulolytic anaerobes and facultative aerobes from soils and sediments. FEMS Microbiol Lett. 2001; 204, 367-74.

Vispo C, Hume ID. The digestive tract and digestive vunction in the North American porcupine and beaver. Can J Zool. 1995; 73, 967-74.

Wenzel M, Schönig I, Berchtold M, Kämpfer P, König H. Aerobic and facultatively anaerobic cellulolytic bacteria from the gut of the termine Zootermopsis angusticollis. J Appl Microbiol. 2002; 92, 32-40.

Yano S, Murakami K, Swayama S, Imou K, Yokoyama S. Ethanol production potential from oil palm empty fruit bunches in southeast Asian countries considering xylose utilization. J Jpn Inst Energy. 2009; 88, 923-6.

Zhang, Y.H.P., Hong, J., and Ye, X. (2009). Cellulase Assays. In Biofuels: Methods and Protocols, JR Mielenz, ed. (New York, NY: Springer) pp. 213-230.

Zhu YM, Maltern M, Torry-Smith M, McMillan JD, Stickel JJ. Calculating sugar yields in high solids hydrolysis of biomass. Biores Technol. 2011; 102, 2897-903.

Chapter 6 - Statistical medium optimization for the increased production of recombinant phytase in the fed-batch cultivation of *Pichia pastoris*

Authors: Potvin G, Li Z, Zhang Z

6.1 Abstract

Using a 20-run central composite design, standard BSM media was optimized for the production of recombinant phytase by *Pichia pastoris* in fed-batch cultivations using glucose as the carbon source. The phytase activity in the supernatant of the cultures at the end of 48 hour cultivation runs was modeled as a function of the medium composition, and this model was successfully validated. Using Inductively Coupled Plasma spectrometry (ICP), residual elements were quantified in the cultivation broth at the end of each run to confirm that differences in final cell density and enzyme activity were not due to nutrient depletion, but indeed associated with the variations in medium composition. The optimized media contains significantly lower concentration of potassium and magnesium sources (28.7 g/l and 4.8 g/l respectively), as well as a reduced concentration of trace salts, and based on experimental results, either provides the same enzyme yields as the standard media, but at a lower cost, or significantly outperforms it under the same conditions.

6.2 Introduction

The methylotrophic yeast *Pichia pastoris* has garnered immense interest as a versatile recombinant protein production platform since its discovery and characterization in the 1970s. This system is particularly valuable for the high cell densities achievable on defined media, for its strong methanol-inducible (pAOX1) and constitutive (pGAP) promoters, the high obtainable yields of intracellular and secreted proteins, and its strong preference for respirative, as opposed to fermentative, growth. *P. pastoris* systems have successfully been used for the large-scale production of hundreds of products, including therapeutics and industrial enzymes (Macauley-Patrick et al. 2005, Potvin et al. 2012).

The particular enzyme of industrial interest in this work is phytase, which catalyzes the hydrolysis of phytic acid, the main form of phosphorus storage in plants (Wodzinski and Ullah 1996). Since monogastric crop animals such as swine, poultry and fish are unable to extract

sufficient phosphorus from phytic acid-rich feeds, phytase is commonly used as an additive, which not only increases the bioavailability of phosphorus for the animals by hydrolyzing phytic acid, but also eliminates the need for feed supplementation with inorganic phosphorus, which is poorly absorbed and associated with a host of environmental concerns (Haraldsson et al. 2005, Vohra et al. 2006).

Several 'upstream' strategies are available to increase the biomass concentration or enzyme yields of cultures by modifying their underlying genetic or metabolic pathways, but from a purely bioprocessing standpoint, these objectives are typically achieved by optimizing operating conditions and parameters, or by optimizing the composition of the growth media, the latter of which is considered here. Although most of the medium optimization work of this type is performed in flasks (Li et al. 2007, Gao and Shi 2013) due to obvious time- and equipment-related advantages, results obtained in such experiments are not always transferable to larger scale (i.e. bioreactor-based) fed-batch or continuous cultures, which can exhibit completely different protein production profiles (Bawa et al. 2014).

In this work, the composition of Basal Salt Media (BSM), commonly used in yeast-centered industrial bioprocesses, has been systematically optimized for the production of pGAP-mediated recombinant phytase by *P. pastoris*. All experimental runs were performed in fed-batch bioreactors with glucose as the carbon source, using industrially-relevant parameters. Although the developed protein yield model and optimized medium composition may vary based on the operating conditions, carbon source, specific strain used, and nature of the enzymes produced in other applications, they may prove useful as a valuable starting point for product yield improvement studies in similar industrial bioprocesses.

6.3 Materials and Methods

6.3.1 Strain

A strain of zeocin-resistant *P. pastoris* expressing and secreting a recombinant phytase under the control of the constitutive GAP (glyceraldehyde-3-phosphate dehydrogenase) promoter (pGAP) was developed and provided by Zell Technologies Inc and used for all experimental runs.

6.3.2 Strain Upkeep and Inoculum Preparation

Yeast cultures were streaked on fresh YPD plates (10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose, 15 g/l agar) supplemented with zeocin every 2 weeks and incubated for 2 days at 30˚C. Reactor inoculum was prepared by inoculating 100 ml of YPD media with a colony from one of these plates, and grown in a shaker at 30˚C and 250 rpm until an OD_{600} of 0.8 was reached. 50 ml of liquid culture was used to inoculate each reactor.

6.3.3 Media

Cultures were grown in BSM media (0.6 g/l $CaSO_4 \cdot 2H_2O$, 42.9 g/l KH_2PO_4, 5.17 g/l $(NH_4)_2SO_4$, 14.33 g/l K_2SO_4, 0.5 ml/l H_2SO_4, 5.71 g/l $MgSO_4 \cdot 7H_2O$ and 4 ml/l PTM1 Trace Salts) or variations thereupon (see section 6.4 for experimental design). Reactors contained 1L of the appropriate media at the beginning of each run, containing 5 g/l of glucose. Feed solutions for fed-batch cultivations consist of the appropriate media, supplemented with 600 g/l of glucose.

6.3.4 Bioreactor

Fed-batch cultivations were performed with New Brunswick Scientific (NBS) Bioflo 110 3L (2L working volume) bioreactors equipped with dual stainless-steel Rushton impellers. Reactors were kept at 30˚C, sparged with 3-5 l/min air to maintain a dissolved oxygen (DO) concentration of 30%, with the agitation rate cascaded to the DO, as measured online by NBS DO probes. pH was regulated by the addition of 30% ammonium hydroxide, with a feed rate controlled by online pH monitoring. Process control and monitoring was achieved with the NBS BioCommand software.

6.3.5 Fed-Batch Cultivation

After inoculation, once the initial glucose was consumed, the 60% glucose medium solution was fed at an appropriate rate to maintain residual glucose under 0.5 g/l to avoid substrate-associated growth inhibition. Feed rate and residual glucose concentrations for a typical run are shown in Figure 6.1. Culture samples were taken every hour and centrifuged, the residual glucose in the supernatant was measured using a YSI 2700 analyzer (Transition Technologies), and the feed rate adjusted to maintain a suitable glucose concentration throughout the run. Once the feed rate

reached 0.6 ml/min, it was kept constant until the end of the run. Foaming was controlled using a 4% Antifoam A solution (Sigma Aldrich). Each run lasted 48 hours.

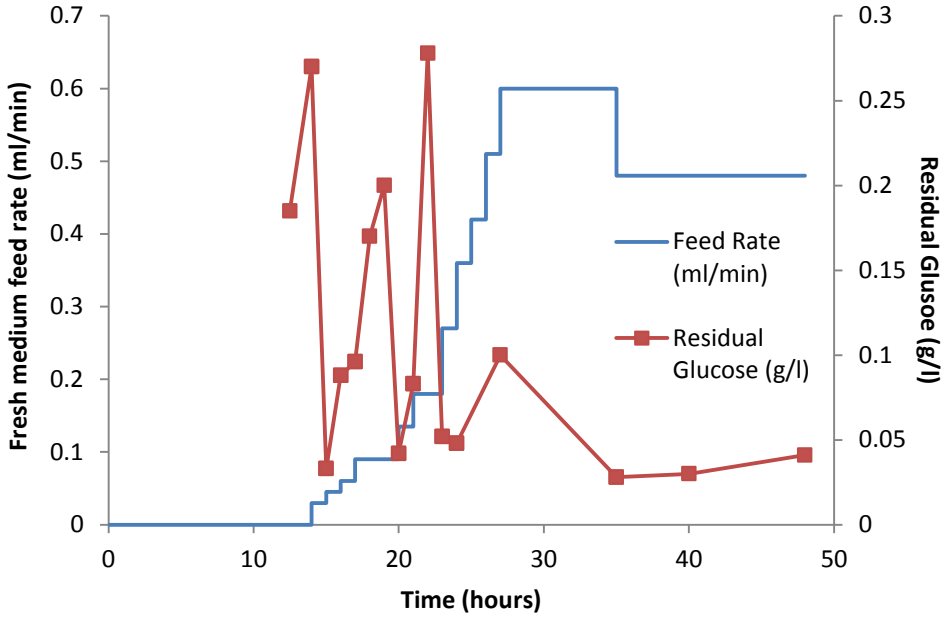

Figure 6.1. *Fresh medium feed rate and residual glucose concentrations for experimental run 19, selected randomly from the replicate runs to illustrate the feeding control methodology and monitoring during a typical run. The feed rate was adjusted every hour in response to the residual glucose concentration, in order to keep it under 0.5 g/l at non-growth limiting concentrations. All other runs exhibit similar profiles. Detection limit for the glucose analyzer: 0.018 g/l (all measured values are above this threshold).*

6.3.6 Cell Density, Enzymatic Activity, and Residual Element Analysis

Cell density was evaluated through OD_{600} measurements using a Biochrom Ultrospec 60 spectrophotometer (1 OD_{600} unit is equivalent to 0.367 g DCW/l, data not shown). Measurement of phytase activity in the supernatant was as previously described (Liu et al. 2011). Briefly, culture samples were centrifuged, and the supernatant used as the enzyme solution in a colorimetric assay using disodium nitrophenyl phosphate (NPP - Sigma Aldrich) as a substrate. Absorbance at 405nm of assayed samples was compared to standard curves to quantify volumetric enzymatic activity. 1 FTU is defined as the quantity of phytase that catalyzes the release of 1 µmol of phosphate from NPP per minute at 37˚C. Residual element concentrations in the growth media were measured by ICP analysis performed Dr. Nimal de Silva in the

Geochemistry Laboratory in the Department of Earth Sciences at the University of Ottawa (Ontario, Canada).

6.3.7 Modeling and Data Analysis

All experimental design, data analysis and model development was done using Design Expert 9 software (Stat-Ease Inc.). See section 6.4 for full experimental design.

6.4 Experimental Design

A 3-factor, 20-run central composite design was used to model extracellular enzyme activity as a function of the 3 variable parameters of interest, namely the concentration of KH_2PO_4, $MgSO_4 \cdot H_2O$ and the trace elements solution (PTM1). The high, low and central levels explored in this study, chosen as a range around the standard component concentrations, are listed in Table 6.1. It should be noted that the concentration of the calcium source ($CaSO_4 \cdot 2H_2O$) is kept proportional to the $MgSO_4 \cdot 7H_2O$ concentration, as the latter interferes with the chelating effects of the former, which is favorable for the growth of *P. pastoris*, as demonstrated in previous work (Zhang et al. 2007, Liu et al. 2011).

Table 6.1. Levels of each parameter studied in the central composite design

	Factor	Level				
		1.680 (+α)	**1**	**0**	**-1**	**-1.680 (-α)**
A	KH_2PO_4 (g/l)	49.33	42.90	33.45	24.00	17.56
B	$MgSO_4 \cdot 7H_2O$ (g/l)	6.63	5.71	4.36	3.00	2.08
	$CaSO_4 \cdot 2H_2O$ (g/l)	1.272	1.00	0.60	0.20	0.00
C	PTM1 (ml)	6.19	5.00	3.25	1.50	0.31

The 20 experimental runs, including the 6 center-point replicates, are listed in Table 6.2. Runs were performed in random order. All variables are coded according to:

$$X_i = \frac{x_i - x_o}{\Delta x_i} \qquad (1)$$

where X_i is the coded value of the independent variable, x_i the actual value of the variable, x_o the value of the variable at the center point, and Δx_i the step change. The data was fitted to a quadratic polynomial of the form:

$$Y = \beta_0 + \Sigma\beta_i X_i + \Sigma\beta_{ii} X_i^2 + \Sigma\beta_{ij} X_i X_j + \beta_{ijk} X_i X_j X_k \quad (2)$$

The model was analyzed at the 90% confidence level for significance and lack of fit. The predictive value of the reduced model was validated with experimental results at the predicted optimum media composition.

Table 6.2. Experimental runs as described by coded variables, final phytase activity in culture supernatants, and final biomass concentration

Experimental	Factor			Final phytase activity (FTU/ml)	Final Biomass Concentration (OD_{600})
	A	B	C		
1	-1	-1	-1	780.83	207
2	-1	-1	1	424.54	304
3	-1	1	-1	1342.16	239
4	-1	1	1	669.51	107
5	1	-1	-1	694.12	167
6	1	-1	1	249.54	214
7	1	1	-1	930.25	194
8	1	1	1	1203.43	143
9	-α	0	0	1244.91	203
10	α	0	0	322.58	247
11	0	-α	0	528.87	117
12	0	α	0	669.51	328
13	0	0	-α	1399.88	488
14	0	0	α	1475.96	378
15	0	0	0	1376.51	292
16	0	0	0	803.90	296
17	0	0	0	1458.13	286
18	0	0	0	1201.28	292
19	0	0	0	1372.92	276
20	0	0	0	808.64	264

6.5 Results and Discussion

6.5.1 Cell Growth

The growth curves for all experimental runs are shown in Figure 6.2, as represented by the OD_{600} of periodic samples. For the first 20h of each run, the growth profiles are quite similar, but they diverge in the second half the cultivation. The main objective of the present work is to optimize enzyme production per unit volume of the cultivation broth, which is a parameter only partially correlated to cell density, and more directly associated with the metabolic response of the cells when grown in different growth media. The growth curves are included here, however, to illustrate the considerable effect that variations in medium composition can have on process yields, and to show that higher enzyme yields are not necessarily the result of higher cell densities. The good reproducibility of the process further lends credence to these observations. It should also be noted that lower maximum cell densities are not associated with nutrient depletion, as confirmed by the ICP analysis results shown in section 6.5.3, and are strictly associated with the effects of changing the concentration of the different media components.

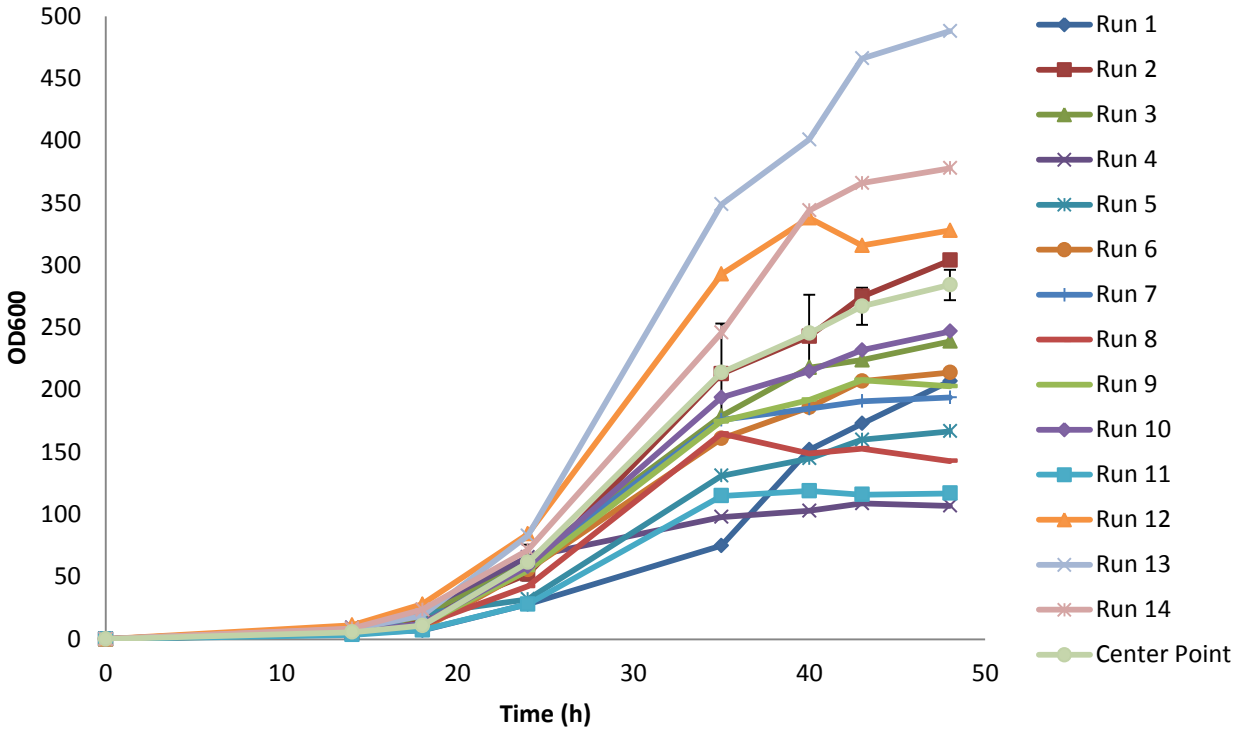

Figure 6.2. *Cell density profiles, as measured by OD_{600} for the 20 experimental 48-hour fed-batch runs. Center Point profile represents the average values of all 6 center point replicates. Error bars indicate standard deviation of cell density measurements for the replicate runs.*

Should the concentration of biomass be of particular interest, as opposed to recombinant product yields, an optimization protocol similar to the one described here could be performed with the organism of interest.

6.5.2 Enzymatic Activity Profiles

The enzymatic activity profiles for all experimental runs are found in Figure 6.3. For the first 20h of each run, the profiles are quite similar, but diverge in the second half of the cultivation. Given the number of runs, the information that can be directly extracted from this figure is limited, but it is included here to once again illustrate the variability of process parameters produced by modifying the composition of the growth media. It is also interesting to note that the results shown here, in conjunction with those in Figure 6.2, indicate that the enzyme activity/cell density ratios are not constant, and that the final cell concentration does not directly correlate to the enzymatic activity of the culture supernatant.

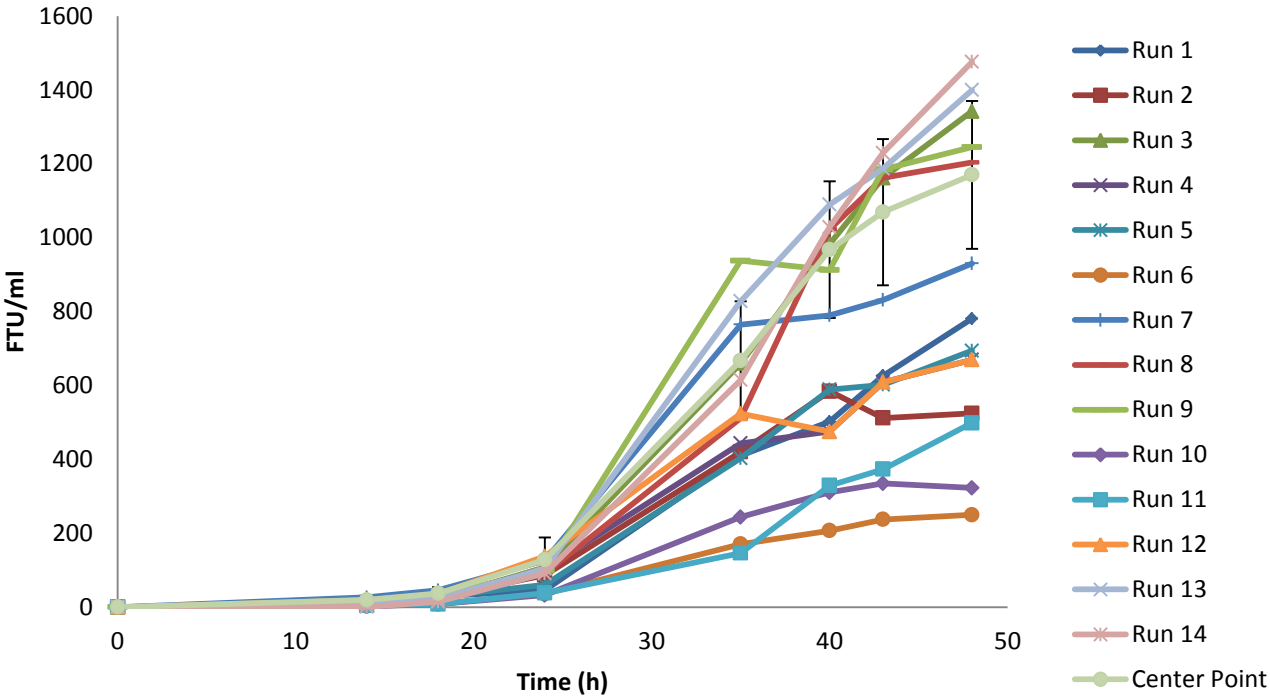

Figure 6.3. *Phytase activity profiles for the supernatants of the 20 experimental 48-hour fed-batch cultivations. Center point profile represents the average values of all 6 center point replicates. Error bars indicate standard deviation of cell density measurements for the replicate runs.*

Although the activity profiles for most runs begin to plateau before the end of the 48 hour run, some runs do not exhibit signs of slowing enzyme production within the time frame of these experiments, and greater enzymatic activities may be achieved by increasing their length. The final enzymatic activity of the culture supernatant is used as the response variable to build the model described and analyzed in section 6.5.4.

6.5.3 ICP analysis of residual elements

Depletion of trace salt elements in the broth can lead to lower final biomass concentrations and enzyme yields (Zhang et al. 2007). To confirm that the variability of the observed results was not the result of such depletion, the residual elements in the final culture broth of each experimental run were quantified by ICP. The results for runs 1, 7, 9 and 13, as well as those of the center point runs, are displayed in Figure 6.4. These runs were selected simply because they had the lowest residual concentrations of most elements, and therefore those of greatest relevance to this concern. Residuals of each element in other runs were of the same order of magnitude as the ones shown, but they are not reproduced here simply for clarity.

Results in Figure 6.4 indicate that no element was completely depleted during any run, and that variations in cell density or enzyme concentrations are therefore not the result of a lack of a limiting component. Other experimental work, as well as the statistical analysis described in section 6.5.4, suggest that even at the higher concentrations studied, the level of trace salts has no significant impact on enzyme yields.

6.5.4 Model of Enzymatic Activity

The data shown in Table 6.2 was fitted to a quadratic model (equation 2) to yield the full coded model (equation 3), where the response variable Y is the predicted enzymatic activity in the culture supernatant in FTU/ml, and A, B and C represent the concentrations of the components of interest. The full coded and unreduced model takes the form:

$$Y = 1174.74 - 201.44A + 163.53B - 0.83C + 50.11AB + 107.13AC + 47.9BC - 166.10A^2 - 231.35B^2 + 65.186C^2 \quad (3)$$

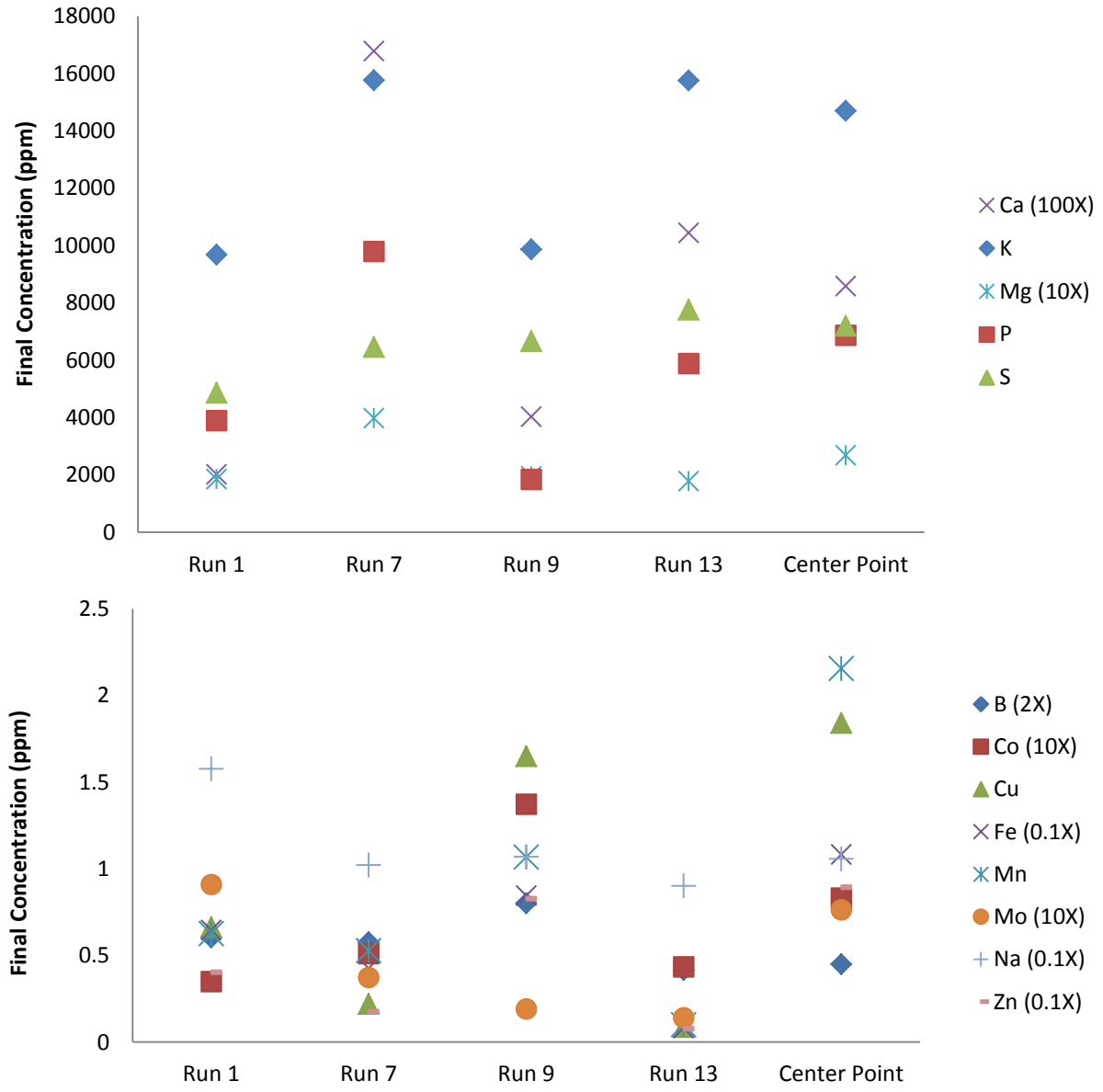

Figure 6.4. *Residual elements (top) and trace metals (bottom) in the culture broth at the end of selected experimental runs. Bracketed multipliers are included next to elements whose values were adjusted to facilitate graphical representation. The runs displayed had the lowest residual concentrations of different media components. Other runs are not shown, but display similar results. Center point represents the average readings of all replicates. Detection limits (ppm): Ca (0.061), K (0.023), Mg (0.022), P (0.021), S (0.067), B (0.001), Co (0.005), Cu (0.001), Fe (0.002), Mn (0.0001), Mo(0.003), Na (0.009), Zn (0.002).*

An ANOVA analysis was performed on the model, which is significant at the 90% confidence level and does not display significant lack of fit. The significance of each term of the model was evaluated, and the results are shown in Table 6.4. Based on these results, the terms associated

with parameter C, namely the trace salts concentration, as well as all the interaction terms are deemed insignificant.

Table 6.3. Evaluation of the significance of model terms. Significant terms (p < 0.1) are in bold.

Model Term	F value	p-value
Overall Model	**3.05**	**0.0488**
A - KH_2PO_4	**6.69**	**0.0271**
B - $MgSO_4$/$CaSO_4$	**4.41**	**0.0621**
C - PTM1	1.125E-4	0.9917
AB	0.24	0.6330
AC	1.11	0.3171
BC	0.22	0.6479
A^2	**4.80**	**0.0532**
B^2	**9.32**	**0.0122**
C^2	0.74	0.4100

The final uncoded predictive enzyme activity model is shown as equation 4, and represented graphically in Figure 6.5. This reduced model does not display a significant lack of fit.

$$\frac{FTU}{ml} = -3203.45 + 107.97[KH_2PO_4] + 1248.92[MgSO_4] - 1.93[KH_2PO_4]^2 - 129.53[MgSO_4]^2 \quad (4)$$

Based on this model, and as seen in Figure 6.5, the predicted optimal concentrations of the media components of interest are 27.9 g/l of KH_2PO_4 and 4.8 g/l of $MgSO_4$ (with a corresponding $CaSO_4$ concentration of 0.7 g/l) which represent 45% and 16% reductions of these components compared to the standard media recipe respectively. Given that the concentration of the PTM1 trace salts is not deemed significant, the lower limit of 1.5 ml salts/L media, a 62.5% reduction compared to the original medium, is taken as optimal solely for economic reasons. This value can likely be reduced further, but such values would fall outside the range of this experimental design.

The model was validated by running four additional fed-batch cultivations using the same protocol as the experimental runs but using the predicted optimal medium composition. The average enzyme activity obtained fall within a 90% predicted interval of the optimal activity value, suggesting the model is predictively useful. The results of these validation runs are found

in Table 6.4, and the extracellular phytase activity profiles for cultures grown in standard and optimized BSM media are found in Figure 6.6.

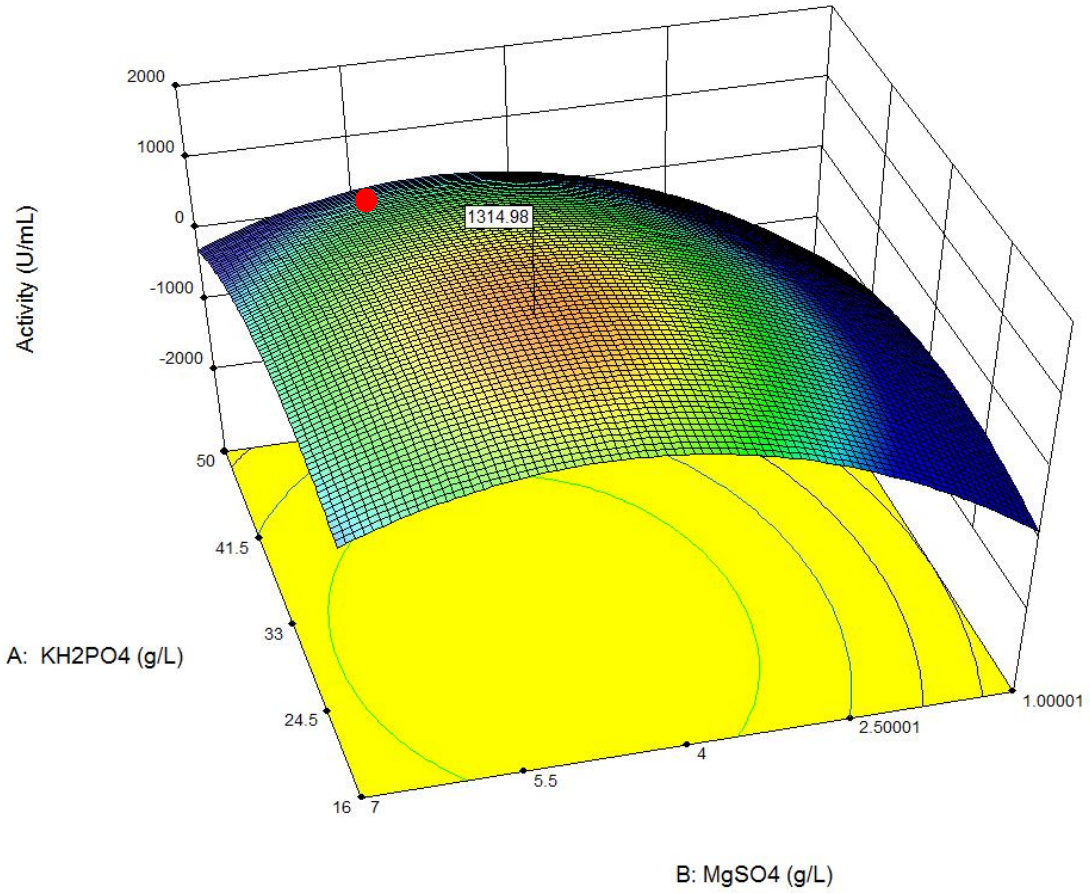

Figure 6.5. *Predicted enzymatic activity based on the model presented as equation 4. The optimum point occurs at 27.9 g/l of KH_2PO_4 and 4.8 g/l of $MgSO_4$. The original concentrations are demarcated by the red dot.*

Table 6.4. Additional fed-batch cultivations to validate the predictive model. All values in FTU/ml.

90% Predicted Interval for Activity (FTU/ml)		
90% PI Low		**90% PI High**
1038.52		1591.44
Experimental Runs for Validation		
Run	**FTU/ml**	**Average (FTU/ml ± std deviation)**
1	1802.3	1569.92 ± 159.03
2	1377.11	
3	1617.98	
4	1482.31	

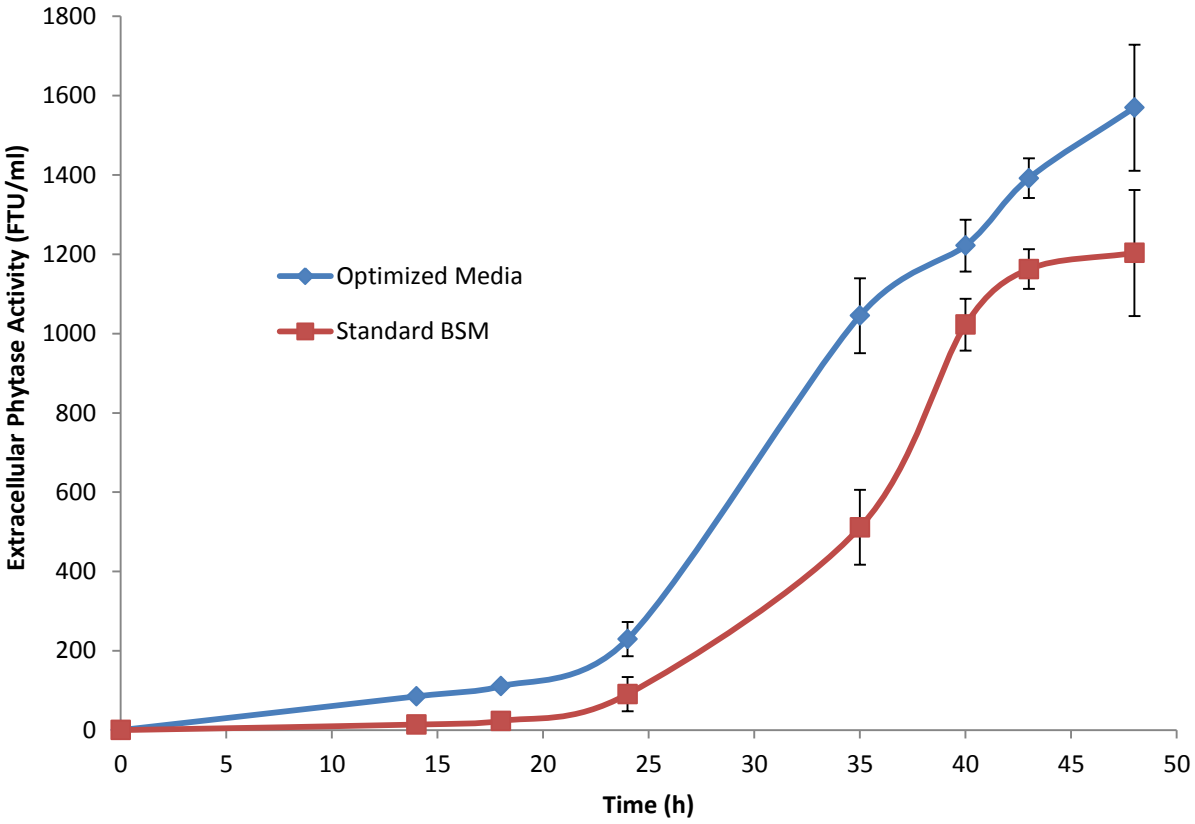

Figure 6.6. *Comparison of the extracellular phytase activity profiles of fed-batch cultures grown in standard BSM and optimized media. Standard BSM profile corresponds to that of experimental run 8, and the optimized media profile is the average of the four validation runs. Error bars represent the standard deviation of the validation runs' time points shown in Table 6.4.*

6.5.5 Discussion

The two significant factors in the present optimization procedure are the concentrations of potassium cations (related to the concentration of KH_2PO_4), and the linked concentrations of $CaSO_4$ and $MgSO_4$ hydrates, each of which is discussed in turn.

Potassium ions are important electrolytes involved in osmoregulation, macromolecule charge balancing and the regulation of ion uptake, and thus play key roles in yeast growth and fermentation under aerobic conditions. High concentrations of K^+, however, can limit the uptake of essential divalent cations, notably Ca^{2+} and Mg^{2+}, as well as interfere with the secretion of recombinant products (Jones and Greenfield 1994). The optimized medium contains a much lower concentration of K^+ than the standard composition, which suggests that this low level strikes the appropriate balance between normal growth-associated metabolic steps involving potassium while maximizing the yields of secreted enzyme. These findings are compatible with

those of previous work involving yeast fermentation using glycerol as a carbon source (Liu et al. 2011).

The concentrations of Mg^{2+} and Ca^{2+} were linked in the present experimental protocol because although they both play key metabolic roles, they also act as competitors for binding to ATP, and the chelating effect of Ca^{2+} can interfere with both yeast growth, and with phytase activity in particular (Fugthong et al. 2010). A proportional relation between the two species of divalent ions that minimized the chelating effects of calcium was determined based on previous work (Liu et al. 2011) and the two compounds were therefore linked here as a single factor. The optimal concentration obtained likely represents the non-limiting threshold of Mg^{2+}-related cofactor activities necessary for cell growth and enzyme production, while minimizing the chelating effects of calcium ions.

The third factor, namely the concentration of trace metals in the culture broth, was deemed not to have a statistically significant effect on enzyme production, and as confirmed by the quantification of residual elements, is, even at the lowest levels studied in this work, not a limiting factor during cultivation. The lower level is therefore deemed optimal for this system for economic reasons, but also for reasons related to process safety and waste treatment, as the preparation of this trace salts solution is time-consuming, and it contains compounds hazardous to health and to the environment in large quantities.

Although the proposed explanations for the variations in enzyme yields are compatible with work previously done in this field, and the results suggest that this optimal medium composition may very well be widely applicable for enzyme production in *Pichia pastoris* systems, it would be interesting, in future work, to verify the applicability of this approach to different systems, as different media components may have very product-specific effects.

The enzyme yields obtained using the optimized media, when using the standard deviation of the replicated validation runs as an approximation of true standard deviation, as shown in Figure 6.5, are significantly higher than those obtained using standard media, which corresponds to experimental run 8. The differences are not significantly different, however, if using the standard deviation of the experimental center point replicate runs as a measure of the true standard deviation. This suggests that the optimized media either has a performance equivalent to the standard one, but at a significantly lower cost due to the lower concentrations of the studied components as well as a reduction in the associated processing and treatment costs, or

significantly outperforms it. Additional experimental runs would be necessary to further characterize the performance of this optimized system in a variety of processes, and determine whether the results are more widely applicable.

6.6 Conclusions

Based on a central composite design, the composition of BSM media was optimized for phytase production in fed-batch cultivations of *Pichia pastoris* using glucose as the sole carbon source. The optimized media contains reduced concentrations of KH_2PO_4 and $MgSO_4 \cdot 7H_2O$ (27.9 g/l and 4.8 g/l respectively) compared to the standard media. The modified media either provides the same enzyme yield as the standard media, but at a significantly lower production and processing cost, or outperforms it. The predictive enzymatic activity model is significant at the 90% confidence level and validated experimentally.

6.7 References

Bawa Z, Routledge SJ, Jamshad M, Clare M, Sarkar D, Dickerson I, Ganzlin M, Poyner DR, Bill RM, Functional recombinant protein is present in the pre-induction phases of *Pichia pastoris* cultures when grown in bioreactors, but not shake-flasks, Microb. Cell Fact. 2014; 13, 17.

Fugthong A, Boonyapakron K, Sornlek W, Tanapongpipat S, Eurwilaichitr L, Pootanakit K, Biochemical characterization and in vitro digestibility assay of *Eupenicillium parvum* (BCC17694) phytase expressed in *Pichia pastoris*, Prot. Express. Purif. 2010; 70, 60-7.

Gao MJ, Shi ZP, Process Control and Optimization for Heterologous Protein Production by Methylotrophic *Pichia pastoris*. Chinese J. Chem. Eng. 2013; 21, 216-26.

Haraldsson AK, Veide J, Andlid T, Alminger ML, Sandberg AS, Degradation of Phytate by High-Phytase *Saccharomyces cerevisiae* Strains during Simulated Gastrointestinal Digestion, J. Agricult. Food Chem. 2005; 53, 5438-44.

Jones RP, Greenfield PF, A review of yeast ionic nutrition: growth fermentative requirements, Proc. Biochem. 1994; 4, 48-59.

Li PZ, Anumanthan A, Gao XG, Ilangovan K, Suzara VV, Duzgunes N, Renugopalakrishnan V, Expression of recombinant proteins in *Pichia pastoris*, Appl. Biochem. Biotechnol. 2007; 142, 105-24.

Liu M, Potvin G, Gan Y, Huang Z, Zhang Z, Medium Optimization for the Production of Phytase by Recombinant *Pichia pastoris* Grown on Glycerol, Int. J. Chem. Reactor Eng. 2011; 9, A86.

Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM, Heterologous protein production using the *Pichia pastoris* expression system, Yeast 2005; 22, 249-70.

Potvin G, Ahmad A, Zhang Z, Bioprocess engineering aspects of heterologous protein production in *Pichia pastoris*: A review, Biochem. Eng. J. 2012; 64, 91-105.

Vohra A, Rastogi SK, Satyanarayana T, Amelioration in growth and phosphorus assimilation of poultry birds using cell-bound phytase of *Pichia anomala*, World J. Microbiol. Biotechnol. 2006; 22, 553-8.

Wodzinski RJ, Ullah AHJ, Phytase, Adv. Appl. Microbiol. 1996; 42, 263-302.

Zhang W, Inan M, Meagher MM, Rational Design and Optimization of Fed-Batch and Continuous Fermentations, in: J.M. Cregg, Pichia Protocols, second ed., Humana Press, New Jersey, 2007, pp. 43-65.

Chapter 7 – Screening of Alternative Carbon Sources for Recombinant Protein Production in *Pichia pastoris*

Authors: Potvin G, Defela A, Zhang Z

7.1 Abstract

Seventeen carbon sources were screened to identify those with the potential to support pGAP-regulated recombinant phytase production by *Pichia pastoris*. Of these, four, namely glucose, glycerol, fructose and ethanol, supported cell growth and enzyme production, and the performance of the latter two were analyzed. Ranges of acceptable residual carbon source concentrations, i.e. those at which no substrate-related growth inhibition occurred, were determined in batch experiments, and used to design fed-batch bioreactor-based processes. In fed-batch cultures, fructose supported higher biomass concentrations and equivalent extracellular enzyme activities than glucose. The same metrics for the cultures grown on ethanol were comparable to those of the cultures grown on glucose, but the fermentation time required to achieve them was significantly longer.

7.2 Introduction

Driven by a continuously growing demand for recombinant products in the biopharmaceutical, cosmetic, and food industries, the development and optimization of efficient and robust biological systems for recombinant protein production remains a highly active research area. The methylotrophic yeast *Pichia pastoris*, given its versatility, is one of the most widely used of these production platforms in industrial processes. This system is particularly valued for its ability to achieve high cell densities, for its strong promoters, most notably the tightly regulated methanol-inducible alcohol oxidase 1 promoter (pAOX1), and the constitutive glyceraldehyde-3-phosphate dehydrogenase promoter (pGAP), and its capacity for effective foreign protein secretion (Damasceno et al. 2012, Potvin et al. 2012, Ahmad et al. 2014). As with all eukaryotic organisms, the ability to perform post-translational modifications such as glycosylation and disulfide bond formation, allows for the production of biologically active proteins, including therapeutics for animal and human consumption.

Several methods are available to improve the economics of bioprocesses through increases in biomass production and product yields, which include, most commonly, upstream strain

engineering for improved growth and production characteristics, and the optimization of process parameters, be it the medium composition or operational considerations. As the size and variety of the available collection of production-ready microorganisms increase, so do the options at one's disposal when it comes to bioprocess design. One area of interest, and the subject of the present work, is the identification of novel carbon sources that can sustain cell growth and protein production at levels that are comparable to, or that exceed, those of standard systems. The successful identification of these systems not only provides flexibility in terms of process design and protects industrial operations against fluctuations in substrate market prices, but would also present a rationale to investigate the use of unusual nutrient sources, such as available unused process or waste streams, for the incorporation of value-added secondary bioprocesses in these operations.

In this work, seventeen carbon sources were screened to determine their ability to sustain pGAP-regulated recombinant phytase production in *P. pastoris*. The effect of these carbon sources on cell growth and enzyme yields are characterized, first in flask experiments, and then in more industrially-relevant fed-batch bioreactor fermentation processes. Phytase catalyzes the hydrolysis of phytic acid, the main form of phosphorus storage in plants. It is commonly used as an animal feed additive to increase the bioavailability of phosphorus by releasing it from phytic acid, which is indigestible by monogastric crop animals. Phytase is used here as an easily assayable recombinant product, but the objective is to establish viable (i.e. economically justifiable and/or competitive with standard reagents) carbon source alternatives for the production of a wide range of recombinant proteins under the control of the pGAP promoter in *P. pastoris* systems.

7.3 Materials and Methods

7.3.1 Strain and strain upkeep

A strain of zeocin-resistant *P. pastoris* expressing and secreting a recombinant phytase under the control of the constitutive pGAP promoter was provided by Zell Technologies Inc. and used for all experimental runs. Yeast cultures were streaked on fresh YPD plates (10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose, 15 g/l agar) supplemented with zeocin every 2 weeks and incubated for 2 days at 30˚C.

7.3.2 Media

Experimental cultures were grown in Basal Salt Medium (BSM - 0.6 g/l $CaSO_4 \cdot 2H_2O$, 42.9 g/l KH_2PO_4, 5.17 g/l $(NH_4)_2SO_4$, 14.33 g/l K_2SO_4, 0.5 ml/l H_2SO_4, 5.71 g/l $MgSO_4 \cdot 7H_2O$ and 4 ml/l PTM1 Trace Salts) containing the stated concentration of the appropriate carbon source. Bioreactors start with 1L of BSM containing 5 g/l of the appropriate carbon source, with the feed solutions for fed-batch cultivation containing 600 g/l of the appropriate carbon source.

7.3.3 Flask Cultivations

All flask experiments were conducted in 250 ml Erlenmeyer flasks containing 100 ml of BSM medium supplemented with the appropriate carbon source. Each flask was inoculated with 1 ml of yeast grown in YPD to an OD_{600} of 0.8, and unless otherwise stated, incubated for 96 hours in a shaker at 30˚C and 250 rpm. Samples of each culture were periodically collected and analyzed for cell density and extracellular phytase activity.

7.3.4 Bioreactors

All bioreactor runs were performed with New Brunswick Scientific (NBS) Bioflo 110 3L (2L working volume) bioreactors equipped with dual stainless-steel Rushton impellers. Reactors were kept at 30˚C, sparged with 3-5 l/min air, and a dissolved oxygen (DO) concentration of 30% was maintained, with the agitation cascaded to the DO, measured online by NBS DO probes. pH was monitored by online NBS pH probes, and controlled by the addition of 30% ammonium hydroxide.

7.3.5 Fed-batch cultivations

Bioreactors were inoculated with 50 ml of yeast culture grown to an OD_{600} of 0.8 in BSM medium supplemented with the appropriate carbon source. Once the initial carbon source was consumed, a 60% carbon source medium was fed at an appropriate rate to maintain residual concentrations of the carbon source within the determined non-inhibitory concentration range, by periodically adjusting the feeding rate based on the measured residual carbon source concentrations. Culture samples were collected every hour, centrifuged, the residual glucose and

ethanol in the culture supernatant was measured using a YSI 2700 analyzer (Transition Technologies Inc.), the fructose using an Agilent Technologies 1100 series HPLC with ion exclusion column, and the feed rate adjusted to maintain a suitable concentration. Once the feed rate reached 0.6 ml/min, it was kept constant until the end of the run. Foaming was controlled using a 4% Antifoam A solution (Sigma Aldrich Inc.) using online foam sensors. Each run lasted 48 hours, with the exception of the ethanol runs which lasted 78 hours.

7.3.6 Sample Analysis

Cell density was evaluated through OD_{600} measurements using a Biochrom Ultrospec 60 spectrophotometer. One OD_{600} unit corresponds to 0.367 g DCW/l (data not shown). Measurement of phytase activity in the supernatant was as previously described (Liu et al. 2011). Briefly, culture samples were centrifuged, and the supernatant used as the enzyme solution used in a colorimetric assay using disodium nitrophenyl phosphate (NPP - Sigma Aldrich) as a substrate. Absorbance at 405 nm of assayed samples was converted to volumetric enzymatic activity using equation 1, determined from a standard curve drawn with enzyme samples of known activity.

$$\text{Phytase Concentration (FTU/ml)} = (A_{sample} - A_{blank})*D/0.01643 \qquad (1)$$

One FTU is defined as the quantity of phytase that catalyzes the release of 1 µmol of phosphate from NPP per minute at 37˚C.

7.4 Results and Discussions

7.4.1 Carbon Source Screening

Seventeen carbon sources, listed in Table 7.1, were screened to determine whether they could support pGAP-mediated recombinant phytase production in *P. pastoris*. Cultures were grown in flasks containing BSM supplemented with 2 g/l of the appropriate carbon source, and samples were periodically collected and analyzed to establish growth and production profiles. Of the carbon sources listed, four, namely glucose, glycerol, fructose and ethanol, led to appreciable cell growth and enzyme production. Glucose and glycerol are the most commonly used carbon

sources in industrial *P. pastoris* bioprocesses, and are included in this study to assess the competitiveness of the alternatives. The maximum cell densities reached on all other carbon sources except for the disacchardies was much lower than those obtained on the four best carbon sources (less than 50% of cell densities on glucose) and the data is therefore not analyzed further here. It should be noted that the disaccharides were included on the list of studied carbon sources as negative controls, as the strain of *P. pastoris* used in this work is unable to hydrolyze them into metabolizable components.

Table 7.1. Carbon sources (2 g/l) screened, organized by type of compound. Sources shown to significantly sustain growth and enzyme production are in bold.

Type of Compound	Compound
Monosaccharide	**Glucose**
	Fructose
	Galactose
Disaccharide	Sucrose
	Lactose
	Maltose
Alcohol	Methanol
	Ethanol
	Glycerol
	i-Propanol
	sec-Butanol
	tert-Butanol
	n-Propanol
	n-Butanol
Carboxylic Acids	Acetic Acid
	Propionic Acid
	Formic Acid

7.4.2 Biomass Production

The growth profiles for the flask cultures grown on the four viable carbon sources are shown in Figure 7.1. Based on these results, the lag phase and maximum growth rates of the cultures grown on fructose, glycerol and glucose are not significantly different, but the final cell densities of the glycerol and fructose cultures are slightly higher than those of the glucose cultures, but only at the 96 hour mark. Glucose and fructose are both metabolized through the glycolytic pathway, with fructose being phosphorylated to the fructose-1,6-biphosphate intermediate by phosphofructokinase. Fructose metabolism, therefore, requires minimal metabolic modification

in order to provide the requirements for growth, and the similarities in biomass production are expected. Cultures grown on ethanol appear to have slightly longer lag phase and significantly slower growth rate than the other three cultures, but sustained a higher final cell density.

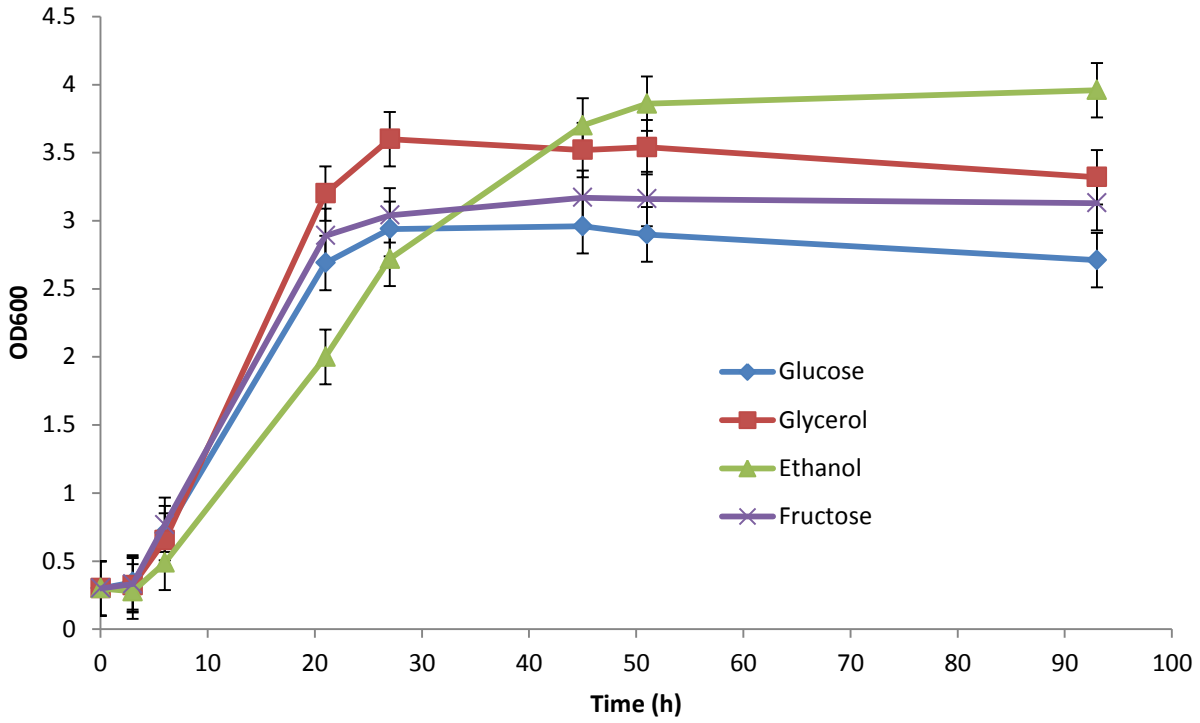

Figure 7.1. *Growth curves for P. pastoris grown in 2 g/l glucose, glycerol, fructose and ethanol for 96 hours. 1 unit of OD at 600 nm corresponds to a concentration of 0.367 g DCW/l. Error bars represent the standard deviation of triplicate runs. Data points correspond to analyzed samples.*

Although these observations are supported by the bioreactor data presented in subsequent sections, the experiment should be repeated with additional time points to better characterize the growth curve. These differences may be explained by the time required for the upregulation of the ethanol metabolic pathways, mainly those responsible for the conversion of ethanol to acetate, which is subsequently processed into acetyl-CoA before entering the citric acid cycle, as well as the slower energy-production kinetics of these pathways compared to those of their glycolytic counterparts (Inan and Meagher 2001). Although the initial cellular growth rate is reduced during ethanol utilization, results suggest that the conversion of ethanol to biomass is more efficient than the other carbon sources on a per-weight basis. This is likely due to the fewer branching points of the ethanol metabolic pathway, leading to fewer intermediates being involved in processes not directly related to cell growth and energy production than the other

carbon sources' pathways. Based on flask experiments, in terms of biomass accumulation, fructose appears to be a suitable replacement for glucose or glycerol should a cost-effective fructose source be available, while the suitability of ethanol in batch operations would depend on the cost-balance between the higher biomass yields and longer production times.

7.4.3 Enzyme Yields

Although biomass production is a parameter of interest, obtaining high recombinant protein yields is usually of greater importance in any product-oriented bioprocess. Although the pGAP promoter is constitutively regulated, its activity is modulated by a variety of transcription factors acting on different elements in the 5' UTR, and is affected by cultivation conditions including medium composition and available oxygen. Although a detailed discussion of pGAP regulation is outside the scope of this paper, the interested reader may refer to the recently published book by Seidler (Seidler 2013). Given the dependence of pGAP on cultivation conditions and parameters, the recombinant product yields and cell density are only partially correlated, and the phytase activity profiles must therefore be drawn for all new carbon sources and sets of operating conditions. These profiles are shown in Figure 7.2 to confirm that the carbon sources of interest also support high enzyme productivity.

Based on these results, there are no significant differences in the phytase activity of the culture broth between glucose, glycerol, ethanol and fructose at the end of the 96 hour cultivations (selected to determine the longer term enzyme degradation effects related to the carbon source used). The enzymatic activities of the glucose, glycerol and fructose cultures peak around the 24 hour mark, and the ethanol culture around 48h, which approximately corresponds to the times of the associated cell density peaks shown in Figure 7.1. The activity to cell concentration ratios for each culture are, however, not constant. Following the peak, broth activity seems to decrease at approximately the same rate for each culture, suggesting that this decline is associated with proteolytic degradation of the enzymes by constitutively expressed proteases.

Assuming that these results are representative of batch processes, glucose, glycerol and fructose are interchangeable and better than ethanol in terms of product yields for shorter cultivations, but for longer processes ethanol and fructose, if available at economically competitive prices, are suitable alternatives.

7.4.4 Inhibitory Concentrations

Most industrial bioprocesses are designed around fed-batch or continuous fermentation processes in which the biomass production, enzyme yields and metabolite profiles can vary significantly when compared to those of their batch counterparts (Bawa et al. 2014). In most processes, high residual carbon source concentrations are associated with substrate-related growth inhibition, which may impact recombinant product yields, and substrate concentrations must therefore be monitored and kept within an optimal range. To determine the optimal concentration ranges, i.e. those in which enzyme yields will be maximized while avoiding negative impacts on growth, flask cultures were grown on different concentrations of ethanol and fructose, and the relevant parameters were measured. Results are shown in Figure 7.3.

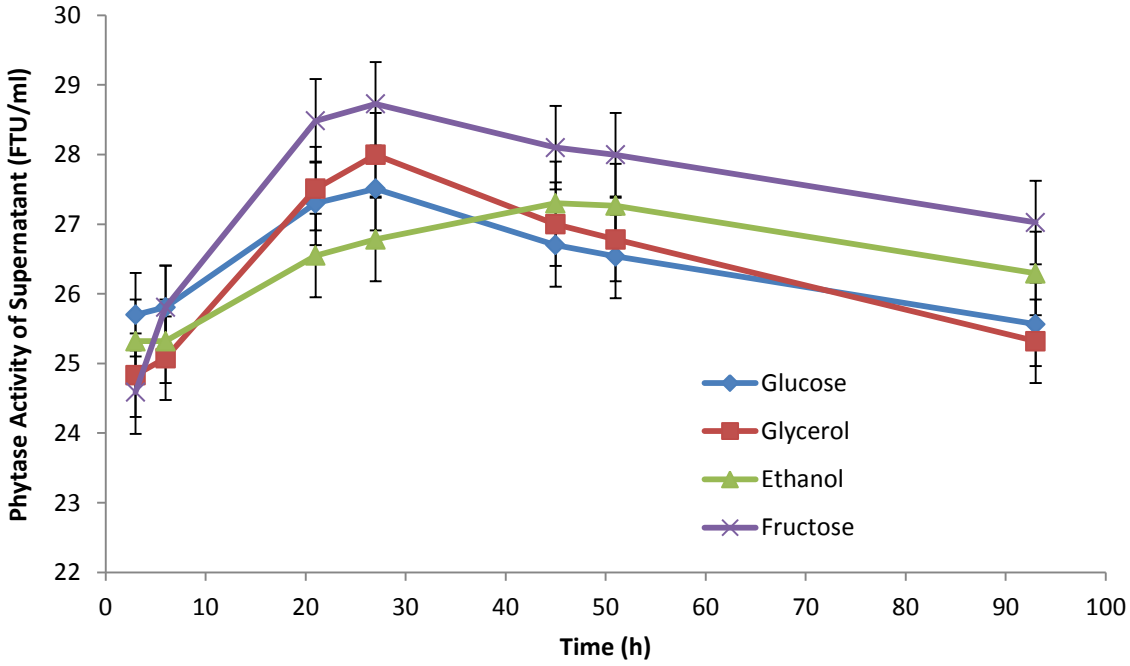

Figure 7.2. *Enzymatic activity profiles for flask cultures of P. pastoris grown on 2 g/l glucose, glycerol, fructose or ethanol for 96h. Data points correspond to analyzed samples. Lines are theoretical and fitted to collected data solely to emphasize the activity profiles. Error bars represent the standard deviation of triplicate runs.*

For cultures grown on ethanol, the highest initial exponential growth rates are obtained for the cultures grown on concentrations between 1 g/l and 5 g/l, with no significant differences in the duration of the lag phase (Fig. 7.3A). Concentrations lower than 0.5 g/l deplete too quickly to support growth at the maximum growth rate, and some growth inhibition is observed at concentrations of 10 g/l. In terms of enzymatic activity profiles (Fig. 7.3C), no significant

difference is observed between the different cultures after 24h, and the higher production observed for the cultures grown on 5 g/l and 10 g/l of the substrate is likely only associated to growth, with the lower concentrations of carbon source being depleted early in the process (data not shown), which is not a concern in fed-batch processes in which the residual concentration can be maintained accurately.

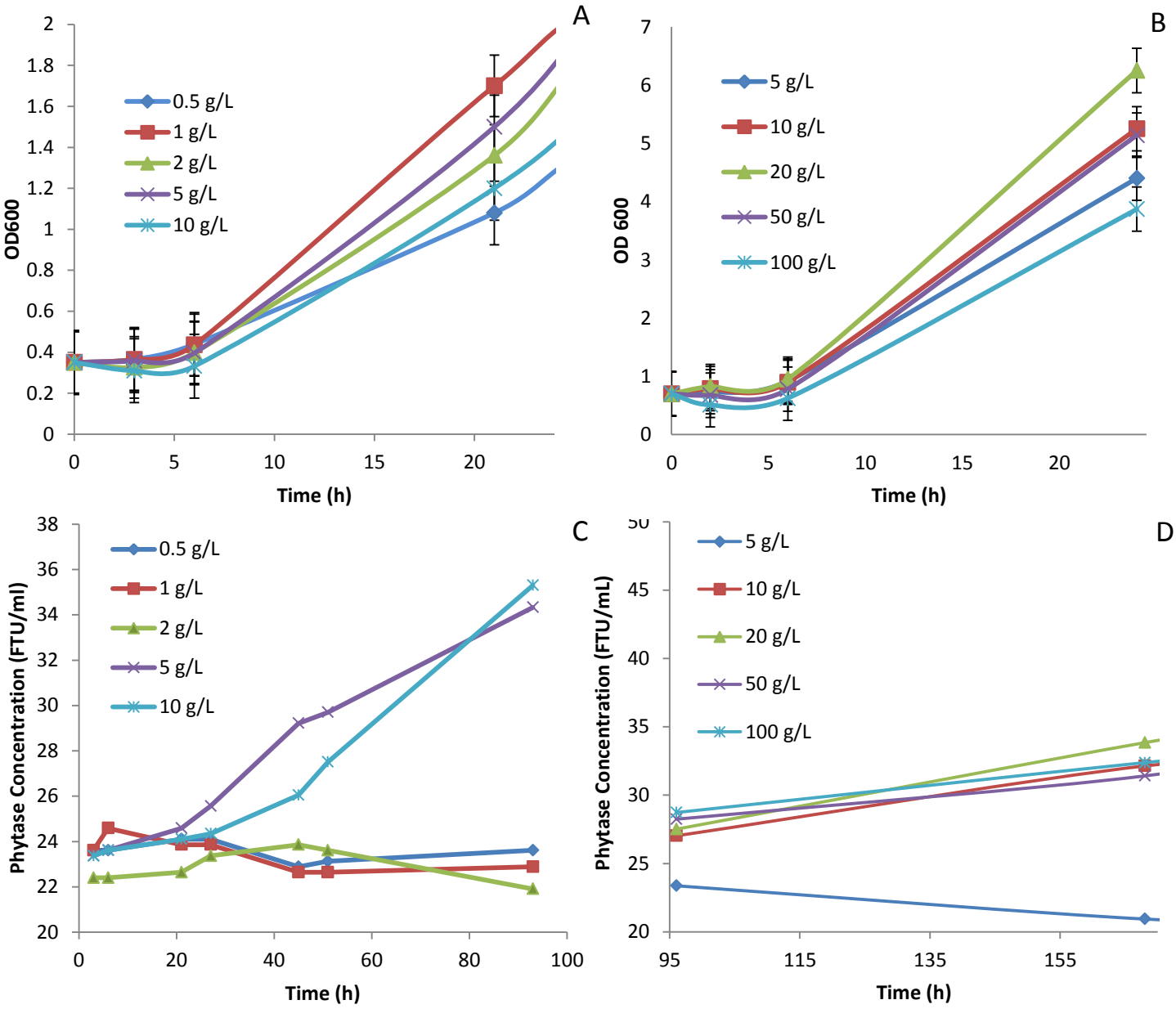

Figure 7.3. *Growth and extracellular enzymatic activity profiles of P. pastoris flask cultures grown on different concentrations of ethanol (A and C) and fructose (B and D) respectively. 1 unit of OD_{600} is equivalent to 0.367 g DCW/l. Error bars represent the standard deviation of triplicate runs. Enzyme profile data has not been replicated, and only used to confirm enzyme production at concentrations sufficient to support cell growth over the length of the runs.*

Based on the growth and enzymatic activity profiles shown in Figures 7.3B and 7.3D respectively, *P. pastoris* has a much higher substrate tolerance to fructose than to ethanol, and indeed to glucose, which is itself usually kept below 1 g/l to avoid substrate inhibition associated with the repression of several metabolic pathways in yeast (Papagianni et al. 2007).

The maximum initial growth rate and enzyme production rates are obtained for the cultures with fructose concentrations between 10 g/l and 20 g/l. The 5 g/l culture does not have sufficient fructose to sustain the growth and enzyme production of the culture for the duration of the experiment.

Although it is difficult to extract process characteristics from batch experiments that will subsequently be applicable to fed-batch or continuous processes, the data shown in Figure 7.3, especially at the beginning of each experiment when conditions are relatively constant, is useful in selecting process parameters. For subsequent fed-batch bioreactor cultivations, based on the available data, the residual concentrations of ethanol, fructose, and glucose were maintained between 1-5 g/l, 10-20 g/l and 0.5-1 g/l respectively.

7.4.5 Fed-batch bioreactor fermentation

Based on the data collected during flask experiments, the performance of fed-batch bioreactor systems using ethanol or fructose as sole carbon sources was evaluated and compared to that of cultures grown on glucose. The results are shown in Figure 7.4. The fructose and glucose cultures lasted 60 hours, while the ethanol culture was allowed to last 78 hours, which is the time necessary for a volume of feed solution equivalent to that of the glucose cultures to be added into the reactor, while maintaining the residual ethanol concentration within the appropriate range.

Based on the results in Figure 7.4, cultures grown on fructose produce significantly more biomass than those grown on glucose, but equivalent quantities of enzyme in terms of activity. In glucose cultures, residual concentrations are kept below 1 g/l, and a significant portion of this glucose is directed to pathways not directly associated with growth, with the rest processed in the glycolysis/citric acid cycle pathways. Given the good tolerance of the cultures to fructose, the residual fructose concentration in the cultivation broth was allowed to reach values of up to 20 g/l. This high concentration gradient and nutrient abundance may explain the higher biomass

yields as a greater concentration of carbon source is available to the growth associated pathways, without compromising any other metabolic process necessary for cell upkeep.

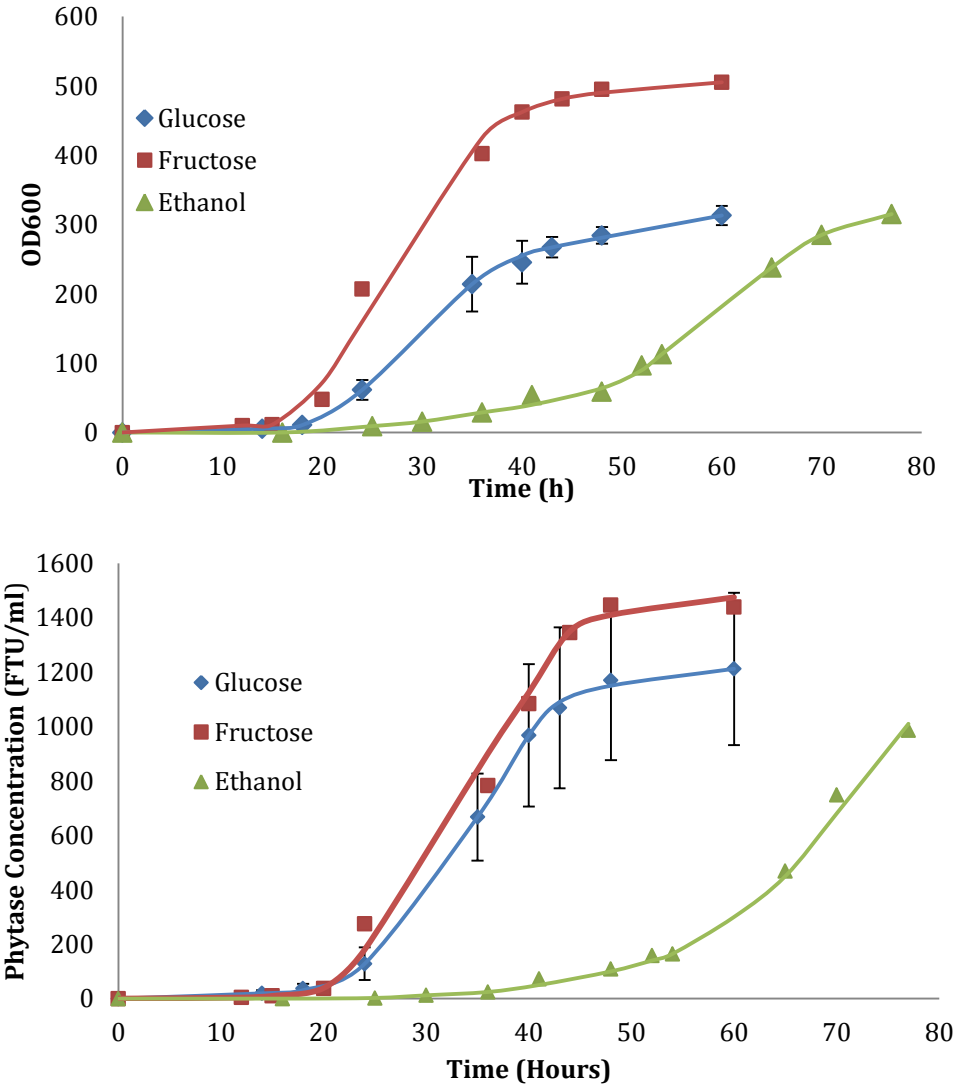

Figure 7.4. *Growth (top) and extracellular enzymatic activity (bottom) profiles of P. pastoris fed-batch bioreactor cultures grown on glucose, fructose and ethanol as sole carbon sources, with the feed rate of a 60% feed solution adjusted to maintain residual substrate concentrations between 0.5-1 g/l, 10-20 g/l and 1-5 g/l respectively. One unit of OD corresponds to 0.367 g DCW/l. Glucose values represent the average of 6 experimental runs, with the error bars corresponding to the standard deviation of each point. Single runs on fructose and ethanol were performed.*

The large error bars associated with the enzymatic activity profiles are due to the very high sensitivity of the activity assay to variations in media composition at a given point, and based on the grouping of the results into two high-reproducibility if separated groups (two runs very close

to 800 FTU/ml, and four runs very close to 1400 FTU/ml, not explicitly shown), to two distinct metabolic steady-states achieved during growth as a result of small variations in operating conditions. Cultures grown on ethanol have similar biomass and enzyme yields to those of the glucose cultures, but require significantly more time to reach those levels, which again is likely due to the slower kinetics of ethanol metabolic pathways and a much longer lag phase associated with a reorganization of the pathways associated with growth.

7.4.6 Comment on process viability

Based on experimental results, both ethanol and fructose may prove to be viable carbon source alternatives to glucose for *P. pastoris* cultures, but only under certain conditions. Fructose could be preferred to glucose in processes in which yeast biomass production is the primary objective, with the added benefit of less stringent concentration monitoring and control, but the market price of fructose, though comparable, is typically higher than that of glucose, making the latter the more common candidate for bioprocesses designed around the production of recombinant proteins. Although ethanol does produce yields comparable to those of glucose, the longer time required to achieve these yields may make these processes unattractive to most operations, at least as a primary design consideration. It should be noted that the results listed here are limited to the production of enzymes using the pGAP promoter, and given the different metabolic pathways involved in the use of these different carbon sources, the enzyme profiles obtained may not be representative of those of other systems.

These process considerations are based on the assumption that the carbon sources investigated would serve as primary high-purity carbon sources. Since fructose is extensively used in food bioprocessing, due in large part to its very high solubility, and ethanol is a very common industrial solvent and that process waste streams rich in ethanol are common in several processes, it would be interesting to investigate the use of these by-product or solvent streams as feed solutions for bioprocesses. The work described here suggests that the carbon sources can support cell growth and enzyme production, and so the use of such waste streams to feed secondary bioprocesses may provide opportunities to add value to industrial operations, reduce waste treatment costs, and have the potential to improve the environmental sustainability of some processes.

7.5 Conclusions

Out of the 17 alternative carbon sources screened in this study, only glucose, glycerol, ethanol and fructose supported cell growth and pGAP-mediated enzyme production in *P. pastoris*, and the performance of the latter two was characterized. Ranges of carbon source concentrations that do not lead to significant substrate-related inhibition were determined. In fed-batch bioreactor-based cultivations, fructose produces significantly more biomass than glucose cultures, and equivalent extracellular enzyme activities. Biomass and enzyme yields of cultures grown on ethanol are comparable to those of the cultures grown on glucose, but require a significantly longer fermentation time due a longer lag phase and slower growth rate.

7.6 References

Ahmad M, Hirz M, Pichler H, Schwab H. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production, Appl. Microbiol. Biotechnol. 2014; 98, 5301-17.

Bawa Z, Routledge SJ, Jamshad M, Clare M, Sarkar D, Dickerson I, Ganzlin M, Poyner DR, Bill RM, Functional recombinant protein is present in the pre-induction phases of *Pichia pastoris* cultures when grown in bioreactors, but not shake-flasks, Microb. Cell Fact. 2014; 13, 17.

Damasceno LM, Huang, CJ, Batt CA. Protein secretion in *Pichia pastoris* and advances in protein production, Appl. Microbiol. Biotechnol. 2012; 93, 31-9.

Inan M, Meagher MM, The effect of ethanol and acetate on protein expression in *Pichia pastoris*, J. Biosci. Bioeng. 2001; 4, 337-41.

Liu M, Potvin G, Gan Y, Huang Z, Zhang Z, Medium optimization for the production of phytase by recombinant *Pichia pastoris* grown on glycerol, Int. J. Chem. Reactor Eng. 2011; 9, A86.

Papagianni M, Boonpooh Y, Mattey M, Kristiansen B, Substrate inhibition kinetics of *Saccharomyces cerevisiae* in fed-batch cultures operated at constant glucose and maltose concentration levels, J. Ind. Microbiol. Biotechnol. 2007; 34, 301-9.

Potvin G, Ahmad A, Zhang Z. Bioprocess engineering aspects of heterologous protein production in *Pichia pastoris*: A review, Biochem. Eng. J. 2012; 64, 91-105.

Seidler NW, Basic biology of GAPDH, Adv. Exp. Med. Biol. 2013; 985, 1-36.

Chapter 8 – Conclusions and Recommendations

The work described in this thesis explored several aspects of bioprocessing, both on the upstream side, concerned with the development of novel recombinant protein expression platforms or the isolation of novel genes with products possessing characteristics of interest, and on the downstream side, through the improvement of fermentation-based bioprocesses. These two general approaches to bioprocess engineering are certainly not independent from one another; it is possible and indeed desirable, to design recombinant strains or biological systems with the final large-scale production process in mind, and to always keep in mind the constraints and characteristics imposed by the microorganisms of interest when designing and operating these processes. Although the work presented spans a relatively wide range of disciplines, they all share a similar objective, namely the improvement of the use of biological systems as production platforms for valuable, industrially-relevant products.

The use of microalgae as platforms for the production of recombinant proteins is a burgeoning field, and given the economic, logistical and environmental advantages of this technology, one that promises to continue to undergo a rapid development. Although the current work, focusing on the development of *Chlamydomonas reinhardtii* strains expressing enzymes of interest to the animal feed industry, did not produce strains able to produce significant product yields, several recombinant strains, having integrated gene constructs designed with novel approaches, were successfully developed. Although, at the time of writing, this is not an uncommon outcome of this type of work, several strategies are available that may increase the production of recombinant proteins in microalgae, including the inclusion of endogenous introns within the heterologous gene sequence, the fusion of the gene construct to a full or partial endogenous gene for co-expression, or simply the screening of a wider range of promoter or insertion loci. Although all of these strategies are logical next steps, this undertaking falls outside of the scope of this doctoral work, and may be pursued by future researchers, given the research capacity and expertise of our group developed during this work.

Given that the computational tools necessary for the *de novo* creation of customized recombinant proteins with specific enzymatic functions are not yet available, the improvement of protein-based processes usually involves either the modification of existing enzymes using protein engineering approaches, or a bio-prospecting approach based on the identification of novel gene candidates through the selective screening of biomass of interest. This latter approach

was used to identify novel cellulase genes from the microflora of the digestive tract of a Canadian beaver. One of the important bottlenecks in the cellulosic bioethanol process is the effective digestion of cellulose substrates, and the identification of enzymes that could potentially alleviate these difficulties is of considerable interest. The screening approach successfully identified and characterized a novel β-glucosidase gene from an isolated strain of *Bacillus thuringiensis*. Since only preliminary characterization was performed, and given that the results obtained were different than those of other published studies on this subject in terms of pH and temperature optimae, it would be interesting in future work to study the product more closely, characterize its method of action, and compare the resulting enzymatic activity model to those of similar enzymes. Such information would prove valuable to future protein engineering projects using cellulase enzymes.

In terms of process engineering, the continued improvement and optimization of bioprocesses to obtain increased product yields ensures that operations maintain their economic competitiveness. Two separate projects with this objective were performed. In the first, the composition of standard BSM was systematically optimized for the production of recombinant phytase with *Pichia pastoris*, and the optimized media produced significantly more enzyme than the standard composition, while also containing significantly lower amounts of medium components, which lowered the price of process inputs. Since the medium composition can have a significant effect on protein expression and characteristics, it would be interesting to investigate whether the improved performance could be replicated for the production of other pGAP-mediated genes in yeast systems, or to fine-tune the model using more reliable on-line glucose measuring methodologies.

The second process-related project was the screening of unconventional carbon sources for candidates that could sustain growth and enzyme production using the same *P. pastoris* strain. Growing the cultures on fructose produced significantly more biomass than the cultures grown on glucose or glycerol, and equivalent extracellular activities. Cultures grown on ethanol could produce equivalent biomass concentrations and enzyme activities to those produced by cultures grown on standard carbon sources, but required significantly more time to achieve this. Although neither of these carbon sources is likely to replace glucose or glycerol as primary substrates, the identification of alternatives provides flexibility in terms of process design. It would be interesting, for example, to investigate the feasibility of using fructose-rich waste streams,

common in the food processing industry, or ethanol-rich industrial effluents, as the carbon source for value-adding secondary bioprocesses, which would generate additional revenue and potentially reduce environmental remediation and/or waste treatment costs.

Bioprocessing is a dynamic field that increasingly inserts itself into the more traditional chemical engineering disciplines. As the available genetic and metabolic toolsets continue to expand, and as we continue to explore and harness the biodiversity of the systems at our disposal, the options made available in the use of bioprocesses will increasingly lead to breatkthroughs and accomplishments in terms of economic performance, environmental sustainability, and in the variety of products we can effectively manufacture.

Appendix A.1 – Full Gene Construct Sequences

The full sequences for all optimized genes and gene constructs are included here. All gene constructs would be sub-cloned to transformation or cloning vectors. Stock *E. coli* cultures containing each vector/contruct are preserved at -80˚C for future use. The PCR primers corresponding to the annealing sites highlighted in these sequences are listed in Appendix A.2.

C1 (Phytase from *Yersinia intermedia* optimized with method 1. Flanking and internal primer sites in red, restriction enzyme sites in blue, Factor Xa protease cleavage site in yellow, his-tag in green)

5′-
gaccgacatcatcagtttgaccccatggatactattacagtagatagtttacgtttatcagtattaacattaattt
taaatagttatgctattagtgcagctccagttgcaattcaaccaacaggttatactttagaacgtgttgtaatttt
atcacgtcatggtgttcgttcaccaacaaaacaaactcaattaatgaatgatgttacacctgatacttggccacaa
tggcctgttgctgcaggttacttaacaccacgtggtgcacaattagttacattaatgggtggtttctatggtgatt
atttccgttcacaaggtttattagctgcaggttgtccaactgatgctgttatttatgctcaagcagatgttgatca
acgtacacgtttaactggtcaagcattccttgatggtattgcaccaggttgtggtttaaaagtacattatcaagct
gatttaaaaaaagttgatccattatttcatccagttgatgctggtgtttgtaaattagattcaacacaaactcata
aagctgttgaagaacgtttaggtggtccattaagtgaattatcaaaacgttatgctaaaccatttgctcaaatggg
tgaaatttaaattttgctgcatctccttactgtaaatcattacaacaacaaggtaaaacatgtgattttgctaac
ttcgcagctaataaaattacagttaacaaaccaggtacaaaagtttcattatctggtccattagcattatcatcta
cattaggtgaaatttttttattacaaaattcacaagctatgcctgatgttgcttggcatcgtttaacaggtgaaga
taattggatttcattattatcattacataatgctcaatttgatttaatggcaaaaacaccttatattgctcgtcat
aaaggtacaccattattacaacaaattgaaactgctttagttcttcaacgtgatgctcaaggtcaaacattaccat
tatcacctcaaacaaaaattttattcttaggtggtcatgatacaaacattgctaatattgcaggtatgttaggtgc
taactggcaattaccacaacaacctgataatacaccacctggtggtggtttagttttcgaattatggcaaaaccca
gataatcatcaacgttatgtagctgttaaaatgttctatcaaacaatgggtcaattacgtaatgctgaaaaattag
atttaaaaaacaatccagctggtcgtgttcctgtagcaattgatggttgtgaaaatagtggtgatgataaactttg
tcaacttgatacattccaaaaaaaagtagcacaagctattgaacctgcttgtcatattattgaaggtcgtcatcat
catcatcatcattaatctagagtagttcgtatccgttaggctgg – 3′

C2 (Phytase from *Yersinia intermedia* optimized with method 2. Flanking and internal primer sites in red, restriction enzyme sites in blue, Factor Xa protease cleavage site in yellow, his-tag in green)

5′-
gaccgacatcatcagtttgaccccatggatactatcactgtagatagtttacgccttagcgtattaacacttatcct
aaatagttatgcaattagtgctgcacctgttgctatccaacccaccggttatactcttgaacgcgttgttattctta
gtcgtcatggtgttcgtagccctacaaaacaaactcaacttatgaatgatgttacacctgatacctggcctcaatgg
cctgtagctgctggataccttacaccccgcggtgctcaacttgttactcttatgggtggattctatggtgattactt
ccgtagtcaaggtcttctagctgctggttgccctactgatgcagttatttatgcacaagctgatgttgatcaacgca
cccgtcttaccggtcaagctttcctagatggaatcgctcctggttgtggattaaaagtacattatcaagcagatctt
aaaaaagttgatcctttatttcatcccgtagatgcaggtgtttgtaagcttgatagcactcaaacacataaggcagt
tgaagaacgcttgggtggtcctcttagtgaattaagtaaacgttatgcaaagccctttgctcaaatgggtgaaattt
taaattttgcagcttctccttactgtaaatcattacaacaacaaggtaaaacatgtgattttgctaactttgctgca
aataagattaccgttaacaagcctggtactaaagtaagcctaagtggacctttagctttatcatcaacacttggtga
aatttttctttgcaaaattcacaagcaatgcctgatgttgcttggcatagacttaccggagaagataattggatta
gccttcttagccttcacaatgcacaatttgatcttatggctaaaactccttatattgcacgtcataagggtactcct
cttttacaacaaattgaaactgctctagtactacaacgtgatgcacaaggtcaaactcttcctctttcacctcaaac

aaaaattttattcctaggtggacatgatactaacattgctaatattgcaggaatgcttggtgcaaactggcaacttc
ctcaacaacccgataatacacctcctggtggtggacttgtattcgaattgtggcaaaaccctgataatcatcaacgt
tatgtagcagttaaaatgttctatcaaactatgggtcaacttcgcaatgcagaaaaattggatttaaaaaacaatcc
tgcaggtagggtacctgttgctatcgatggttgtgaaaatagtggtgatgataaactatgtcaactagatacattcc
aaaagaaagtagcacaagcaattgaacctgcatgccatattattgaaggtcgtcatcatcatcatcatcattaatct
agagtagttcgtatccgttaggctgg - 3′

C3 (Phytase from *Citrobacter amalonaticus* optimized with method 1. Flanking and internal primer sites in red, restriction enzyme sites in blue, Factor Xa protease cleaving site in yellow, his-tag in green)

5′-
gaccgacatcatcagtttgaccccatgggaatacattacttttcgtttaattatgtttattttcatgtttggttc
tttcccattacaagctgaagttccagatgatatgaaattagaacgtgttgttattgtaagtcgtcatggtgtacgtg
caccaacaaaattcacaccattaatgcaagaaattacaccttactattggccacaatgggatgttccattaggttgg
ttaacagctcgtggtggtgaattagttacagaaatgggtcgttatcaacaaaaagtattaattgataacggtgtttt
agaaagtaatgtatgtccatcaccagaacaagttgcagttattgctgatacagatcaacgtactcgtaaaacaggtg
aagcatttttagctggttttgctccaggttgtaaaaataaagttcattatcaaaaagatcatgataaaaaagatcct
ctttttaatccagtaaaaatgggtgtttgtgcttttaatgtacaaaaaactcaagaagctattttaacacgtgctga
aggtaacattgaacgttacactcaacgttatgattctgcattccgtactttagaacaagttttaaatttctcacgtt
cagcagcatgtcgttcagcaagtcaatctggttgtacattaccaggtacattaccttcagaattaagagtttctgct
gatacagtttcattatctggtgcttggagtctttcttcaatgttaacagaaatttttttattacaagaagctcaagg
tatgccagaagttgcttggggtcgtattcatggtgaaaaagaatggacagctttattaagtttacataatgctcaat
ttgatcttttacaacgtactccagaagttgctcgtagtcgtgcaacaccattattagatttaattagtgaagcatta
gttagtaatggttcaacagaaaatcattacggtattaaattaccagtttcattattatttattgctggtcatgatac
aaatcttgcaaatttaagtggtgtatttgatcttaactggtctttacctggtcaaccagataatacacctcctggtg
gtgaattagttttcgaacgttggacacgtgttagtgataacactgattggattcaaatttcatttgtttttcaaact
cttcaacaaatgcgtaaatttaaacctttttcatcttcatctttaccaaacaaaattgttcttacattaccatcttg
tcaagataaaaatcctgaaggtatgtgtccattaaaacattttattgatattgttcaaacagcacgtattccacaat
gtgcagttatggctgatgtaaaccgtattgaaggtcgtcatcatcatcatcatcattaatctagagtagttcgtatc
cgttaggctgg - 3′

D1 protein (product of endogenous *psbA* gene for restoration of photosynthesis when inserted under control of *psbD* promoter in ppsbD transformation vector, flanking and internal primer sites in red, restriction enzyme sites in blue)

5′-
gaccgacatcatcagtttgaccccatgggacagcaattttagaacgtcgtgaaaattctagcctatgggctcgttt
ttgtgagtggatcacttcaactgaaaaccgtttatacatcggttggttcggtgtaatcatgatcccatgtcttctta
ctgcaacatcagtattcatcatcgctttcatcgctgctccgccagtagacatcgatggtatccgtgaaccagtttca
ggttctcttctttacggtaacaacatcattacaggtgctgtaatcccaacttctaacgcaatcggtcttcacttcta
cccaatttgggaagctgcttctcttgacgagtggttatacaacggtggtccttaccaacttatcgtttgtcacttcc
ttctaggtgtatactgctacatgggtcgtgagtgggaattatctttccgtttaggtatgcgtccttggatcgctgta
gcttactcagctccagtagctgcagcttcagctgtattcttagtttaccctatcggccaaggttcattctctgacgg
tatgcctttaggtatctctggtactttcaacttcatgatcgtattccaagcagaacacaacatccttatgcacccat
tccacatgttaggtgttgctggtgtattcggtggttcattattctcagctatgcacggttctttagttacttcatct
ttaatccgtgaaacaactgaaaacgaatcagctaacgaaggttaccgtttcggtcaagaagaagaaacttacaacat
tgtagctgctcatggttactttggtcgtctaatcttccaatacgcttctttcaacaactctcgttcattacacttct
tcttagctgcttggccggtaatcggtatttggttcactgctttaggtttatcaactatggcattcaacttaaacggt
ttcaacttcaaccaatcagtagtagactcacaaggtcgtgtactaaacacttgggcagacatcatcaaccgtgctaa
cttaggtatggaagtaatgcacgagcgtaacgctcacaacttccctcttgacttagcttcaactaactctagctcaa
acaactaatctagagtagttcgtatccgttaggctgg - 3′

X1 (Xylanase from *Actinomadura sp 114* optimized with method 1. Flanking and internal primer sites in red, restriction enzyme sites in blue, Factor Xa protease cleaving site in yellow, his-tag in green)

5′-
gaccgacatcatcagtttgaccccatgggattcaagctccagcacgtcctaaaagtcgtcgtcgtgttcgtgctcg
tcttcttgctagtggtgttagtgctgtagcaatggttgctgcaacaactgttcttccaggtacagctcatgcagcta
ttacaactaatcaaacaggttggcataatggttatttttactcattctggacagattcacaaggtacagtttcaatg
gaacttggttcaggtggtaattattcaacatcttggtataatacaggtaactttgttgctggtaaaggttggaatcc
aggtggtcgtcgtacagttaattattcagcttcttattcaccatctggtaattcatatcttacactttatggttgga
cacgtaatccacttgttgaatattacattgtagaatcatggggttcttggcgtccaacaggtacttataaaggtaca
attacttcagatggtggtacatatgatatttaccaaacaatgcgttataatgctccatcaattgaaggtattcgtac
atttccacaatattggtcagttcgtcaatcaaaacgtacatcaggtactattacatcaggtaatcattttgatgctt
gggcacgttatggtattaatcttggttcacatgattatatgattatggctacagaaggttatcaatcatctggttat
tcaaatgttacagtatggattgaaggtcgtcatcatcatcatcatcattgatctagagtagttcgtatccgttaggc
tgg - 3′

X2 (Xylanase from *Actinomadura sp 114* optimized with method 2. Flanking and internal primer sites in red, restriction enzyme sites in blue, Factor Xa protease cleaving site in yellow, his-tag in green)

5′-
gaccgacatcatcagtttgaccccatgggatgattcaagctccagctcgtccaaagtctcgtaggcgtgttcgtgc
taggcttcttgcttctggggtttctgctgttgctatggttgctgctactactgttcttccaggtacagctcatgctg
ctattacaacaaaccaaacaggttggcataacggttatttttattctttttggacagattctcaaggcactgtttct
atggaacttggttctggtggtaactattctacatcttggtataacactggtaactttgttgctggtaaaggttggaa
cccaggtggccgtaggactgttaactattctgcttcttattctccatctggtaactcttaccttacactttatggtt
ggactcgtaacccacttgttgaatattatattgttgaatcttggggttcctggcgtccaacaggtacttataaaggt
actattacatctgatggtggtacttatgatatttatcaaactatgcgttataacgctccaagtattgaaggtattcg
tacatttccacaatattggtctgttcgtcaaagtaaacgtacttctggtacaattacaagtggtaaccattttgatg
cttgggctcgttatggtattaaccttggttctcatgattatatgattatggctacagaaggttatcaatcttctggt
tatagtaacgttactgtttggattgaaggtcgtcatcatcatcatcatcattgatctagagtagttcgtatccgtta
ggctgg - 3′

X3 (Xylanase from *Alicyclobacillus* optimized with method 1. Flanking and internal primer sites in red, restriction enzyme sites in blue, Factor Xa protease cleaving site in yellow, his-tag in green)

5′ -
Gaccgacatcatcagtttgaccccatgggactgatacttatcgtaataatccttcattaagtgaacgttaccgtcc
atactttcgtattggtgctgctgttaatgctaaatcacttaatacacatcgtgatttattagttacacattttaaca
gtgttacagcagaaaacgaaatgaaatgggaagaaattcatccagaacaagatcgttacgaattcgcaaaagctgat
gcattagttatttttgctcgtgaacatggtatgtttgttcgtggtcatacattagtttggcatattcaaacaccagc
tgcagtttttcttgatgatttaggtcaaacagctacagctgctgttgttgaacgtcgattagaagaacatgttgcaa
cacttcttggtcgttaccataacgatatttacgattgggatgttgctaacgaagcagttgttgatgctggtacaggt
tttttacgtgatagtcgttggttacaaacacttggtgatgattacattgctaaagcttttcgtattgcacatcaagc
tgcaccagatgcacttttattctacaacgattacaatgaaactaaaccagataaatcagaacgtatttacaaacttg
ttgcaggtttattagatgaaggtgttccaattcatggtattggtatgcaaggtcattggatgttagatgatccagct
ttagatgaaattgaacgtgctattgatcgttacgcttcattaggtgttcatcttcatattacagaattagatgtttg
tgtttatggtaatgttcatggtacaggtggtcaatcacaagaagttttaccatacgatgatgaacttgctaaacgtt
tagctgaacgttatcgttctcgtttttcatcattacgtgctcgtaaagatgttatggaaagtgttacattctggggt
gttgctgatgatgatacatggcgtgataacttccctgttcgtggtcgtaaagattggccacttttatttgatgttaa
ccatggtccaaaacaagctttttggagtgttgttgaatttattgaaggtcgtcatcatcatcatcatcattgatcta
gagtagttcgtatccgttaggctgg - 3′

ppsbD transformation vector. Any sequences in frame inserted at NcoI/XbaI site will be expressed under the psbD promoter. Primer sites in red, restriction enzyme sites in blue, *psbD* promoter and 5' UTR in green, *psbD* 3' UTR in yellow, black flanking sequences target inserted sequence downstream of *psbA* gene by homologous recombination.

5′-
gaccgacatcatcagtttgaccgagctcataaagctttgtaaatactaaatgaataaactcacttctttttttatta
ccattagataataaaagatacaagcccgaagggcttgggggaattcgcacaaaaagcttacatcgaagtcacttaaa
cacatatcacatatacgctcctaaatctaagggtttataagcgtatatgttatggctaaacggtgtgtagctagcat
cagtgttgtataagttctacttaaatatccacactaaggaatgcgtgtcggtgcagaaagcatcgctactggtaaat
tattggatttcaccacagcttgggagggagctacaaatccaacaacaaagacaaaaatcccaataaacaacactgtc
acttgtgttacaacaaaacattttaaattttaaatttttttcctaattatatatttttacttgcaaaatttataaaaat
tttatgcatttttatatcataataataaaacctttattcatggtttataatataataattgtgatgactatgcacaa
agcagttctagtcccatatatataactatatataacccgtttaaagatttatttaaaaatatgtgtgtaaaaaatgc
ttatttttaattttatttttatataagttataatattaaatacacaatgattaaaattaaataataataaatttaacg
taacgatgagttgtttttttatttttggagatacacgcaccatggtctagaattttttatttttcatgatgtttatgtg
aatagcataaacatcgtttttatttttttatggtgtttaggttaaatacctaaacatcattttacatttttaaaatta
agttctaaagttatcttttgtttaaatttgcctgtgctttataaattacgatgtgccagaaaaataaaatcttagct
ttttattatagaatttatctttatgtattatatttttataagtaataaaagaaatagtaacatactaaagattaaagc
tgttcgaagaaaagcttgatctagcattaaattaattcgatctataagacaagcaaaactttcttcttcgttgacta
aaaccaaccattgtgtgaccactttattgtcaacaaaacaatttttttcatgaaggatttcttgtaacaaacaaaag
aaatcctttaacaaaataggatttaaatcagccatacttaacaccaaaagatctaagtcttggaaaagcattaaaag
tgctcctttagtattattaattggcaatttttgaaatgctttttttatagtaccataaagtattttcttggcttctt
ctgataaactataagataagcttaaattaaccaagaaagctagtcctagttgtggaggaggtaccgtagttcgtatc
cgttaggctgg - 3′

Appendix A.2 – Complete List of PCR Primers

All PCR primers used in the work described in this document are listed here in the 5'-3' orientation. All primers were designed using Primer Premier 5 software. 'F' and 'R' denote forward and reverse primers respectively.

Notation indicating relevant gene constructs:

C1 – *Yersinia intermedia* phytase – Optimization Method 1
C2 – *Yersinia intermedia* phytase – Optimization Method 2
C3 – *Citrobacter amalonaticus* phytase – Optimization Method 1
X1 – *Alicyclobacillus* xylanase – Optimization Method 1
X2 – *Alicyclobacillus* xylanase – Optimization Method 2
X3 – *Actinomadura* xylanase – Optimization Method 1
D1 – Protein D1 from *Chlamydomonas reinhardtii*

Internal Check Primers (to amplify fragments of relevant gene sequences to confirm presence/absence):

C1CheckF: TATTGCAGGTATGTTAGG
C1CheckR: ATATGACAAGCAGGTTCA
C2CheckF: ATGGGTGGATTCTATGGT
C2CheckR: CCAGGAGGTGTATTATCG
C3CheckF: TGGGATGTTCCATTAGGT
C3CheckR: AGAAAGACTCCAAGCACCAGATA
D1CheckF: TGTAATCCCAACTTCTAAC
D1CheckR: ATGATGTCTGCCCAAG
X1CheckF: CGTGTTCGTGCTCGTCTT
X1CheckR: CCACCTGGATTCCAACCTTTACC
X2CheckF: AAAGTCTCGTAGGCGTGTT
X2CheckR: CCAAGTTCCATAGAAACAGTGCC
X3CheckF: CGTTACCGTCCATACTTT
X3CheckR: TCGTTAGCAACATCCCAATCGTA

Flanking p-546 (primers anneal to regions flanking the gene insertion site in the p-546 plasmid, optimally designed for each inserted gene sequence):

flp546C1aR: GGCGAAAGGGGGATGTG
flp546C1aF: CAGCGCGCATTACCCTC
flp546C1bR: CGAAAGGGGGATGTG
flp546C1bF: CCTCACTAAGGGAACAA
flp546C2R: GCGAAAGGGGGATGTG
flp546C2F: CGTTTGATTTTTTGTGGTAT
flp546C3R: CGAAAGGGGGATGTG
flp546C2F: CTGGAGCTCGTCCTATT
flp546DX1R: TGGCGAAAGGGGGATGT
flp546DX1F: CTCACTAAGGGAACAAAAG
flp546DX2R: GAAAGGGGGATGTGCTG
flp546DX2F: CCTCACTAAGGGAACAA
flp546DX3R: TAACGCCAGGGTTTTCC
flp546DX3F: GTACATAAATGTGCTAGG

Flanking ppsbD (primers anneal to regions flanking the gene insertion site in the ppsbD plasmid, optimally designed for each inserted gene sequence):

ppsbDFlankingR: TTTATCAGAAGAAGCCAAGA
ppsbDFlankingF: CACTAAGGAATGCGTGTCGG
flppsbDX1R: ATAGTTTATCAGAAGAAGCCA
flppsbDX1F: CGAAGGGCTTGGGGGAAA
flppsbDX2R: ATTTTTCTGGCACATCGTA
flppsbDX2F: AAAGATACAAGCCCGAAGG
flppsbDX3R: TATAGTTTATCAGAAGAAGCCA
flppsbDX3F: AAAGATACAAGCCCGAAGG
flppsbDC1R: TTGGTTTTAGTCAACGAAG
flppsbDC1F: AAGATACAAGCCCGAAGG
flppsbDC2R: TTTATCAGAAGAAGCCAA
flppsbDC2F: GAAAGCATCGCTACTGGT
flppsbDC3R: GGTACCTCCTCCACAACTAG
flppsbDC3F: AAAGATACAAGCCCGAAGG

Miscellaneous Primers

Artificial primer binding sites flanking all gene constructs for full sequence amplification:

Fphy: GACCGACATCATCAGTTTGACC
Rphy: CCAGCCTAACGGATACGAACTAC

Primers to amplify 16S rRNA gene fragment for sequencing and identification of microbial species:

LD27F: AGAGTTTGATCCTGGCTCAG
LD1525R: AAGGAGGTGATCCAGCC

Forward Primers to replace 5' NcoI restriction site with NdeI site to remove duplicate initiation codon in p-546 plasmid constructs:

FNdeC1: CAGTTTGACCCATATGGATACTATTACAGTAG
FNdeC2: CAGTTTGACCCATATGGATACTATCACTGTAG
FNdeC3: CAGTTTGACCCATATGAATACATTACTTTTTC