DEPARTAMENTO DE CIÊNCIAS DA VIDA

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Pesquisa de estirpes de Saccharomyces cerevisiae para expressão de proteína recombinante

Dissertação apresentada à Universidade de Coimbra para cumprimento dos requisites necessários à obtenção do grau de Mestre em Bioquímica, realizada sob orientação científica do Doutor Pedro Castanheira (BIOCANT) e do professor Doutor Euclides Pires (Universidade de Coimbra)

Paulo Filipe Ribeiro do Espírito Santo

2013

Agradecimentos

Este espaço dedico a todas as pessoas que me acompanharam durante a minha vida académica e pessoal, que me proporcionaram bons momentos, e com os quais não teria chegado aqui.

Gostaria de agradecer ao professor Doutor Euclide Pires e ao professor Doutor Carlos Faro por terem proporcionado todas as condições necessárias para a realização deste trabalho na Unidade de biotecnologia molecular no Biocant.

Ao Doutor Pedro Castanheira deixo um especial agradecimento por me proporcionar a possibilidade de trabalhar numa das áreas mais bonitas da ciência. Obrigado pelo apoio, orientação, e paciência durante este último ano.

À professora Doutora Isaura Simões e à Carla agradeço por todos os esclarecimentos, conselhos e sugestões disponibilizados ao longo deste trabalho.

Gostaria também de agradecer a todo o pessoal do Biocant, em especial, a todas as pessoas que fazem parte do grupo de Biotecnologia Molecular. Ao Rui, Marisa, André Soares, Rita Leal, Ana Sofia Lourenço, Daniela, Paulo e à Joaninha). Obrigado por todos os momentos divertidos, pelo fantástico ambiente que existe dentro e fora do laboratório, pelos vossos conselhos e opiniões que me ajudaram neste trabalho ate ao fim.

Ao André da genómica mando também um agradecimento especial pela disponibilidade e ajuda no meu trabalho.

Não me tendo esquecido de ti, agradeço te pela tua paciência eterna, momentos de parvoíce e principalmente pela pessoa 5 estrelas e meia que sempre foste para mim. Ficas ca dentro Pedro Curto.

Aos meus colegas de Coimbra, com quem cresci desde que entrei para este mundo académico, com quem partilhei imensos bons momentos, e poucos maus, Agradeço do fundo do coração ter vivido convosco (Pedro Farias, José Codêsso, Carlos Custodia, Denis Brito, Cláudia Saraiva Luís Miguel, Sara Handem) Peço desculpas antecipadas a quem me esqueci, certamente não é por mal

A casa 23 que me albergou durante 6 anos da minha vida não seria um sítio espetacular de viver sem as pessoas que por ela passaram. Aos que ainda la habitam, obrigado André Mendes por me educares, e João Lima por seres uma pessoa impecável.

Aos meus amigos da minha santa terrinha deixo também a minha homenagem, pelas risadas e palhaçadas, copos, jantares, festas e sobre tudo pela amizade. Não podendo evitar, agradeço especialmente ao Rui Moura. És e sempre serás o amigo que fica no peito.

Ao pessoal das futeboladas, agradeço os momentos de divertimento neste belo desporto.

Não ficas para última, porque essa parte é para os meus pais. Mas a ti Raquel, agradeço-te por existires na minha vida. Por seres a fonte que gera a minha alegria e boa disposição, que parece que é uma das minhas qualidades. Pelo teu carinho e amor, pela tua paciência e dedicação, e pela força que me deste desde os nossos primeiros dias ate hoje. Obrigado

Como isto já esta a ficar muito grande vou finalizar. Agradeço à minha família, em especial aos meus pais, isto é fruto do vosso esforço e apoio que nunca me faltou. Obrigado por me ajudarem a tornar num Homem. Sem me esquecer de ti, Jorge, o próximo a chegar aqui és tu, terei sempre ambas as minhas mãos abertas para te ajudar no quer que seja.

Obrigado a todos

Table of contents

Agradecimentos i

Table of contents v

Abstract ix

Resumo xi

Abbreviations xiii

CHAPTER 1 15

Introduction 15

1.1. EXPRESSION SYSTEMS 17

1.1.1. E. coli expression system 18

1.1.2. Other prokaryotic systems 21

1.1.3. Yeast Pichia pastoris 22

1.1.4. Filamentous Fungus – Factories to secrete recombinant protein 23

1.1.5. Insect cell expression system 24

1.1.6. Mammalian – A relative mature expression system 25

1.1.7. Trangenic animmals expression system 26

1.1.8. Trangenic plant expression sytem 26

1.1.9. Cell-Free systems 27

1.1.10. Saccharomyces Cerevisiae Expression system 28

1.2. OBJECTIVES 36

CHAPTER 2 37

Materials and methods 37

2.1. MATERIAL AND METHODS 39

2.2. METHODS 39

2.2.1. PCR 39

2.2.2. CPEC cloning 41

2.2.3. Gibson Assembly™ and In-Fusion®HD 41

2.2.4. Vector assembly. Restriction-Ligation cloning 42

2.2.5. Saccharomyces deep well preparation 42

2.2.6. Carrier DNA preparation 43

2.2.7. Yeast medium sclae tranformation 43

2.2.8. 96-well Microtiter tranformation 44

2.2.9. High throughput activity assay 44
2.2.10. Small-scale screening for protein expression 45
2.2.11. Protein Expression and Purification 45
2.2.12. SDS-PAGE and Western Blot analysis 46
2.2.13. Enzyme activity assay 47
CHAPTER 3 49
Results and Discussion 49
3.1. VECTOR ASSEMBLY 51
3.1.1. PCR. Sequence-independent cloning 52
3.1.2. Circular Extension polymerase cloning - CPEC 54
3.1.3. Gibson and In-Fusion 55
3.1.4. Digestion-Ligation cloning 57
3.2. S. CEREVISIAE TRANSFORMATION – LIAC/SS-CARRIER/PEG METHOD 60
3.3. SMALL SCALE PROTEIN EXPRESSION IN LAB STRAINS 61
3.4. LARGE SCALE PROTEIN EXPRESSION AND PURIFICATION 64
3.5. 96 MICROTITER WELL TRANSFORMATION PROTOCOL - OPTIMIZATION 71
3.6. HIGH THROUGHPUT SCREENING – CONTROLS 73
3.7. HIGH THROUGHPUT TRANSFORMATION 74
CHAPTER 4 77
Conclusions and Future Remarks 77
4.1. CONCLUSIONS AND FUTURE REMARKS. 79
REFERENCES 81

Abstract

Recombinant DNA expression constitutes a major approach in gene function studies that naturally complement genetic and genomic research. Expression systems provide an invaluable tool for investigating the roles of novel gens, either in their original cellular environment or in specialized host organisms, to express high quantities of recombinant proteins for biochemical studies and structural determination, or even for industry applications or medical applications. Higher yield of proteins are achieved in prokaryotic systems and E. coli is the traditional host for producing recombinant protein. Its drawback of poor post-translational modifications when compared to a eukaryote organism, gave S.cerevisiae a chance to stand out as expression system. It's well known genetics, influence on mankind's economics and culture in beer wine and bread making during centuries, its GRAS status, fast and easy growth to high densities with low cost medium, along with its eukaryote machinery made it one of the most appreciated expression systems used today. Genetic engineering is one of the best tools to improve already existent strains, however yeast libraries harbour huge biodiversity with uncharacterized strains from witch much profit can be taken. New enzymes, new organic molecules and new strains with specific enhanced capacities can be lying in this libraries.

In this project, *S. cerevisiae* library of wine and vineyeard strains was screened to discover and validate a S. cerevisiae expression strain. The screen was prepared first by assembling an expression vector with the pAMT20 shuttle vector as backbone, *mus musculus* salivary α-Amylase as reporter enzyme, KanMX4 as selection marker, and *K. Lactis* α-mating factor prepro peptide as secretion leader. Transformation was adapted and carried on in a high throughput manner to screen the library and activity seen as halos formed around colonies in agar plates supplements with starch. Two commercial lab strains were used as reference for protein production levels, and to do so, a large-scale expression in liquid medium was performed and protein purified from supernatant in various steps, and activity measured to set a threshold for later comparison with the yeast candidates selected from the library. From 400 strains, 3 candidates stood out the average and are serious candidates to proceed with validations as a strain to produce recombinant protein.

Keywords: *S.cerevisiae*, High-throughput, Screening, Expression system, Yeast library, protein purification, α-Amylase

Resumo

A expressão de ADN recombinante é uma abordagem abundantemente utilizada para estudar a função de genes complementando as áreas de investigação genética e genómica. Os sistemas de expressão representam ferramentas valiosas para investigar a função de novos genes, expressando-os em grandes quantidades, quer nos hospedeiros aos quais pertencem ou em sistemas heterólogos, para estudos de caracterização bioquímica, determinação de estrutura 3D, e ate para serem usados na indústria ou na área da medicina. Os sistemas procariotas como a tradicional *E. coli* produzem maiores quantidades de proteína recombinante mas apresentam algumas limitações nas modificações pós-traducionais de proteínas. Face às limitações do sistema em *E.coli,* a *S. cerevisiae* destacou-se como sistema de expressão heteróloga simples e robusto que permitia modificações pós-traducionais em proteínas. Acresce ainda o facto de ser microrganismo que reúne imensa informação acerca do seu genoma, é usado como modelo de estudo de mecanismos eucariotas em diversas áreas da ciência, teve influência preponderante na cultura e economia humana durante seculos como a produção de pão cerveja ou vinho, o seu estatuto GRAS e o facto de atingir grandes densidades celulares de forma rápida e pouco dispendiosa tornou-o rapidamente num sistema robusto e de uso rotineiro no laboratório. A engenharia genética é uma ferramenta interessante para melhorar as características de algumas estirpes já existentes, no entanto existem outras estipes de leveduras que nunca foram vistas ou caracterizadas, que estão escondidas por de trás da enorme biodiversidade existente em bibliotecas criadas a partir do isolamento de estirpes no seu meio natural. Estas bibliotecas têm um potencial enorme para se descobrir por exemplo novas estirpes com capacidades peculiares e melhores, novas enzimas, novas moléculas orgânicas entre outras.

Este projecto, tirou partido do facto de dispor de uma libraria de leveduras isoladas de mostos de vinho ao longo de vários anos, à qual se fez uma seleção de estirpes com capacidade de produzir proteína heteróloga a níveis superiores à média. Para esta selecção criteriosa construiu-se um vetor, baseado no pAMT20, contendo uma alfa amílase salivar de *mus musculus* que serviu proteína repórter de actividade das colonias recombinantes, a cassete KanMX4 que é uma marca de selecção que confere resistência ao antibiótico G418, e uma sequência do prepro peptído responsável pela secreção de uma feromona com origem no *K. lactis*. O protocolo de transformação foi ajustado para funcionar em larga escala, e os testes de actividade foram avaliados como o tamanho de halos formados por colonias transformadas em meio solido com amido. Duas colonias de laboratório foram usadas como controlos no que diz respeito à quantidade de proteína produzida. Para isso fez-se uma expressão em meio liquido em larga escala, e purificou-se a proteína recombinante do sobrenadante e mediu-se a

actividade. Esta actividade serviria como patamar de expressão proteica para comprara mais tarde com os níveis de produção de proteína recombinante de uma potencial estirpe isolada da biblioteca. De 400 estirpes evidenciaram se 4 estirpes que são fortes candidatos para continuar para a fase de validação de estirpe para expressão de proteína heterologa

Abbreviations

ACMNPV - Autographa californica multinucleopolyhedrovirus

ADH2 Alcohol Dehydrogenase II

ADHI - Alcohol Dehydrogenase I

AMY - α-Amylase

BHK Baby Hamster Kidney

CHO -Chinese Hamster Ovary cells

CIP - Calf Intestinal alkaline Phosphatase

CPEC - Circular Polymerase Extension Cloning

EDTA - Ethylene Diamine Tetra Acetic

EMEA - European Medicines Agency

ER - Endoplasmic Reticulum

FDA - U.S. Food and Drug Administration

GAL1, GAL10 - Galactose promoters

GRAS – Generally Regarded As Safe

K. lactis - Kluyveromyces lactis

LAB - Lactic Acid Bacteria

LDP - Low Density Precipitate

LiAc – Lithium acetate

OD600 – Optical density measured at wavelength of 600nm

PCR – Polymerase Chain Reaction

PDB - Brookhaven Protein Data Bank

PEG – Polyetilenoglycol

PrAF – Precipitated protein with 90% ammonium sulphate, After filtration

SDS-PAGE – Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

TE - 10 mM Tris-HCl pH 7.5, 1 mM EDTA

YEP – yeast episomal plamid

αMF - αMating factor secreting leader (αMF)

CHAPTER 1

Introduction

1.1. Expression systems

The expression of recombinant proteins in heterologous systems has proved invaluable for the determination of the biochemical and structural properties of enzymes. This is a pre-requisite for performing many biochemical studies and to enrich our knowledge of gene function, which too often relies on theories derived from sequence alignment or whole-organism genetic studies (Srivatsan, Han et al. 2008). Once a protein has been successfully expressed and purified, this opens the door to several new experiments including testing for enzyme activity, interactions studies with other proteins or nucleic acids, antibody generation, and ultimately structure determination (Vedadi, Arrowsmith et al. 2010).

The production of recombinant protein has to follow an economic and qualitative rational, which is dictated by the characteristics and anticipated application of the compound produced. For the production of technical enzymes or food additives, gene technology must provide an approach which has to compete with mass production of such compounds from traditional sources. As a consequence, production process have to be developed that employ high efficient platforms and that lean on use of inexpensive media components in fermentation processes. For the production of pharmaceuticals and other compounds that are considered for administration to humans, the rationale is dominated by safety aspects and a focus on the generation of authentic products (Gellissen 2006).

Some of the systems presented are distinguished by an impressive track of record as producers of valuable proteins that have already reached the market, while others are newly defined systems that have yet to be established but demonstrate great potential for industrial applications. There is clearly no single system that is optimal for all possible proteins, but all have special favourable characteristics and limitations/drawbacks (for resumed information consult table 1). There exists although a correlation between the complexity of a particular protein and the complexity and capabilities of an expression platform. Single-subunits proteins can easily be produced in bacterial hosts, whereas proteins demanding presence of several disulphide bonds or authentic complex mammalian glycosylation necessitate higher eukaryotes as host.

The Gram-negative *Escherichia coli* was the first organism to be employed for recombinant protein production because of its long tradition as a scientific organism, the ease of genetic manipulation, and the availability of well-established fermentation procedures. Typically over 90% of all proteins deposited in the Brookhaven Protein Data Bank (PDB) (http://www.rcsb.org) whose

structures have been determined have been expressed in *E. coli.* However, the limitation in secretion and the lack of glycosylation impose restrictions for general use.

To overcome these problems a variety of expression systems have been devised. These are more suitable for expression of eukaryotic proteins because they possess the most appropriate chaperones to aid protein folding, contain homologous binding partners that may help stabilize proteins that naturally exist as part of a multicomponent complex, and have also the capability to perform the full range of post-translational modifications. These include expression in yeast, such as Saccharomyces and Pichia, which provide a similar intracellular environment to mammalian cells for the processing of proteins and require simple media and reach high cell densities (Buckholz and Gleeson 1991, Romanos, Scorer et al. 1992, Cregg, Vedvick et al. 1993). Filamentous fungi which have the capacity to produce high amounts of secreted protein and are of great use in food industry (Srivatsan, Han et al. 2008). Insect cells such as *Spodoptera* and *Trichoplusia* have been shown to overproduce recombinant proteins, generally competent in the glycosylation, and possess and efficient protein transport system (Matsuura, Possee et al. 1987, King and Possee 1992). Mammalian cells, [e.g., Chinese hamster ovary cells (CHO)] are capable of faithfully modifying heterologous compounds according to mammalian pattern. Animal transgenic systems are efficiently used as bioreactors for production of therapeutic proteins with high value for pharmaceutical use (Echelard 1996, Soler, Le Saux et al. 2005). Transgenic plant system offer the comfortable option to produce and store recombinant protein in certain organs and are easy to scale-up. Lastly, the peculiar in-vitro expression system, with the potential to produce otherwise toxic proteins for living cells, the cell free systems.

All these protein expression systems will be reviewed in more detail below. Microbe have been used to produce a myriad of primary and secondary products to benefit mankind for decades. Molecular biology and genetic engineering have been the major driving force in research and the production of high levels of proteins, and changed the faces of pharmacology, medicine and industry.

1.1.1. E. coli expression system

The Gram-negative bacterium *Escherichia coli* was not only the first microorganism to be subjected to detailed genetic and molecular biological analysis, but also the first to be employed for genetic engineering and recombinant protein production. Our knowledge of its genetics, molecular

biology, growth, evolution and genome structure has grown enormously since the first compilation of a linkage map in 1964 (Neidhart 1996).

From a model organism for laboratory-based basic research, *E. coli* has evolved into an industrial microorganism, and now the most frequently used prokaryotic expression system. It has become the standard organism for the production of enzymes for diagnostic use and for analytical purpose, and even used for the synthesis of pharmaceutical proteins, provided that the desired product does not consist of different multiple subunits or require substantial post-translational modification (e.g. Glycosylations).

The ease of manipulation, its ability to grow rapidly to high densities in simple inexpensive substrates together with its well-characterized genetics made *E. coli* by far the most widely employed host. A huge body of knowledge and experience in fermentation and high-level production of proteins has grown up during the last 40 years. Many strains are available which are adapted for the production of proteins in the cytoplasm or periplasm, and hundreds of expression vectors with differently regulated promoters and tags for efficient protein purification have been constructed.

1.1.1.1. *E. coli* strains

Theodor Escherichia described back in 1885, what was to become the most widely used microorganism, and the cell line *E. coli* K12 was isolated in 1992 and deposited as “K-12” at Stanford university in 1925. A pedigree and description of various standard strains, including *E. coli* B strain, commonly used all over the world, was first published in 1972 (Bachmann 1972). Nowadays, many strains are available such as BL21 and derivatives (deficient strain in proteases), Origami (K-12 derivate, with mutations to enhance disulphide bond formation in cytoplasm), Rosetta (BL21 derivate, suitable for eukaryotic protein production with rare codons in *E. coli*) and many others are available to offer a large spectrum of features to express the desired recombinant protein.

1.1.1.2. Expression vectors

Many commercial expression systems designed for various applications and compatibilities are available such as the T7 based pET expression system (Novagen), systems using the λPL promoter/cI repressor (e.g., Invitrogen pLEX), Trc promoter (e.g., Amersham Biosciences pTrc), Tac promoter (e.g., Amersham Biosciences pGEX) and hybrid lac/T5 (e.g., Qiagen pQE) promoters are common (Sørensen and Mortensen 2005)

The pET expression system is the most powerful system yet developed for the cloning and expression of recombinant proteins in *E. coli*. Target genes are cloned in pET plasmids under control of strong bacteriophage T7 promoter and expression is induced (with IPTG) by providing a source of T7 RNA polymerase in the host cell. T7 RNA polymerase is so selective and active that almost all of the cell's resources are converted to target gene expression and the desired

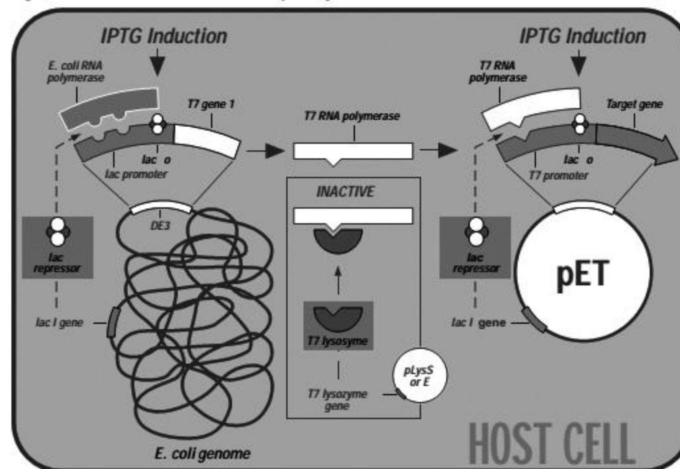

Figure 1 – Control elements of the pET system. **Schematic illustration of host and vector elements available for control of T7 RNA polymerase levels and the subsequent transcription of a target gene in a pET vector. T7 RNA polymerase gene is under the control of the lacUV5 promoter. The pLys plasmids encode T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase, and thus reduces its ability to transcribe target genes in uninduced cells**

1.1.1.3. Protein production in *E. coli*

Proteins in *E. coli* can be produced in the cytoplasm, targeted to the periplasmic space, or as inclusion bodies. When targeted to the cytoplasm, protein yields are the highest and fusion proteins may be used to improve solubility, purification, used as reporters, or protect against degradation by intracellular proteases (Janiyani, Bordelon et al. 2001, Magliery, Wilson et al. 2005). There are many fusion partners described in literature and actively used in laboratory or industry and can be reviewed in (LaVallie and McCoy 1995).

Proteins destined for the periplasmic space and outer membrane must pass through the cytoplasmic membrane. An N-terminal signal peptide fused to the protein is used to target proteins to the secretory pathway in *E. coli*. The periplasmic space also has its own chaperones to assist in protein folding and to generate disulphide bonds (Hayhurst and Harris 1999), essential to some proteins which cannot fold in the reducing environment of the cytoplasm, or to protect the desired product from proteases because periplasm itself has a different set of proteases and are less abundant (Raivio and Silhavy 2001). Disulphide bonds usually do not form in the cytoplasm, but mutation in target genes,

which encode proteins that maintain a reductive environment, allowed for disulphide bond formation and some host strains (Origami) have been established.

Inclusion bodies do have advantages of protection against proteolytic degradation, and easy separation and purification from other cell material (Rudolph and Lilie 1996). Heterologous proteins produced as inclusion bodies are inactive, aggregated and insoluble, usually possessing non-native intra and inter-molecular disulphide bonds and unusual free cysteine, and solubilisation steps are required (Fisher, Sumner et al. 1993, De Bernardez Clark 1998).

Even being the oldest proteins expression system, it is still the most used, and developments and engineering strategies to improve further this prokaryotic capacity to express and process recombinant protein are still going on (Sørensen and Mortensen 2005).

1.1.2. Other prokaryotic systems

Another growing field is biotechnological engineering of gram positive bacteria such Lactic acid bacteria (LAB), including a large number of Gram-positive cocci or bacilli. LAB are widely used in the food industry for the production and preservation of fermented products. Due to their GRAS (generally regarded as safe) status, a great deal has been done to genetically modify them to improve their traits. They grow to high densities in simple media, can secrete high levels of well folded protein, and over the last decades a number of genetic tools have been developed for lactic acid bacteria, so homologous and heterologous proteins could be expressed.

High-level production of proteins in lactic acid bacteria has been obtained using L. lactis constitutive promoters (Kuipers, de Ruyter et al. 1997). Several inducible expression promoters have been developed and through these promoters, gene expression can be controlled. The well characterized and commonly used controllable expression system, designed for *Lactobacillus lactis*, is the nisin-controlled gene expression (NICE) system, in which nisin serves as an inducer (De Ruyter, Kuipers et al. 1996, Kuipers, de Ruyter et al. 1998). The NICE system has proved to be highly versatile, and its use in pharmaceutical, medical and food technology fields is most promising (Le Loir, Nouaille et al. 2001). A similar system is also available for *Bacillus subtilis*, the SURE system, with the advantage of achieving higher expression levels are compared to NICE (Bongers, Veening et al. 2005).

Some disadvantages mark this hosts as expression systems such as instability of plasmids or excessive protease secretion, which degrade the protein of interest (He, Brückner et al. 1991). But, some of these gram positive bacteria are used in industry in production (mainly for homologous

expression) of enzymes such as proteases for detergents (Maurer 2004), and amylases for starch and baking (Sanchez, Prim et al. 2002).

1.1.3. Yeast Pichia pastoris

The ability of certain yeast species to utilize methanol as a sole source of carbon and energy was discovered approximately 30 years ago by Koichi Ogata (1969). In the early 1980s, Researchers at SIBIA isolated the AOX1 gene (and its promoter) and developed vectors, strains, and methods for molecular genetic manipulation of *P. pastoris* (Cregg, Barringer et al. 1985, Ellis, Brust et al. 1985, Tschopp, Brust et al. 1987, Cregg, Madden et al. 1989). The first step in the metabolism of methanol is the oxidation of methanol to formaldehyde, generating hydrogen peroxide in the process, by the enzyme alcohol oxidase (AOX). This first step occurs in specialized organelles called peroxisomes. This peroxisomes sequester toxic hydrogen peroxide and keep it away from the rest of the cell. There are two genes that code for AOX – *AOX1* and *AOX2* – but the *AOX1* gene is responsible for the vast majority of alcohol oxidase activity in the cell (Ellis, Brust et al. 1985) The *AOX1* gene is tightly (at the level of transcription (Koichi Ogata 1969, Tschopp, Brust et al. 1987, Cregg, Madden et al. 1989) regulated and induced by high levels of methanol. GAP (glyceraldehyde 3-phosphate) is also one of the choice promoters for inducible expression.

Pichia expression systems can either express proteins intracellularly or secreted into medium, benefiting the latter option from the fact that *P. pastoris* secretes only low levels of endogenous proteins (Tschopp, Sverlow et al. 1987). Vectors such as pPICZ, pHIL, pGAPZ, pFLD, and pPIC (all from Invitrogen) are some of the vectors available for recombinant protein expression in this system either for intra or extracellular expression

Disadvantages of this system are the N-glycosylation of proteins (Hyperglycosylation), which differ from higher eukaryotes (Higgins and Cregg 1998, Cregg 1999, Cereghino and Cregg 2000), and are potentially allergenic when applied in therapeutics.

Nevertheless it has been used as a tool in biotechnology, and to date well over 550 heterologous proteins have been expressed in *P. pastoris (see* http://www.kgi.edu/faculty-and-research/profiles/james-m-cregg.html). Significant advances in the development of new strains and vectors, improved techniques, and the commercial availability of these tools coupled with a better

understanding of the biology of *Pichia* species have led to this microbe's value and power in commercial and research labs alike.

1.1.4. Filamentous Fungus – Factories to secrete recombinant protein

Filamentous fungi are extraordinary organisms widely impact on many aspects of our lives. These organisms are characterized as having branched filamentous structures or hyphae having typical diameters of 2–18 um, with (higher fungi) or without (lower fungi) cross-walls or septae. Higher fungi include *Aspergillus, Penicillium, Trichoderma* and *Fusariumspecies*. Lower fungi include *Rhizopui* and Mucorspecies. Filamentous fungi are chemo-organotrophs meaning they obtain their energy and carbon by oxidation of organic compounds.

Filamentous fungi such as *Aspergillus niger*, *Aspergillus oryzae*, and *Trichoderma reeseiare* are able to produce and secrete large concentrations of enzymes into the environment (e.g. amylases, amyloglucosidases, proteases, cellulases, pectinases, laccases/ligninases, phytases, lipases, microbial rennets and glucose oxidase) (Iwashita 2002, Wang, Ridgway et al. 2005).

The known high productivity characteristics of filamentous fungus are in parte related to their inherent abilities to grow at high rates and to high biomass densities supported by low cost substrates in relative simple fermenters. This made them especially interesting targets for recombinant protein production because of their demonstrated capacity to hyperproduce and secrete enzymes. As final products the secreted enzymes can be applied in the agricultural, food and nutrition, biomedical and pharmaceutical, and energy and industrial sectors (Schuster, Dunn-Coleman et al. 2002).

Foreign genes are normally incorporated via plasmids into chromosomes of the filamentous fungi where they integrate stably into the chromosome as tandem repeats providing superior long-term genetic stability. When integrated in the fungal genome the recombinant DNA is not lost because of unequal partitioning, as may occur with the self-replicating plasmids used to produce recombinant proteins in bacteria or yeast

Secreted yields of some heterologous proteins have been comparatively low in some cases (Iwashita 2002). Such differences are consequence of RNA instability (Verdoes, Punt et al. 1995) and high protease activity (Gouka, Punt et al. 1997), and they share the same limitations in N- glycosylation's as other yeasts posing a problem if this feature is required for a give protein. Recombinant protein concentrations have been increased by fusing the gene of interest with genes of fungal origin ((Gouka,

Punt et al. 1997, Punt, van Biezen et al. 2002). Such approaches have been effective with some target heterologous proteins but not with others

1.1.5. Insect cell expression system

Baculoviruses are insect pathogens that regulate insect populations in nature and are being successfully used to control insect pests. Examples include control of the codling moth, *Cydia pomonella*, on apple (Lacey and Unruh 2005), the velvet bean caterpillar, *Anticarsia gemmatalis*, in soybean fields (Moscardi 1999) and the cotton bollworm, *Helicoverpa armigera* (Sun and Peng 2007). Typical properties of baculoviruses, such as the high level of very late gene expression, make them highly suitable as vectors for foreign gene expression.

In recent years, the baculovirus expression vector system has become a well-established and popular method for producing high yields of structurally, functionally and antigenic authentic foreign proteins in insect cells (Kost, Condreay et al. 2005). Traditionally, when making a recombinant baculovirus, the target gene is cloned into a transfer vector which contains sequences that flank the polyhedrin (*polh*) locus in the virus genome. The viruses generally used are *Autographa californica* multinucleopolyhedrovirus (AcMNPV) or in a less extend Bombyx mori (Bm) NPV. Homologous recombination between vector and virus genome occurs, producing recombinant budded virus which are harvested from the culture medium and used to infect insect cells. Some commercial expression systems are already in the market and used to produce recombinant protein such as *flash*BAC system (Oxford Expression Technologies Ltd), BacMagic (Merck) and BaculoOne (PAA) systems.

The insect cells more commonly used are the Sf21 cells (O'Reilly, Miller et al. 1994), originally derived from the pupal ovarian cells of *Spodoptera frugiperda* (fall army worm), Sf9 cells (Vaughn, Goodwin et al. 1977), which are a clonal isolate of Sf21, and High Five cells (BTI-TN-5B1-4; Invitrogen). Insect cell culture medium utilizes a phosphate buffering system, rather than a carbon-based buffers (used in mammalian cells) and do not need CO_2 incubators.

Insect cells, have a more advanced post-translation machinery, capable of producing high levels of recombinant mammalian derived proteins (Agathos 1991) or proteins to use in pharmaceutical applications, and that they produce high levels of target proteins (Maiorella, Inlow et al. 1988).

Insect cell systems however, do have some shortcomings such as particular patterns of post-translational processing that differ from mammals (Bisbee 1993), some cases of inefficient secretion

or aggregates formation. These problems occur because it still is a recent expression system and future experiments and improvements will overcome these problems.

1.1.6. Mammalian – A relative mature expression system

Mammalian cells constitute a demanding system for the production of heterologous proteins. The need for specialized media and sufficient oxygen supply, low densities and slow growth kinetics, and a high sensitivity of the cells to mechanical stress are obstacles which must be overcome in routine fermentations (Wurm 2004). Furthermore, mammalian cells are potential targets for adventitious viral agents, and processes based on such cells must be rigorously monitored. Despite these difficulties, mammalian cells are the preferred production systems for the synthesis of authentic glycoproteins intended for administration to humans.

The ease of production in bacterial systems must be counterbalanced against the need to dissolve and renature misfolded, aggregated, insoluble protein. In contrast, the chaperone system in mammalian cells ensures that proteins are secreted in correctly folded form. Whereas eukaryotic microbial systems such as yeasts are also capable of modifying recombinant proteins by proteolytic processing of precursors, formation of disulphide bridges and phosphorylation, only mammalian cells are able to glycosylate proteins in the patterns characteristic of higher eukaryotes, yielding products that are identical to their natural human counterparts (Wu, Yeh et al. 2010).

Normally to produce recombinant proteins in mammalian cells, a coding sequence is inserted between a strong promoter and a polyadenylation sequence contained in an expression vector (e.g. pEGFPN1, Clonetech; pCMVseaphyg, ProBioGen), and transfected into the suitable cell line. The vector and cell line form an integrated system designed to produce maximal yields of a give protein. Cell lines used to produce therapeutic proteins in fermentation processes are CHO (Chinese hamster ovary), BHK (baby hamster kidney) and NS0 cells (mouse myeloma) and Sp2/0 (myeloma cells). More recently HEK 293 (derived from human embryonic kidneys cells) are used for transient protein expression. The establishment of a recombinant mammalian production line takes a full year, therefore transient protein expression in mammalian cells (e.g. HEK293 and CHO) became desirable for rapid protein production. They produce less amount of proteins, but the levels are sufficient to satisfy screenings of biotherapeutics, or structural and functional studies.(Edwards and Aruffo 1993, Hanai, Nagata et al. 2004).

Despite continuous advances being made in basic research for this system, the mechanisms of expression and secretion remain poorly understood.

1.1.7. Trangenic animmals expression system

Modern state-of-the-art molecular biological techniques allow us to transform recombinant DNA into fertilized animal ovum, and then these DNA components can be integrated into host genome at certain frequency, expressing the foreign protein in specific tissues or organs of the body depending on the promoters used. Up to now, a variety of transgenic animal species are being used to produce recombinant proteins (Echelard 1996) or being built as disease models (Fan and Watanabe 2003).

A general approach is to target the expression of the desired protein to the mammary gland using regulatory elements derived from a milk protein gene and then collect and purify the product from milk (Clark 1998). Promoter sequences from a number of different milk protein genes have been used to target expression to the mammary gland, although significant problems remain with regard to achieving transgene expression levels consistent with commercial exploitation (van Berkel, Welling et al. 2002). Besides the mammary gland of transgenic animal, urine, blood, egg white, seminal plasma and silk worm cocoons are also for targeted protein expression. Goats, pigs, rabbit, sheep and some aquatic animals are being developed as production system. The individual animal expression system requires a relatively longer production period and higher costs than above-mentioned expression systems. So this system can express foreign proteins mainly for medical purposes, but not for industrial or environmental applications due to its relative high cost and low yield.

1.1.8. Trangenic plant expression sytem

For recombinant proteins production, the use of plants, as compared to that of living animals and animal cell cultures, is much safer and less expensive, requires less time, and is superior in terms of storage and distribuiton.

Plants shares some architectural and functional similarities with animal cells. Hence they constitute and optimal system in to express heterologous proteins that require complicated post-translational modifications, such as some glycoproteins, bioactive peptides, and drugs. In plants,

heterologous proteins expressed can be localized to different organs of the plant by controlling the tisse-specific regulatory sequences involved in gene expression. One can even maipulate the time of expression to specific growth stages or secrete the desired protein utilizing secretory signal sequences. As protein factories, plants can be grown in the field, providing a very inexpensive source of material (only water, minerals and sunlight) compared to any organims grown in fermenters. The scale-up is also very easy. More plants can be grown very easily, increasing total protein yield. Finally, expression of proteins in plant seeds results in a unit of production (seed) in which proteins are extremely stable, readily stored , and generally easier to extract and purify.

Transgenic plants can be produced in two ways. One way is to insert the desired gene into a virus that is normally found in plants, such as the tobacco mosaic virus in the tobacco plant (Casper and Holt 1996). The other way is to insert the desired gene directly into the plant DNA. plants can acquire some new or enhanced characteristics by the expression of target genes, such as disease resistance (Coca, Bortolotti et al. 2004), stress tolerance (Murakami, Tsuyama et al. 2000), enhanced growth (Holmberg, Lilius et al. 1997) and improving nutrition (Yoshihara, Takaiwa et al. 2005).

This system has some disadvantages, including low transformation and expression effeciencies, different glycosilation pattern from that of animal-derived cell expression systems (Breiteneder, Krebitz et al. 2012) and a controversial safety (Schauzu 2004)

1.1.9. Cell-Free systems

Since the pioneering studies conducted by Nirenberg and Matthaei more than four decades ago (Nirenberg and Matthaei 1961) cell-free protein synthesis has been a valuable tool for understanding how mRNAs are translated into functional polypeptides. In vitro translation systems are based on the early demonstration that cell integrity is not required for protein synthesis to occur, and are also called “uncoupled translation”. In its simplest form, translation can be accomplished using a crude lysate from any given organism (that provides the translational machinery, accessory enzymes, tRNA and factors) in combination with exogenously added RNA template, amino acids and energy supply.

The most popular sources of the lysates for cell-free systems are prepared based on Escherichia coli, wheat germ and rabbit reticulocytes. Similar to in vivo systems, the choice of the system should be determined by origin and biochemical nature of the protein and the specifics of the downstream application. Simply put, *E. coli* based systems provide higher yields of protein suitable for

structural studies (Kigawa, Yabuki et al. 1999), and Eukaryotic-based systems, which are less productive, provide a better platform for functional studies. (Jermutus, Ryabova et al. 1998, Endo and Sawasaki 2003, Jackson, Boutell et al. 2004). This is not necessary true for all cases, and thus similar to in-vivo systems, adjustments can be made.

The key goal for this systems is to synthesize biological active proteins, so protein folding and achieving post translational modifications are the current issue. Exogenous chaperon addition to a cell free reaction mixture led to increase folding levels of recombinant protein (Ryabova, Desplancq et al. 1997, Jiang, Ookubo et al. 2002). For glycosylation, developed strategies so far account for supplementing extracts with microsomal fractions to promote protein translocation to these compartments (Walter and Blobel 1982), adding non-natural amino acid linked to a monosaccharide moiety (works as a primer for glycosylation) along with glycotransferases (Zhang, Gildersleeve et al. 2004) has been tested with mild results. These results however can be the drift to meliorate and perfectioning of the cell free systems.

Until now, these systems have found a large variety of low and high throughput applications suitable for functional and structural proteomics (Kigawa, Yamaguchi-Nunokawa et al. 2002, Klammt, Lohr et al. 2004) and many other applications (for review see Katzen, Chang et al. (2005)).

1.1.10. Saccharomyces Cerevisiae Expression system

The Yeast *Saccharomyces cerevisiae* is widely recognized and used as a robust host for recombinant protein expression. It was the first eukaryote whose genome was fully sequenced (Goffeau A 1996), and has been engineered to express a wide variety of recombinant proteins for the last three decades. The first vaccine effective against human viral infections, hepatitis B (McAleer WJ 1984), was produced intracellularly in recombinant *S. cerevisiae*. The popularity of *S. cerevisiae* in basic and applied research is undoubtedly influenced by its classification as GRAS by the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMEA) (Huang CJ 2010).

This microbe provides a huge amount of information about its genetics, physiology, and biochemistry as well as genetic engineering and fermentation that have accumulated over time. Several databases such as the *Saccharomyces* Genome Database (http://www.yeastgenome.org/) and the Comprehensive Yeast Genome Database (http://mips.gsf.de/genre/proj/yeast/) contain an enormous amount of information concerning *S. cerevisiae* genes, open reading frames, and gene

products, as well as networks of protein interactors (General Repository of Interaction Datasets [http: //www.thebiogrid.org/]). The European S. *cerevisiae* Archive for Functional Analysis (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/) as well as the Japanese Yeast Genetic Resource Centre (http://yeast.lab.nig.ac.jp/nig/index_en.html) collect and store many useful tools for the yeast scientific community, e.g., strains and plasmids generated during various projects

When expressing recombinant proteins in heterologous hosts, a set of problems can appear, which might lead to unsuccessful results. Recent eukaryotic expression systems lack comparable accumulated information with regard to host physiology, genetics, proteins translation, protein interaction, secretion pathways etc., which can pose problems when poorly understood and become difficult to overcome.

Saccharomyces accumulated data makes it an appealing host to establish an expression system for recombinant protein production, because much is known, and common problems have been solved, or can be solved with available information. Some example of products on the market which are made in *S. cerevisiae* include insulin, hepatitis B surface antigen, glucagons, granulocyte macrophage colony stimulating factor (GM-CSF), hirudin, and platelet-derived growth factor (Demain AL 2009).

1.1.10.1. History of saccharomyces in industry

In most societies, fermented beverages and foods have an important role because of their economic and cultural importance and the development of fermentation technologies is deeply rooted in their history. Archaeologists have found evidence for the production of a fermented beverage in China at 7000 BC (McGovern, Zhang et al. 2004), and of wine in Iran and Egypt at 6000 BC and 3000 BC, respectively. 2003 (McGovern, Hartung et al. 1997, Cavalieri, 2003 #292). It's tolerance to high ethanol concentrations (Casey and Ingledew 1986), relatively low pH and inhibitory environments of raw materials in large fermentation tanks (Almeida, Modig et al. 2007) made S. *cerevisiae* the most attractive microorganism, and today Baking, wine making, brewing, and production of bioethanol constitute the majority of S. *cerevisiae* biotechnological industry. Knowledge of the genetics of industrial yeast strains lags behind that of laboratory strains of *S. cerevisiae* however, there is an increasing interest in the functional genomics of these strains (de Winde 2003) and efforts are being made to improve strains to satisfy the industry demands (Nevoigt 2008).

1.1.10.2. Saccharomyces a model eukaryote

The safety of the system is also guaranteed, including absence of endotoxins and oncogenes. In addition, it has proven to be a reliable model for mammalian disease and metabolic pathways, being frequently used as a eukaryote model to understand the role of some proteins in their native host (mammalian cells) (Romanos, Scorer et al. 1992), such as models is for G-protein-coupled pathway (Mentesana, Dosil et al. 2002).

Apart from *S. cerevisiae* culture simplicity, rapid growth, and low cost, which are similar with those of *E. coli*, the possibility to express proteins into different compartments, intracellular (soluble or membrane bound), and extracellular, a whole set of promoters, expression vectors and other functional studies are available to facilitate protein recombinant expression.

The expression cassettes can be integrated in episomal type vectors with low (CEN/ARS) or high copy number (2µ), allowing for different expression levels, and these vectors are generally hybrid between yeast-derived and bacterial sequences. The bacterial segment harbours elements required for plasmid propagation in an Escherichia coli host, such as an ORI and a sequence conferring resistance against a specific antibiotic like ampicillin. The yeast part may lean on two basic options: A first possibility leads to constructs which result in an episomal fate (YEp type plasmids) of the transformation plasmid. Such vectors replicate independently of the chromosomal DNA. 2µ plasmids are present in most *Saccharomyces* strains, making 2µ-based plasmids by far, the most used expression vectors in yeasts

Alternatively, vectors are mitotically stable integrated into the host's genome (YIp type) targeting the foreign DNA to specific genes of the host by homologous recombination. Integrated vectors are present in low copy numbers (Hinnen, Buxton et al. 1995).

1.1.10.3. Genetic tools for protein expression

Yeast mRNA promoters consist of at least three elements which regulate the efficiency and accuracy of initiation of transcription (Struhl 1989). Upstream activation sequences (UASs), TATA elements and initiator elements. It is advantageous to use yeast-derived (homologous) rather than heterologous sequences, because the former are more efficient and heterologous elements will sometimes not work in yeast. Constitutive promoters are derived from genes of the glycolytic pathway, because these lead to high-level transcriptional expression. On the other hand, regulated promoters can be controlled by controlling the availability of certain nutrients. This allows to increase yeast cell

mass prior to heterologous gene expression, so that the cell population can be optimized before the regulated promoters are turned on.

The first promoters used were from genes encoding abundant glycolytic enzymes in yeasts like, alcohol dehydrogenase I (ADHI) (Hitzeman, Hagie et al. 1981), phosphoglycerate kinase (PKG) (Hitzeman, Leung et al. 1983), glyceraldehyde-3-phosphate dehydrogenase (GPD) (Bitter and Egan 1984) or pyruvate kinase (PYK1), These promoters are constitutive, their activity may be enhanced/induced by addition of glucose to medium and despite being poorly regulated, and they are the most powerful of S. *cerevisiae*.

On the other hand, Galactose promoters (GAL1 and GAL10) are the most powerful tightly-regulated promoters of S. *cerevisiae.* The genes encoded by the GAL family are involved in metabolizing Galactose, they remain tightly repressed in the presence of glucose and are induced by galactose. Manipulations in order to improve the Galactose-regulated system have been carried to enhance the expression of heterologous genes, but thigh regulation was lost (Baker, Johnston et al. 1987). Some strategies fused GAL promoters to glycolytic to combine thigh regulation of GAL, with the high activity of glycolytic promoters (Partow, Siewers et al. 2010)**.**

Another well studied and well regulated gene is the promoter of the acid phosphatase (POH5). It is negatively regulated by inorganic phosphate and switching to a low-phosphate medium induces the promoter. Since the PHO5 promoter is not very strong, relative to glycolytic, its effective regulation trough POH5 has also been used to confer regulation to glycolytic promoters (Hinnen, Meyhack et al. 1989). MET3 and CUP1 are another example of negatively regulated promoters, by methionine and Ca^{2+} ions respectively (Mascorro-Gallardo, Covarrubias et al. 1996, Mao, Hu et al. 2002).

Repressible and strong promoters examples are the ADH2 (Alcohol dehydrogenase II, both powerful and tight regulated), SUC2 and CYC1 (CYC1 is a weak promoter) which are repressed by glucose presence in the medium (Price, Taylor et al. 1990). However this is a disadvantage in industrial processes, because maintaining high concentrations of glucose are expensive and difficult because of high cell density. Galactose regulated promoters also have a disadvantage in industrial processes, because of the need to add inducer to the medium, which is not economically compatible with large fermentation tanks.

Selection markers are another important component present in an expression vector, and yeast selections usually relies on complementation selection. It explores the host incapacity to synthetize a specific organic molecule (auxotrophic selection), essential for growth, such as LEU2, TRP1, URA3 or HIS3 which encode for leucine, tryptophan, uracil and hisitdine respectively. Another type of selection possible for yeasts is the use of antibiotics, also called dominant selection, and some examples of frequent use are the aminoglycoside antibiotic G418 (Agaphonov, Romanova et al. 2010), hygromycin

B (Gritz and Davies 1983), chloramphenicol (Hadfield, Cashmore et al. 1986), and neomycin (Prezant, William jr et al. 1996).

One appreciated feature of recombinant protein expression systems is ability in secretion of the target protein to facilitate purification and separate from all the cell material. The enriched endomembrane system of yeasts allows proteins to be secreted into the extracellular environment, as well post-translational modifications that are sometimes required to produce active recombinant protein *S.Cerevisiae* only secretes naturally 0.5% of its proteins, but it can be increased several fold so that secreted proteins are almost free of contaminations with other cell material. As higher eukaryotes, protein secretion of yeast is directed by an amino-terminal sequence (secretion signal sequence) that, targets proteins to surface, to a vacuole or are excreted. The well-studied leader to secrete *saccharomyces cerevisiae* mating pheromone, prepro-sequence of MFα1, is extensively used with reports of efficient secretory activity. Tags fused to proteins in *S. cerevisiae* normally serve the purpose of purification, or detection assays, such as GFP (Niedenthal, Riles et al. 1996).

1.1.10.4. *S. cerevisiae* Transformation

Genetic transformation of *Saccharomyces cerevisiae*, was first reported by Oppenoorth (Oppenoorth and van 1960) but until today, several techniques have been developed, with high efficiency of transformation and ease of manipulation. The removal of yeast cell wall by enzymatic treatment, using gut juice of snail" to yield protoplasts was first observed in 1957 (Eddy and Williamson 1957). But transformation of yeast cells with recombinant DNA became feasible for the first time in 1978 by treating spheroplasts with calcium and PEG (Hinnen, Hicks et al. 1978). Years later, intact yeast cell transformation was achieved and along came an innovating technique, which used monovalent cations in combination with PEG and a heat shock at 42 ºC to stimulate DNA uptake by yeasts (Ito, Fukuda et al. 1983). Today, different methods exist and are employed according to need. The LiAc/ssDNA/ PEG method (improvement from the original Li^+ method) is the most widely used with highest efficiencies, Electroporation is fast and easy to manage (Dower, Miller et al. 1988), Glass bead method, with lower yields but very simple and low cost (Costanzo and Fox 1988), biolistic transformation (Sanford 1988) and spheroplasting used to transform YACs (Burgers and Percival 1987) are the main techniques used in routine transformations of yeasts.

1.1.10.5. *S. cerevisiae* post-translational modifications

S. cerevisiae, can secrete recombinant proteins to the culture medium, intracellular proteins are usually properly folded, and as other eukaryotes, it is also capable of performing most posttranslational processing typical of mammalian cells. However, differences in glycosylation in proteins expressed in yeast, compared to those of mammalian cells, limit their use. N-glycosylation of proteins produced by yeasts are high-mannose (with more than 3 mannose residues) or hypermannose (more than 6 mannose residues) types, with terminal α-1, 3 linkages (Gellissen 2000). Such forms are very immunogenic to mammals (Ko, Hahm et al. 2002). Moreover, O-glycosylation by yeasts contains only mannose residues (Gellissen 2000). Cell engineering has been utilized for obtaining nonimmunogenic glycoproteins from yeasts. Namely, Chiba et, al. introduced the gene of a α1, 2-mannosidase with an ER retention signal in a *S. cerevisiae* mutant that had disrupted the genes of several mannosyltransferases (Chiba, Suzuki et al. 1998). Unmodified yeasts are suitable for the production of proteins that do not require mammalian-type glycosylation.

Another important molecular aspect of recombinant proteins expressed in yeast are the features of post-translational processing and modification processes specific to yeast, which provide an eukaryotic mechanism to fold and modify proteins. Moreover, intracellular proteins are usually properly folded. As other eukaryotes, yeasts are also capable of performing most posttranslational processing typical of mammalian cells

Unmodified yeasts are suitable for the production of proteins that do not require mammalian-type glycosylation. One of these proteins is insulin, which has been commercially produced in *S. cerevisiae* after enhancing its folding and secretion capacities through genetic engineering (Kjeldsen 2000).

New yeast strains, never described before, with peculiar features are also a good to propel the establishment of an expression system. Recent progress in molecular microbial ecology has shown that microbial diversity is far greater than that reflected in laboratory strain collections (Courtois, Cappellano et al. 2003). Microorganism libraries have huge biodiversity, and recent studies have begun to confirm the enormous potential for discovering new enzymes small molecules or strains with enhanced capacities (Short 1997, Rondon, August et al. 2000). Biodiversity in libraries of wine making Saccharomyces cerevisiae is starting to get interest because of the potential they represent in the wine industry (Valero, Cambon et al. 2007), but this does not mean that other potentialities, such as good hosts to establish as protein expression systems, might advent from screenings.

Recombinant DNA expression constitutes a major approach in gene function studies that naturally complement genetic and genomic research. Well-regulated expression systems provide an

invaluable tool for investigating the cellular roles of novel genes either in their original cellular environment or in specialized host organisms (Porro, Sauer et al. 2005). These systems can be utilized to observe the biological effects of the controlled expression (or lack of it) of a given DNA sequence. Very often they also provide the means to produce and purify a desired gene product, opening the way to a comprehensive analysis and manufacture of proteins of biotechnological interest (Quintero, Maya et al. 2007).

All expression systems have advantages and limitations, and the variety of systems available allow to compensate each other's flaws, overcoming problems in this area of science, to the benefit of mankind. Table 1 and 2 resume the features of the main expression systems used today.

Table 1 Characterization of expression system in common use

Host system	Cell growth	Cost medium	Expression level	Post-translational modifications					
				Protein folding	N-linked Glycosylation	O-linked Glyc.	Phosphorylation	Acetylation	Gamma-carboxylation
E.coli	Rapid (30 min)	Low	High	Refolding usually required	None	No	No	No	No
Yeast	Rapid (90 min)	Low	Low-High	Refolding may be required	High mannose (mostly)	Yes	Yes	Yes	No
Insect cells	Slow (18-24h)	High	Low-High	Proper folding	Simple, no sialic acid	Yes	Yes	Yes	No
Mammalian cells	Slow (24h)	High	Low-moderate	Proper folding	Yes	Yes	Yes	Yes	Yes

Adapted from (Yin J 2007)

Tabela 1 - Applications of expression system in common use

Host system	Advantages	Disadvantages	Common applications
E.coli	Scale-up, low cost and time, easy operation	Protein solubility, may require protein-specific optimization	Structural analysis, antibody generation, funtional assays, proteins interactions
Yeast	Eukaryotic protein processing, scalable up to fermentation simple media requiremens	Fermentation require for very high yield, growth conditions may require optimization	Structural analysis, antibody generation, funtional assays, proteins interactions
Insect cells	Near mammalian protein processing, greater yield than mammalian system	More demanding culture conditions	funtional assays, proteins interactions, antibody generation
Mammalian cells	Highest level protein processing	Relatively low yield, more demanding culture condition	Funtional Assays, proteins interactions, antibody generation

Adapted from (Yin J 2007)

1.2. Objectives

Saccharomyces cerevisiae is a robust microorganism used for decades all over the world for recombinant protein expression. On the other hand, libraries of Wine making *S. cerevisiae* harbour immense potential and huge biodiversity of yeast strains. Regarding this, the main objective of this work was to establish a high throughput screening method to research the biodiversity of a *S. cerevisiae* library isolated from wine practice over the years, and ultimately to isolate potential candidates for further characterization, to be validated as recombinant protein expression systems to enrich the laboratory means, and further away, to use in industry

CHAPTER 2

Materials and methods

2.1. Material and methods

The majority of reagents used in the following procedures were obtained either from Sigma-Aldrich or Merck. DNA inserts were obtained from different plasmids, purchased or availiable in the lab. KanMX4 cassette came from pUG6 vector, *P.pastoris* αMating factor secreting leader (αMF) from pPICKzαA, *K.Lactis's* αMF from pKLAC-1, α-Amylase (AMY) was purchased from *Source Bioscience* (Unigene ID: Mm.439727) and pAMT20 vector (Fig.3A) was kindly provided by *Iha, Hidekatsu* (Iha and Tsurugi 1998). Commercial yeasts used in this work were two lab strains, MATa/MATα W303 (leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11, 15) and CEN.PK2 (MATa/α ura3-52/ura3-52 trp1-289/trp1-289 leu2-3_112/leu2-3_112 his3 Δ1/his3 Δ1 MAL2-8C/MAL2-8C SUC2/SUC2). The Yeast library was created out from Alentejo, Bairrada, Douro and Dão Portuguese Appellations, by isolation of *S. cerevisiae* from wine yards, wine tanks, and fermentations performed in laboratory, during several consecutive years

2.2. Methods

2.2.1. PCR

Majority of PCR reactions were performed using, NEB Phusion® High-Fidelity DNA Polymerase (1 U), 0.5 µM of each primer (forward and reverse), 200 µM dNTP's, 1x Phusion HF buffer. The PCR conditions were: initial denaturation 95 ºC during 5 min, followed by 35 cycles of denaturation 98 ºC during 20 seconds, annealing at 3 ºC under Tm during 30 seconds, extension at 72 ºC at a rate of 30 seconds per kb, a final extension at 72 ºC during 7 min, and stored at 4 ºC until further use. For ligation reactions in pGEM-T Easy, PCR reactions were performed using, GE Healthcare *Taq* DNA Polymerase (1.U), 0.2 µM of each primer, 200 µM dNTP, 1x r*Taq* DNA polymerase buffer. The PCR conditions were: initial denaturation 95 ºC during 2 min, followed by 35 cycles of denaturation 95 ºC during 30 seconds, annealing at 5 ºC under Tm during 30 seconds, extension at 72 ºC during at a rate of 1 min per kb, a final extension at 72 ºC during 10 min, and stored at 4 ºC until further use.

The bands of each amplified DNA insert were gel purified using Invitrogen PureLink® Quick Gel Extraction Kit according to manufacturer's instructions.

Table 2 – Sequence of primers used in sequence-independent cloning. DNA fragment number (Fnº). Primers are represented in 5' to 3' direction. The primer hybridization zone to the correspondent gene is underlined Fnº and numbers above and below the underline correspond to Fnº. Grey boxes represent the overhangs designed in primer. Numbers below and above grey boxes illustrate the matching overlapping sequences. Exception is between insert nº 2 and nº3, where a His-**t**ag designed in the primers (C-terminal tag) is the overlapping sequence. Asterisk represents the stop codon

Gene/DNA insert	Fnº	Primer name	Sequence
α-Mating Factor	1	F1_F	5'-TCA ACT CCA AGC TGA C***TC TAG A***TG AGA TTT CCT TCA ATT TTT ACT G-3' 5 XbaI 1
		F1_R	2 1 5'-GGT CAT ATT GAG CTT CAG CCT CTC TTT TCT C-3'
α-Amylase	2	F2_F	5'-GG CTG AAG CTC AAT ATG ACC CAC ATA CTC AAT ATG-3' 1 2
		F2_R	* His Tag 2 5'-***TCA*** GTG ATG GTG ATG GTG ATG TAT TTT TGA CTC TGC ATG GAT T-3'
ADH1 terminator from pAMT20	3	F3_F	5'-CAT CAC CAT CAC CAT CAC ***TGA*** GTC GAC CTG CAG CCA AGC TAA TT-3' His Tag * 3
		F3_R	4 3 5'-GAG GCA AGC TTA GGT TAT TAC TGA GTA GTA TTT ATT TAA GTA TTG TTT-3'
KanMX4 cassette	4	F4_F	5'-CTC AGT AAT AAC CTA AGC TTG CCT CGT CCC-3' 3 4
		F4_R	4 5 5'-CGT GAG TAT ACT CGA CAC TGG ATG GCG-3'
pAMT20 backbone	5	F5_F	5'-TCC AGT GTC GAG TAT ACT CAC GTG CTC AAT AGT CAC-3' 4 5
		F5_R	1 XbaI 5 5'-GAA GGA AAT CTC A***TC TAG A***GT CAG CTT GGA GTT GAT TGT-3'

Table 3 – Sequence of *Primers used for digestion-ligation cloning*. Primers are represented in 5' to 3' direction. Restriction sites are marked in bold. "F" stands for forward and "R" for reverse

Primer name	Sequence
KanMX4-Nde_F	5'-GGG AAT T**CATATG**AGCTTGCCTCGTCCC-3' NdeI
KanMX4-Bsr_R	BsrGI 5'-GGG AAT TC**T GTA CA**T TAA GGG TTC TCG AGA GCT C-3'
pAMT20-Bsr_F	5'-GGG AAT TC **TGT ACA** TAT ACT CAC GTG CTC AAT AGT CAC-3' BsrGI
pAMT20-Nde_R	NdeI 5´-GGG AAT TC**C ATA TG**T AGG TTA TTA CTG AGT AGT ATT TAT TTA AGT ATT GTT T-3'
αMF-pKlac-Bam-F	5'-**GGATCC** ATG AAA TTC TCT ACT ATA TTA GCC GC-3' BamHI
Amylase-Bgl-F	5´-**AGA TCT** CAA TAT GAC CCA CAT ACT CAA TAT G-3' BglII
Amylase-Sal_R	SalI 5'-**GTC GAC** TCA GTG ATG GTG ATG GTG A-3'

2.2.2. CPEC cloning

All DNA inserts used in CPEC procedures were amplified using New England Biolab (NEB) Phusion® High-Fidelity DNA Polymerase and primers listed in Table 1. For cloning of all PCR products, 50-150 ng linearized vector (longest PCR product) was mixed with all amplified DNA insert to a final ratio of 1:1 or 1:2 (vector:insert). Additionally, Phusion polymerase (1 U), 200 µM dNTP and 1x Phusion HF buffer were added. CPEC conditions were: initial denaturation 95 ºC during 10 min, followed by 25 cycles of denaturation 98 ºC during 20 seconds, annealing at 55 ºC during 30 seconds, extension at 72 ºC at a rate of 30 seconds per kb according to length of the longest piece, a final extension at 72 ºC during 10 min, and correct size product formation confirmed in agarose gel.

2.2.3. Gibson Assembly™ and In-Fusion®HD

Purified PCR products used for these procedures are the same as for CPEC cloning. Gibson assembly was performed in a total volume of 20 µl using: Gibson Assembly Master Mix (1x), 0.2 ρmol for PCR products longer than 1500 bp, and 0.4 ρmol for PCR products with less than 1500 bp Reaction is incubated at 50 ºC during 60 min in a thermocycler, and final product with the correct size confirmed in agarose gel or used for transformation procedure.

For In-fusion, following quantities were used; total volume of 20µl with 1x In-Fusion HD Enzyme Premix, 50 ng for PCR products longer than 1500 bp, 75 ng for PCR products between 1500-500 bp, and 30 ng for PCR products with less than 500 bp Volume reaction is incubated at 50 ºC during 15 min, followed by transformation procedures.

2.2.4. Vector assembly. Restriction-Ligation cloning

All sub-cloning reactions in pGEM®-T Easy Vector used *Taq* polymerase (GE Healthcare) to generate PCR products (primers listed in table 2). Ligation procedures on pGEM-T Easy were performed using NEB T4 DNA Ligase (400 cohesive end U), 1X T4 DNA Ligase Buffer, 25ng pGEM-T east vector, and 75 ng insert DNA. Positive clones were selected by blue/white selection on agar plates with 40 µl of 100 mM IPTG and 25 µl of 50mg/ml X-gal (X-gal and IPTG were spread over the surface of an LB-ampicillin plate and allowed to absorb for 30 minutes at 37°C prior to use.) and analysed by restriction analysis with endonucleases and automated DNA sequencing.

For ligations not involving pGEM-T easy, procedures were as follows: T4 DNA ligase (400 cohesive end U), 1X T4 DNA Ligase Buffer, 1:3 Vector: insert ratio using 100 ng digested vector. Incubation times for all ligations were always performed at 16 ºC O.N. in a water bath.

All restriction analyses were performed using NEB endonucleases enzymes, and NEB buffer selection was set according to NEB Double Digest Finder tool. Incubation were always set at 37 ºC during 3-4 hours, 1 x NEB buffer (1, 2, 3 or 4), restriction enzymes (6U), and BSA (1 x) when needed. In the case of vectors, after double digestion procedure, CIP (5U) were added to the reaction tube, further incubated during 2 hour and purified using Invitrogen PureLink-quick gel extraction kit. DpnI assays procedures were performed as follows: DpnI enzyme (40 U) and 1x NEB buffer, followed by liquid purification using Invitrogen PureLink-quick gel extraction kit. Automated sequencing was performed either in STAB Vida of MACROGEN.

2.2.5. Saccharomyces deep well preparation

Prior to transformation, Saccharomyces library was prepared as following: First, each strain from a main stock (kept at -80 ºC) was streaked on YPD agar and grown $_{\text{O.N.}}$. at 30 ºC. Then, a single colony from each strain was picked and inoculated in a single well (with 800 uL of YPD medium) of a 96 deep well plate, so each well contains a diferent strain. After completed, an over nigh incubation at 30 ºC in a static incubator (Binder BD115 #05-91700) follows, cells are counted (using a Automated cell counter TC10® from Bio Rad) and a new deep well is prepared with 1*10^5 cells per well and stored at 4 ºC until further use.

2.2.6. Carrier DNA preparation

Carrier DNA (2mg/ml) was extracted from liofilized salmon testis. Dissolved DNA (10 mg/ml), it is incubated with Proteinase K (2 mg/ml) for 3-4 h, and extracted twice with a 24:1 chloroform: isoamyl alcohol solution, followed by precipitation at a 1:10 ratio with a 5 M NaCl solution, 2.5 volumes of 96 % ethanol and let to rest at 4 ºC O.N. After precipitation DNA is centrifuged at 14000 x g for 10 min, washed with ethanol 70 % (to remove excess salt), centrifuged again to remove ethanol and dryed at 37 ºC. DNA is resuspended using ddH_2O at 2mg/mL concentration, aliquatoed, boiled at 98 ºC for 20 minuts and stored at -20 ºC. Prior to use, aliquots are boiled for 5 minuts and chilled on ice.

2.2.7. Yeast medium sclae tranformation

Several colonies, with less than two weeks, with 2-3mm diameter were inoculated in 1mL YPD, vigorously vortexed and transferred to an Erlenmeyer flask containing 50 mL YPD. Culture medium was grown O.N. in a rotary shaker (Innova 44 - incubator shaker series) at 30 ºC and 200 rpm. The next day, culture was diluted in 100 mL YPD medium to an OD_{600} of 0.2 and grown with agitation at 30 ºC and 200 rpm approximately 3-5 hours. When OD_{600} reached 0.5-0.6, cells were pelleted at 1000 x g for 5 minutes at room temperature. Supernatant was carefully removed, each pellet resuspended in a total of 12.5 ml of freshly prepared TE/LiAc buffer [(10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100mM lithium acetate (LiAC)], and pooled into one tube. Cells were pelleted again, resuspended in 500 µl of TE/LiAc buffer, and aliquot (150 µl) in Eppendorf tubes.

For transformation procedures 0.5 µg plasmid DNA is added to a single aliquot, following 600 µl of 40 % of PEG solution (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100mM LiAc, 40 % PEG). Cells are carefully mixed, and incubated for 45 min in a rotary shaker (Innova 44 - incubator shaker series) at 30 ºC and 200 rpm. After incubation DMSO is added (final concentration of 10 % v/v), tubes are gently mixed by inversion. Heat shock was performed at 42 ºC in a water bath for 1 h, and cells chilled on ice. 100 or 200 µl of the cell mixture is plated on selective plate and placed at 30 ºC in a static incubator for 3-5 days until large colonies are formed. Alternatively, before plating a Recovery step at 30 ºC and 200 rpm in a rotary shaker may be performed to increase number of transformants.

2.2.8. 96-well Microtiter tranformation

Yeast transformation was performed according to (Gietz and Schiestl 2007) with some modifications. Deepwell plates were prepared as described in previous section were grown overnight and 200 µl off each well were transferred to a 96 well microtiter plates with the help of a multichannel pipette. This results in 10^6 – 10^7 order of cells per well. Cells were pelleted at 2000 x g for 30 min and supernatant was removed with a single shake. All single shake procedures should be performed into a large recipient with paper covering the bottom to avoid splashing.

When peparing the transformation mix without PEG, plasmid DNA was used in 1 - 1.7µg per well, all ressuspending steps were performed at 400 rpm using a microtiter box in a rotary shaker (Infors HT Ecotron - AG-CH4103 Bottmingen) at 30 ºC. Heat shock was performed fo 3 hours, at 42 ºC with multiwell plates inside zipped plastic bags. After heat shock the cells were plated at 1500 x g for 10 min and supernatant removed with a single shake. Cells were resuspended in 50 ul 35 % PEG solution and rested for 1h. After 1 hour, 150 ul YPD without selection marker was added to each well and incubated O.N. at 30 ºC. Finally the 200 µl from the 96-well microtiter plate are transferred to a deep well with 600 µl YPD + 200 ug/mL of the antibiotic G418 and incubated at 30 ºC for 3-5 days.

2.2.9. High throughput activity assay

High troughput assays were performed in a Sciclone ALH 3000 Workstation-Caliper LifeScience robot, with a 96 pin tool. Pin tool was sterilized as followed : First, dippind 3 x in 10 % beleach and dryed by touching a lint free blotting paper, followed by dipping in a water tank, sonicated for 50 seconds, removed and dipped 3 x in the same water (while sonication is on) and dryed out in lint free blotting paper. Finally, the 96 pin tool is dipped 3 x in 96% ethanol and air dryed. For halo assays, 1 µl of culture medium were transferred with the sterelized pin tool to a single well (Nunc® OmniTray Single-Well Plates with Lids, Thermo Scientific) containing YPD agar supplemented with 2% starch, 200 µg/ml G418 and grown O.N. Plates were stained with a solution of 70 % lugol to see halo formation. Lugol contains iodine which is capable of inserting itself inside the 3 dimensional structures of starch mixed in the solid medium. If starch is degraded, iodine is unable to be retained in solid medium and no staing occurs in the area.

2.2.10. Small-scale screening for protein expression

In the small-scale screening of protein expression, a single colony with 2-3 mm diameter was scraped from the plate and resuspended in 10 mL YPD medium supplemented with 200 µg/mL G418, and grown in a rotary shaker at 30 ºC and 200 rpm (Innova 44 - incubator shaker series) $_{O.N.}$.. The next day, starter culture was diluted in 30 mL YPD medium supplemented with 200 µg/mL G418 to an O.D. of 0.2 in 100 ml baffled Erlenmeyer flasks. The first sample (250 µl) was collected when OD_{600}. reached 0.8, and subsequent samples collected two times a day, in the morning and evening. Samples were collected during 72 hours and were immediately denatured in 6x denaturing buffer and stored at room temperature until further use.

2.2.11. Protein Expression and Purification

For large-scale cultures, first a single colony (never older than two weeks) was inoculed into 20 mL YPD medium containing 200 µg/µl G418 and grownh for 6 to 8 hours. Then the wole inoculum was poured into 30 ml YPD medium (with 200µg/µl G418) and grown with agitation at 30 ºC and 200 rpm. Next day, the culture was diluted to an O.D. of 0.2 into 4 x 2.5 L Thomson Ultra Yield Flasks containing 1 L YPD medium each with 60 µg/µl G418, and let to grow during 120 hours in a rotary shaker at 30 ºC and 200rpm.

After culture was grown, cells were pelleted by centrifugation at 5000 x g for 20 min at 4 ºC, and supernatant was carefully removed. Each pellet was resuspended in a total of 10 mL of 20mM phosphate pH 7.5 buffer with 100 mM NaCl and pooled togheter. Resuspended cells were lysed in a Emulsiflex-C3 from Avestin (3 passages at 21000 psi), the lysed cell extract was ultracentrifuged at 104350.1 x g and supernatant was stored at -20 ºC.

The supernantant was concentrated by Tangential Flow Filtration (Sartocon® Slice 200) using a 10 kDa membrane down to 200 ml and ultracentrifuged at 104350.1 x g. Protein was precipitated with 90 % ammonium sulphate in an ice bath (0 ºC) with agitation (salt concentration was calcluated using http://encorbio.com/protocols/AM-SO4.htm calculator), centrifuged at 10000 x g for 20 min, resuspended in 10 ml of 20 mM phosphate pH 7.5 buffer with 100 mM NaCl and filtred with 0.2 µm membrane. Filtred fraction was applied to size exclusion chromatography Superdex 200 26/60 prep

grade (GE Healthcare Life Sciences), previously equilibrated in 20 mM phosphate pH 7.5 buffer with 100 mM NaCl. The column was operated in a AKTA FPLC system (GE healthcare Life sciences) at a flow of 2 ml/min at room temperature, and 10 mL fractions were colected and immediately stored at 4 ºC until further use.

2.2.12. SDS-PAGE and Western Blot analysis

Protein samples were loaded with 6 x loading buffer (0.35 M Tris-HCl/0.28% SDS buffer pH 6.8, 30% glycerol, 10% SDS, 0.6 M DTT, 0.012% Bromophenol Blue), denatured at 90 ºC for 10 minutes and separated by SDS-PAGE. . The SDS-PAGE was done in 12.5 % polyacrylamide gels and run at 150 V at room tempeature, until the dye (Bromophenol blue) exits the gel, using a MiniProtean 3 system (Bio-Rad). The running buffer was 100 mM Tris, 100 mM Bicine with 0.1 % SDS. The gels were stained with a solution containing 50 % methanol, 10% acetic acid and 0.2 % Coomassie Brilliant blue. The removal of the excess staining was accomplished by incubation with a solution containing 25% methanol and 5% acetic acid under agitation at room temperature.

For Western blot analysis, protein separeted by SDS-PAGE were transferred to a PDVF membrane (Roche), previously activated in methanol. The electrotransferences were performed in a Trans-Blot® Electrophoretic Transfer Cell (bio-rad) or in Bio-Rad Mini Protean® tetra system. For Trans-Blot® Electrophoretic Transfer Cell, overnight electrotransferences at 40 V, 11 ºC were performed using 25 mM Tris, 192 mM Glycine and 20 % methanol. Electrotransferences in the mini system were performed either for 2 hours (electrotransference at 100 V, 11 ºC using 10 mM CAPS pH 11 with 10 % methanol) or or overninght at 30 V, 11 ºC using 25 mM Tris, 192 mM glycine and 20 % methanol.

After transfer, PDVF membranes were blocked with TBST buffer (20 mM Tris, 137 mM NaCl, 0.1 % Tween, pH 7.6) containing 5 % skim milk for 1 hour. The membrane were then incubated for another hour with the primary antibody against His-Tag (THE™ His Tag Antibody, mAb, Mouse, GenScript) diluted in TBST 0.5% milk solution (1:10000 dilution). After the first incubation, the membrane was washed with TBST 0.5% milk solution at least 7 times for 5 minutes before incubation with secondary antibody. Incubation with the secondary antibody also lasts 1 hour, in the same solution (TBST 0.5% milk). Secondary antibody used is a Anti-Mouse IgG + IgM alkaline phosphatase linked whole antibody (from goat; Amersham Biosciences) with the same dilution (1:10000). Membranes were washed at least 7 times for 5 minuts with TBST buffer, then developed with ECF ™ substrate (GE Healthcare) for 5 minuts, and revealed in a Molecular Imager FX (Bio-Rad) scanner.

2.2.13. Enzyme activity assay

Agar plates were prepared with 50 mM Hepes pH 6.9 buffer, 5 mM $CaCl_2$, 2 % starch and 20 g/L agar. After solidification, holes were opened in the agar with the help of a 60 mL plastic syringe (without needle) and 15 to 20 µl samples loaded into each holes. Plates were incubated overnight at 30 ºC and stained with 70% lugol solution. Activity was measured as the ratio of the halo/colony size.

CHAPTER 3

Results and Discussion

3.1. Vector assembly

To screen a wine yeast library with vast biodiversity in *S. cerevisiae* strains, for candidate hosts to establish a recombinant protein expression system, an effective and high throughput assay had to be assembled. The screening needed to simple, fast, and one of the best strategies to perform this is testing activity on agar plates. This method provides a quantitative assay, simple, inexpensive, straight forward method to assess the activity of enzymes (Vermelho, Meirelles et al. 1996). Iodine, contained in I_2KI solution (e.g. Lugol), is capable of inserting itself between the secondary structures of starch mixed in agar plates, acquiring a dark blue colour. But if starch is degraded, iodine is not held in the agar because the secondary structure of starch is lost, and no staining occurs. Salivary α-Amylase has such capacity and can be adapted to a high throughput screening, when for instance, is secreted from recombinant yeast grown on solid medium. Variable sized halos will form around colonies, depending on the protein production capacity of yeasts, allowing for a simple way to differentiate potential individuals to establish an expression system.

Mus musculus salivary α-amylase mature form (without native signal peptide) was the enzyme chosen for the activity assay. The α amylase family contains a characteristic catalytic $(\beta/\alpha)_8$-barrel domain, depends on Ca^{2+} ions to be active and has two potential glycosylation sites. It has been previously successfully expressed and secreted in an active form from yeasts (Thomsen 1983, Kato, Ishibashi et al. 2001) and presents itself as two bands in SDS-PAGE, 57 and 60 kDa for non-and glycosylated form respectively.

To complete the screening method, a selective marker is needed to exclude non transformed strains. Normally, selection in yeasts is based in complementation for defective amino acids (auxotrophic selection), but wine yeast are normally prototrophic (Hammond 1995, Benítez, Gasent-Ramírez et al. 1996, Guijo, Mauricio et al. 1997, Pretorius 2000, Dequin 2001) and thus, a dominant marker that confers resistance to an antibiotic was used. KanMX4 cassette confers resistance to Kanamycin in *E. coli* and to G418 antibiotic in yeasts (Agaphonov, Romanova et al. 2010).

Finally, pAMT20 is the vector backbone which will incorporate these genes used in the high throughput screening. It is a 2µm based *E. coli/*yeast shuttle vector with high copy number in both organisms, it has a strong glycolytic promoter (short version of ADH1) and a terminator to stop transcription (tADH1). An inducible promoter would not be adequate for a high throughput screening because of technical difficulties of the process *per se*. The short version of ADH1 promoter, is reported to be fully activated only when glucose is almost depleted from medium, but remains active through

the ethanol phase (Ruohonen, Aalto et al. 1995). It is also reported that the final yield of protein is greater using this version of the promoter, instead of the middle of long version (Monfort, Finger et al. 1999). The modifications in this vector implicated the substitution of the original auxotrophic marker for a dominant one, TRP1 for KanMX4 respectively, the main enzyme α-amylase, and the secretion leader of pheromone α-mating factor (αMF).

The DNA sequences of the selection marker, the enzyme, and the secretion leader encountered themselves in different vectors, and needed to by amplified and purified by PCR and gel extraction respectively before cloning in pAMT20.

3.1.1. PCR. Sequence-independent cloning

The traditional cloning strategies involving sequential steps of DNA amplification, digestion, ligation, subsequent transformation, purification, and restriction analysis are labour intensive and time consuming because only one insert at a time can be cloned on the desired vector. Seamless cloning and assembly techniques allow the assembling of more than one insert at a time, and thus were chosen to assemble five PCR products into one final vector: Linearized pAMT20, KanMX4, α-Amylase, αMF, and terminator (see Figure 3 for illustration).

The constructs were amplified by PCR using the circular plasmids as templates. Primers used in PCR reactions (Table 1) to amplify each construct produce overhangs with overlapping sequences of about 18 to 25 base pairs essential for homologous recombinant techniques. Good amplification yields were obtained (Figure 2), and each band was gel purified, and stored at -20 ºC until further use.

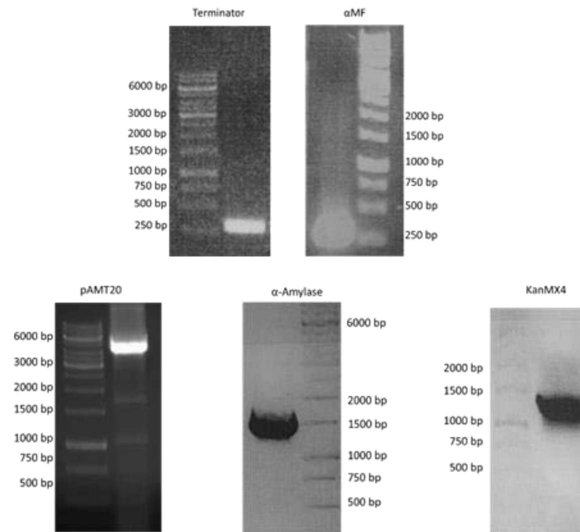

Figure 2 - ***Agarose gel electrophoresis of PCR products.*** *Individual PCR reactions to amplify each of the DNA inserts. Names above Gel lanes correspond the amplified construct, and numbers correspond to 1kb DNA ladder band sizes. Terminator and αMF have around 250 bp; Linearized pAMT20 around 4500 bp; αAmylase and KanMX4 cassette have around 1500 bp.*

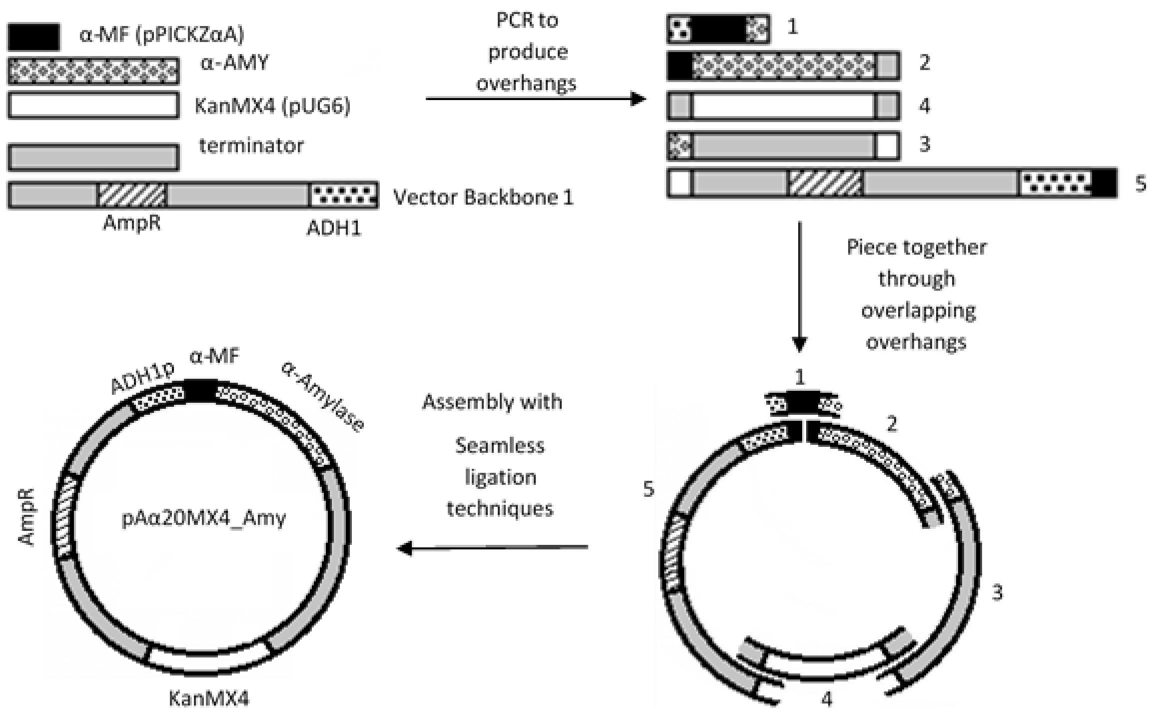

Figure 3– ***Schematic illustration of Seamless ligation cloning principle.*** Figure illustrates how overlapping regions work in the final assemble. These overlapping ends are created simply with primer overhangs.

3.1.2. Circular Extension polymerase cloning - CPEC

Circular Polymerase Extension Cloning (CPEC) is a sequence-independent cloning and is largely based on homologous recombination. For synthetic biology applications involving high-complexity or multi-fragment cloning, sequence-dependent methods are generally inconvenient because they require unique and specific sites in both the insert and the vector in order to generate the initial plasmids. For this reason, the more flexible sequence-independent cloning methods are preferred. However such methods usually require generating complementary single-stranded overhangs in both the insert and vector fragments. CPEC relies entirely on the polymerase extension mechanism, and overlapping regions between insert and the vector fragments after annealing, work as primers and are extended with a polymerase to form a complete circular plasmid and therefore its name (Figure 4). If assembling involves multiple Inserts, overlapping regions need to be designed with approximately same melting temperatures, between 60 and 70 ºC, and cycling steps need to be augmented (up to 25 cycles).

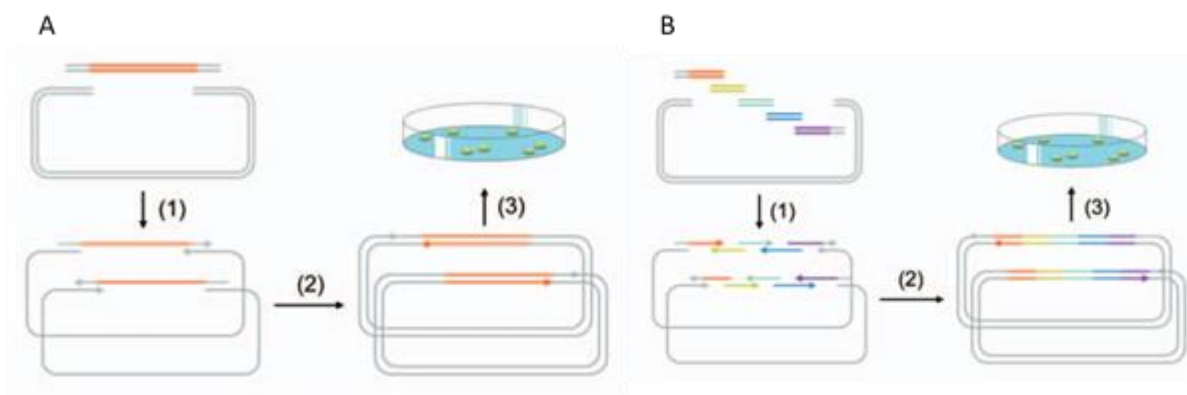

Figure 4 - ***Gene cloning using CPEC***. *A- Schematic representation the purposed mechanism for cloning individual gene. The vector and the insert share overlapping regions at the ends. After denaturation and annealing (step 1), the hybridyzed insert and vector extend using each other as template until they complete a full circle and reach their own ends (step 2). The final completely assembled vector has two nicks, one on each end, marked by the position of the arrow head. They can be used for transformation (step 3). B – Schematic representation of multi-way CPEC. Any of the neighbouring fragments share an overlapping region with identical*

Assembling of all five PCR products in a single reaction was unsuccessful, and after several tries no band could be visualized on agarose gel, neither positive clones generated. Since five overlapping regions are present, and need to anneal at the same time, if only one fragment fails to anneal, final product formation in subsequent cycles will be reduced or inexistent. The first modification realized to produce results was to experiment several annealing temperatures, from 55 until 72 ºC, with and

without annealing ramps (0.1 ºCs^{-1}), but all produced no positive results. Vector concentration increase was also assessed, but the highest concentration used in a single reaction was 150 ng and this could be limiting the success rate of whole procedure. Augmenting vector concentration could lead to positive results.

Likewise assembling five pieces together with our constructs could be too demanding, so fragment numbers were reduced for each reaction. For example, first introducing PCR amplified KanMX4 cassette into the linearized vector (Primers F3_R and F5_F to linearize the vector) generating pA20MX4. After in vivo replication and purification, the next step would join PCR products αMF, αAMY and pA20MX4 in one CPEC reaction, thus assembling less than 4 DNA fragments at a time. The other way around was also used as strategy to circumvent the multi-way CPEC reaction. First αMF, αAMY and pAMT20 assembling through CPEC, *E. coli* replication, purification, and final reaction with pAMT20 (with both α inserts cloned) and KanMX4 PCR product. Again annealing ramps, variation of cycle number (5, 10, 20, and 25), and different polymerases tested produced no other results than negative clones.

The negative clones, when analysed through restriction pointed always to the same issue, which was the fact that only the native vector was being internalized into *E .coli* cells. This native pAMT20 contamination comes probably from the linearized vector sample which was gel purified, because the PCR product and native vector have similar molecular weight in all CEPC assembling. Native vector was probably gel purified along with the PCR amplification linear vector justifying its appearance in clones transformed with CPEC reactions. The use of DpnI could eliminate this contaminant, and increase the frequency of positive clones, but other techniques, such as Gibson, In-fusion and restriction ligation cloning, were set forward to circumvent the problem and CPEC left behind.

3.1.3. Gibson and In-Fusion

These techniques resemble CPEC, in that they depend as well on overlapping overhangs in DNA fragments, but differ in the mechanism used to soe DNA chains together. Gibson Assembly™ and In-Fusion®HD are isothermal single-reaction methods to assemble multiple DNA fragments, which depend on the concerted action of a exonuclease, a DNA polymerase and a ligase (ligase is only present in Gibson). It basically starts with the exonuclease which recesses the double stranded DNA fragments in the solution, yielding single-stranded DNA overhangs that specifically anneal (due to overlapping

sequences), a polymerase which completes the missing strand, and a ligase that covalently joins the DNA fragments (In-fusion relies on in-vivo reparation of the nicked strands, thus it does not contain a ligase) (Figure 5). The reactions are incubated during 60 and 15 min, for Gibson and In-fusion respectively, in a thermocycler and visualized on gel for or transformed directly in competent cells.

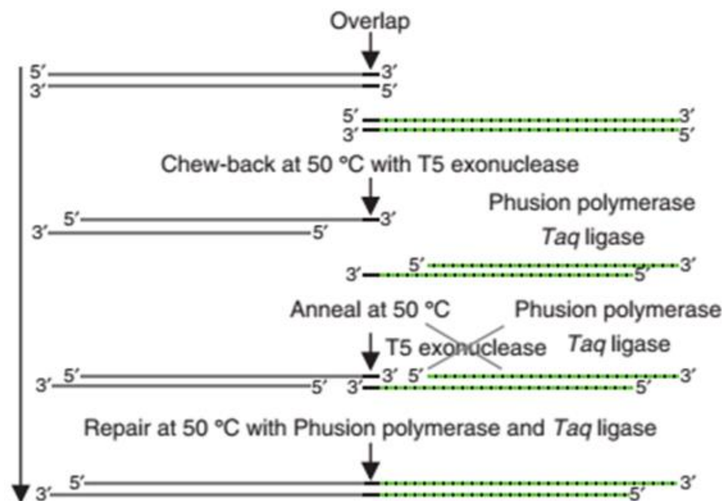

Figure 5 - ***Gibson™ One-step isothermal in vitro recombination.*** *Two adjacent DNA fragments sharing terminal sequence overlaps (black) were joined into one covalently sealed molecule in a one-step isothermal reaction. T5 exonuclease removed nucleotides from the 5′ ends of double stranded DNA molecules, complementary single-stranded DNA overhangs annealed, Phusion DNA polymerase filled the gaps and Taq DNA ligase sealed the nicks. Adapted from (Gibson, Young et al. 2009)*

Similar to CPEC, five fragment assemble turned out to be complicated and strategies, which involved less fragments per reaction, were assessed. In-fusion never produced bands on agarose gel or positive clones in any strategy, but with Gibson, when assembling five DNA pieces, a slight band near 8000 bp was generated twice (in to consecutive experiments). This would be exactly the product we were looking for, so bands were gel purified and used to transform several different competent cells. Disappointingly, no transformants generated, not even electrocompetent cells. Reaction was repeated one last time, but this time reaction mixture directly to transform competent cells, and only negative clones, harbouring the native plasmid, were transformed. Competence of cells might have influenced the lack of positive transformants, or again DNA concentrations should be increased even further (until 200 ng).

At this stage, and to drop out of this negative results cycle, digestion-ligation cloning was set at full speed to construct the vector the fastest way possible, to recover the time lost, and the seamless cloning techniques set aside.

3.1.4. Digestion-Ligation cloning

Vector assembly was successfully achieved using traditional cloning techniques. Figure 8 represents the schematic assembling steps to help understand the full vector assemble. At this stage, primers listed in table 2 and Taq polymerase were used to amplify DNA sequences, PCR products were gel purified and ligations performed using pGEM-T. After restriction analysis and automated sequencing confirmation (Figure 7), Inserts were double digested from pGEM-T easy and gel purified.

Figure 6 does not discriminate sub-cloning steps in pGEM-T easy, but the inserts sub-cloned in this vector were KanMX4 and α-Amylase. KanMX4 was first amplified to generate NdeI/BsrGI restriction sites, gel purified, cloned in pGEM-T easy, confirmed by restriction analysis and sequenced. After confirmation, KanMX4 inserted in pGEM-T easy was digested, separated on agarose gel and band purified. Vector pAMT20 was also amplified to generate same restriction sites through PCR, treated with DpnI and purified using PureLink-quick gel extraction kit. KanMX4 DNA insert with cohesive ends was ligated to digested vector and again analysed by restriction analysis (Figure 6A) and automated sequencing. The same procedure was used to clone α-Amylase in pGEM-T easy. But first amylase was engineered by PCR reactions to introduce a His-Tag in the C-terminal, and a second PCR to introduce restriction sites (BglII/SalI). The choice of using α-MF from pKLAC1 (*K. lactis* pheromone secreting leader), instead of α-MF from pPICZαA (*S. cerevisiae* pheromone secreting leader) was the compatibility of restriction sites between pAMT20 and pKLAC1. Double digested amylase insert was ligated to vector pKLAC1 (previously double digested, treated with CIP, followed by liquid purification) and confirmed as described previously (Figure 6B).

The final step in this strategy was to amplify the α-MF_α-Amylase construct (simultaneously producing BamHI/SalI restriction zones) from pKLAC1_AMY (Figure 6 C), through a PCR reaction and to clone this sequence into pA20MX4 (pAMT20 already containing KanMX4 cassette), confirmed by restriction and automated sequence analysis (Figure 6C, D and 7). After all this laborious process, the final vector, named pAα20MX4_AMY was used for transformation procedures using Yeast lab strains, to assess if it is working properly, to proceed to liquid protein expression and high throughput transformation of the yeast library.

Figure 7 show the automated sequence analysis of sequences that need to be faithfully inserted, namely in-frame cloning of α-Amylase with αMF (Figure 7A), Histidine Tag (essential for later purification procedures) and stop codon (Figure 7B), and selection marker KanMX4 cassette (Figure 7C).

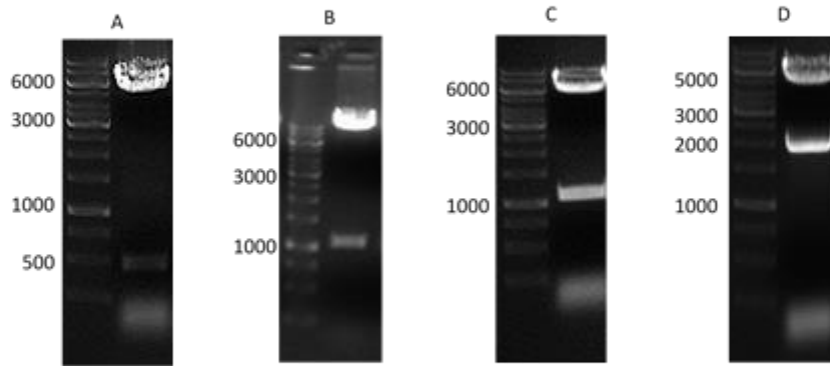

Figure 6 - ***Agarose gel electrophoresis of Restriction analysis for vector assembling.*** *In all examples, left lanes correspond to 1kb DNA ladder. A – KanMX4/pAMT20 restriction analysis with PstI (467 and 5406 bp); B – α-Amylase/pKLAC1 restriction analysis with Msc1 and XhoI (1000 and 9000bp); C – Final assembled vector restriction analysis with ClaI (1113 and 6438 bp); D – 2nd Confirmation of final assembled vector restriction analysis with NcoI (2000 and 5567 bp)*

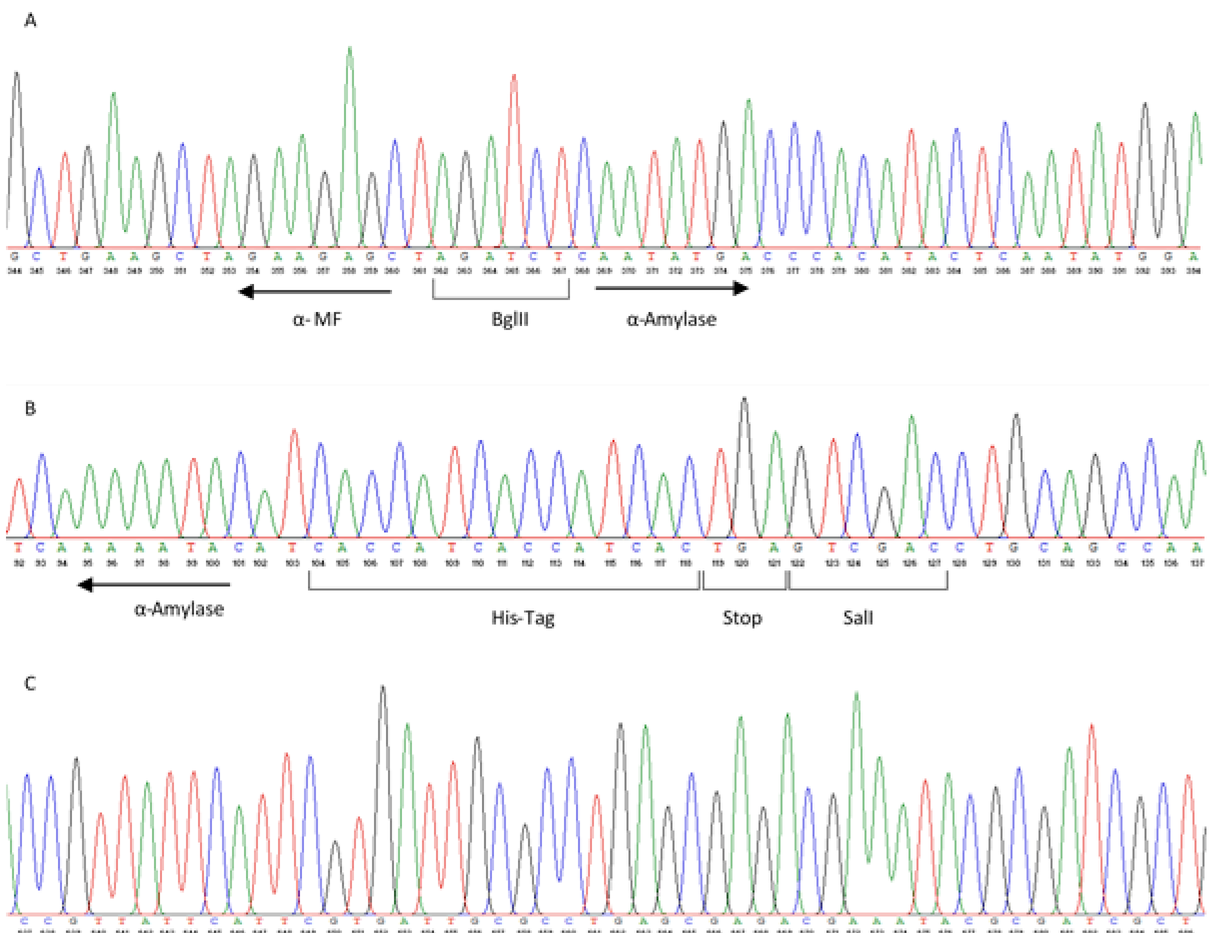

Figure 7 - ***Part of sequencing electrophorectogram of the final vector assembled.*** *A – Sequenced zone of αMF/α-amylase interface, confirms in-frame cloning of α-amylase with αMF in final vector. B – Sequenced zone harbouring the His-tag and stop codon in C-terminal of α-amylase. C – Part of sequencing electrophorectogram of KanMX4 cassette.*

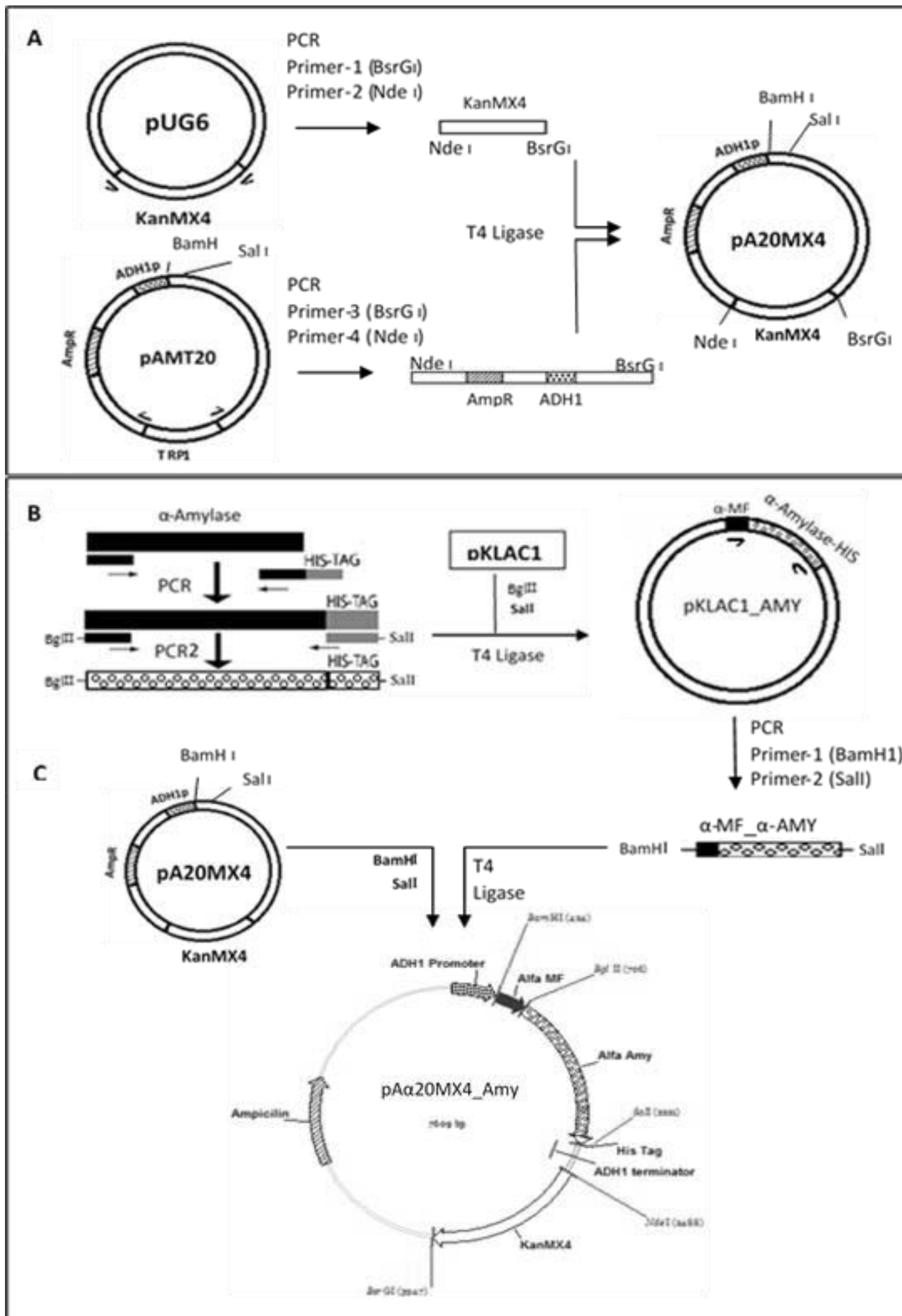

Figure 8 – Vector assembly with Digestion-ligation cloning. A – KanMX4 insertion in pAMT20. Both molecules are first PCR amplified to insert NdeI/BsrGi sites. B – His-tag and restriction sites are introduced in amylase with two consecutive PCR and cloned into pKLAC. Insert αMF αAMY is PCR amplified to insert restriction sites BamHI/SalI. C – final Insert is cloned in pAMT20.

3.2. S. cerevisiae transformation – LiAc/SS-carrier/PEG method

The introduction of exogenous DNA into yeast by transformation has become an essential technique in molecular biology. Transformation is used to investigate the genomics and proteomics of yeast itself and also when yeast is employed as a system to study the genes and gene products of other organisms (Ideker, Galitski et al. 2001). Intact yeast cells can be transformed by a number of procedures: the lithium acetate/single-stranded carrier DNA/polyethylene glycol (LiAc/SS-Carrier DNA/PEG) method (Gietz, Schiestl et al. 1995), electroporation (Becker and Guarente 1991), agitation with glass beads (Costanzo and Fox 1988), and bombardment with DNA-coated microprojectiles (Johnston and DeVit 1996). The most widely used and with easy and practical application in the high throughput transformation is the LiAc PEG (Gietz and Schiestl 2007).

The first attempt to transform yeast with the pAα20MX4_Amy was accomplished with the lab strains, W303 and CEN.PK2 using a medium scale transformation protocol.

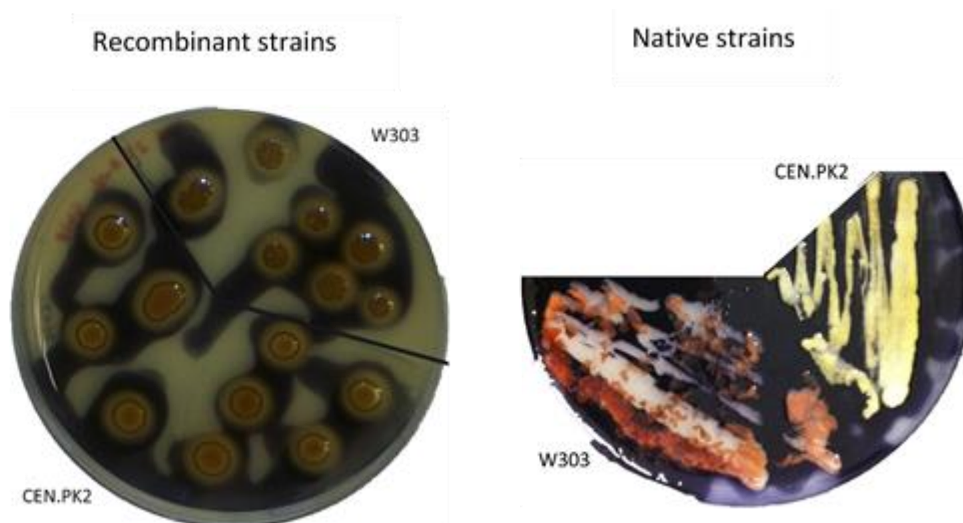

Figure 9 - Lab strains plated on YPD + 2% starch and stained with lugol solution. On the left side are the transformed lab strains, with clear hallo formation around colonies. Negative control on the right side. Native strains do not form halos. The white area visible in the W303 are not degradation of starch. More lugol was added afterwards and those areas stained as well (Data not shown). When stained, the plate was shaken vigorously and colonies were displaced. Adding more lugol stained that area (Data not shown)

This assay shows that α-Amylase is secreted in an active form, thus the vector is working as well as the activity assay. The white area on the negative controls are not starch degradation. More lugol was added later, and those area turned black. If starch degradation had occurred (for example, the strain secretes and endogenous amylase) then the agar + starch medium would not stain, no matter

how much lugol added. The problem results from vigorous plate shaking to distribute lugol all over the area.

Proving that the vector was working allows us to move to protein expression in liquid medium using lab strains. The purpose is to establish a threshold for protein level production in a lab strain, and later compare these levels produced by a selected candidate from the Yeast library. Before moving into large-scale, a small scale expression was performed to assess the time of expression to yield sufficient amount of protein.

3.3. Small scale protein expression in lab strains

Time course for protein expression performed with CEN.PK2 and W303 lab strains was started at the same time by pré-inoculum in a small volume medium. Next day, culture medium was diluted until 0.2, and samples collections started when O.D reached 0.8. Due to imparity in growth rates between lab strains, W303 (which grows at slower rates) time course was delayed 12 hours.

Samples were collected for 72 h (60h in W303), immediately denatured using 6x loading buffer and stored at room temperature. Salivary α-Amylase has 57kDA and is reported in yeasts to be glycosylated in yeast, so two bands could appear in SDS-PAGE and western blot. Figure 10 shows the time course of both lab strains, and some protein in present in the supernatant, but these proteins are probably endogenous secreted proteins or from cell lysis during growth, because Western-blot analysis is clear. Protein at this stage is too diluted, and therefore is not detectable even with a Western-blot analysis.

Next thing to assess was to verify the intracellular content of the collected samples. In theory, most of the amylase produced should be secreted, but apparently this was not the case, and more signal is detected from intracellular fractions than the supernatant (Figure11). We conclude that α-Amylase is being retained inside the cells. Thomsen (1983), used ADH1 promoter and α-amylase native signal peptide to secrete *Mus musculus* α-Amylase produced in *S.cerevisiae.* They reported with this signal peptide over 90% exportation of amylase to the extracellular medium and practically no protein accumulated inside cells. Using only the signal peptide of *K. lactis* could increase secretion levels, but reported literature says otherwise, and excluding some exceptions, α-prepro region is needed to efficiently secrete proteins in *S. cerevisiae* (Li, Xu et al. 2002).

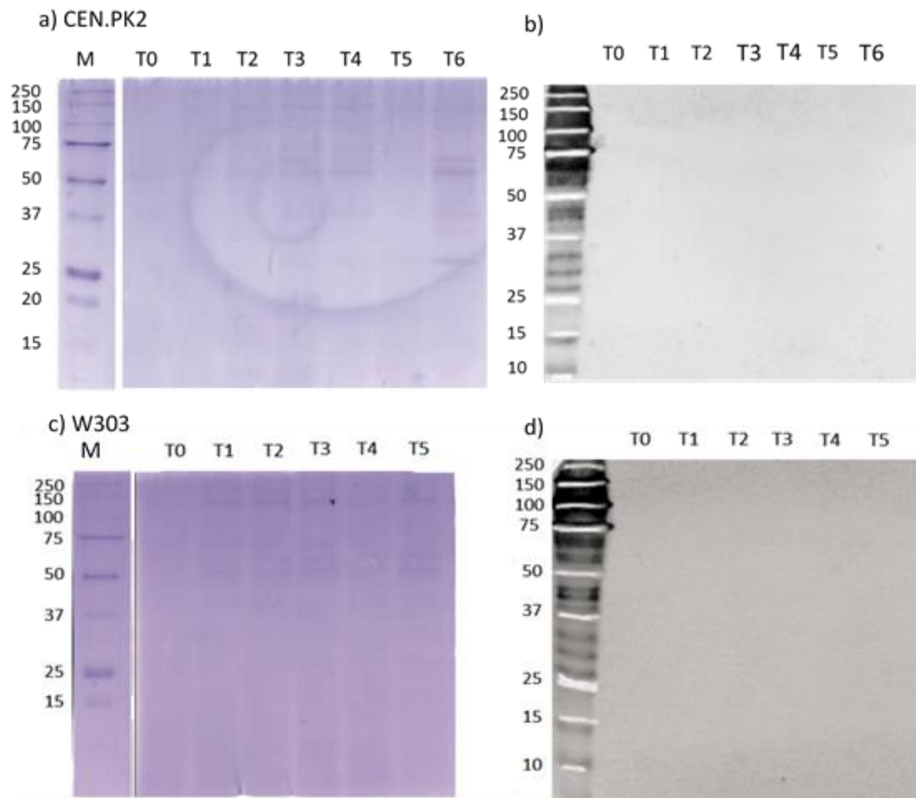

Figure 10 - P**rotein expression timecourse in lab strains analysis.** Only supernatant of culture medium is analysed. 250 µl samples were denatured with 6x loading buffer and 50 µl applied to each lane either for SDS-PAGE followed by Coomassie staining of for western blot analysis. In both cases 12.5% acrylamide gels were used. The number above lanes correspond to times of sample collection. M – molecular weight marker; T0 – OD_{600} 0.8; T1 – 12h; T2 – 24h; T3 – 36h; T4 – 48h; T5 – 60h; T6 – 72h. a) SDS PAGE followed with Coomassie staining. b) Western blot analysis with anti-His tag antibody of the same SDS-PAGE samples of CEN.PK2. c) and d) correspond to SDS-PAGE and Western-blot from W303 respectively.

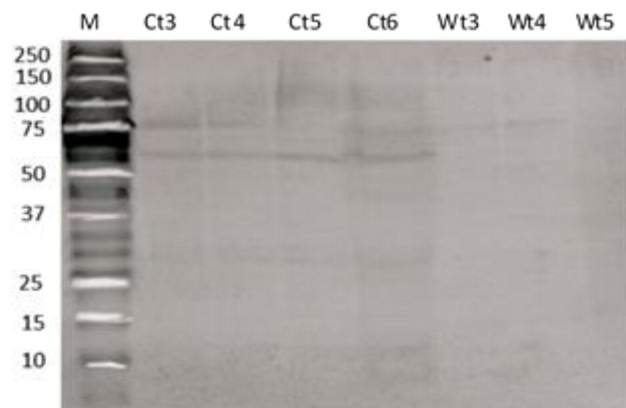

Figure 11 -***Western blot analysis of intracellular protein content.*** *Screening of intracellular samples of both lab strains, CEN.PK2 and W303 for recombinant α-Amylase. Samples were denatured with 6x loading buffer and 50 µl applied to each lane for western blot analysis. 12.5% acrylamide gels were used. The letters above lanes correspond to the strain (C for CEN.PK2, and W for W303) and numbers correspond to times of sample collection. M – Molecular weight marker T3 – 36h; T4 – 48h; T5 – 60h; T6 – 72h*

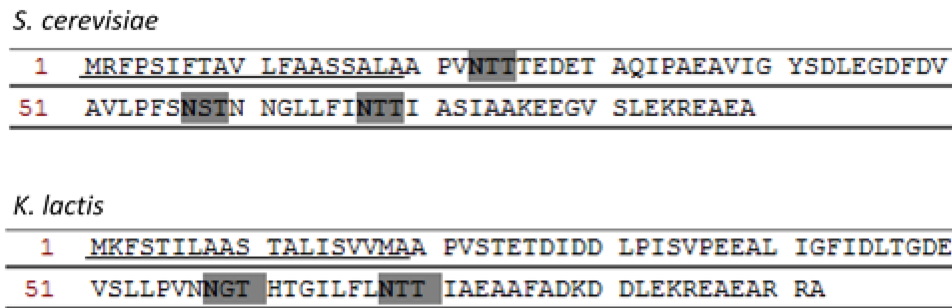

Figure 12 - ***Alfa mating factor secretion leader sequence in yeasts.*** *Sequence of both yeast prepro-αfactor. Underlined are the signal peptides cleaved early in the secretory pathway. Light grey boxes represent N-glycosylation, and asparagine is evidenced in bolt.*

The secretion leader of S. cerevisiae is composed of 89 a.a., that consists of 19 a.a signal peptide and 64 a.a pro-region containing 3 glycosylation sites The nascent polypeptide is translocated into the endoplasmic reticulum (ER), where the signal sequence is cleaved to produce pro-α-factor (Waters, Evans et al. 1988), and core N-linked carbohydrate is added to the three sites within the pro region. *K-lactis* leader only possesses 2 glycosylation sites instead of the tree from *S. cerevisiae* (Figure 12). Prepro region of αMF in *saccharomyces* and *K. lactis* only share 48% of similarity, which is quite low.

Caplan et al. studied the effect of alterations in glycosylation and sequence of pro region from αMF. Their results showed that deleting one two or all three glycosylation sites in the pro-region lead to increasing reduction of alpha factor secretion respectively (with concomitant intracellular accumulation). But some alpha factor could still be secreted. Also, when they altered the pro region sequence (either with insertions or deletions) the same effect as eliminating glycosylation sites occurred (Caplan, Green et al. 1991). Adding this to our information, both glycosylation's and pro region sequence of prepro-αfactor differ in *S. cerevisiae* and *K. lactis*. Together this differences are probably responsible for delaying α-amylase secretion, resulting in intracellular accumulation. Due to time constrains, exchanging the secretion signals in pAα20MX4_Amy was not possible and we had to continue.

As Caplan et al. reported, some Alfa mating factor is still being secreted and thus we decided to proceed to large scale protein expression even with the inconvenience of an adequate leader. Looking at figure 15, CEN.PK2 produces more protein than W303, when comparing for instance, the protein levels at 60h, a band near 60 kDa is visible in CEN.PK2 but not in the W303 lane. The western-blot (Figure 11) shows another band at 75 kDa, which might be a hyperglycosylated form of α-amylase. Despite the increase of 25 % in apparent molecular weight, there are cases reported with an increase in molecular weight of 30 % in hyperglycosylated proteins by *S. cerevisiae*, corresponding to 120

residues of mannose. (Bulter, Alcalde et al. 2003). This supposed hyperglycosylated form is also visible in W303 timecourse.

3.4. Large scale protein expression and purification

Despite the poor secretion of amylase from recombinant yeasts with our plasmid, we still proceeded to a large scale protein expression. This large scale assay was to establish a threshold with the lab strains, to later compare performances of the candidate strains discovered from the library. We started the scale-up with the CEN.PK2 strain. After 120 hours of growth, cell culture was centrifuged and concentrated over several steps, as described in materials and methods. Samples were collected of every step to later evaluate the protein content. Despite the scale-up, recombinant protein levels are barely detected on SDS-PAGE (figure 13). A band near 60 kDa, indicated by the black arrow, could correspond to α-amylase. The western blot shows that α-amylase is mostly accumulated inside the cell, and barely detectable in the supernatant of culture medium. After tangential filtration concentration, some protein signal is detected, but then is lost after concentration by precipitation with 90 ammonium sulphate. The LDP stands for "Low Density Precipitate", and appeared on the surface, when high concentrations of salt was being added to precipitate protein. After 2 hours in high salt concentration, medium was filtered with glass wool, to remove the LDP and procedures carried on. The LDP was resuspended and stored at -20ºC until further use. It would later be discovered that most amylase is being retained in this LDP decreasing protein levels in subsequent purification steps, and no visible protein in lanes PrBF and PrAF.

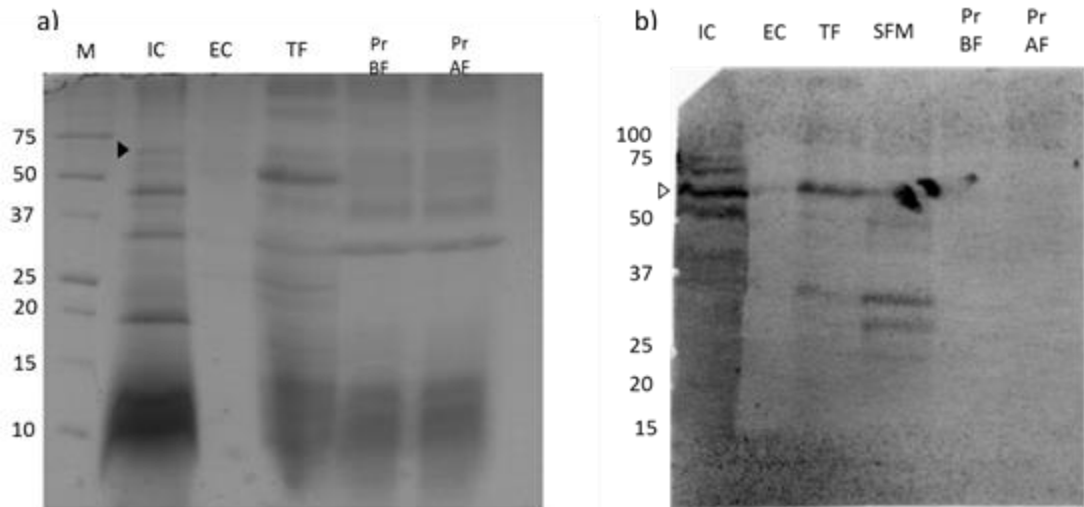

Figure 13 - **α-amylase purification analysis through SDS-PAGE and Western blot**. The samples were denatured with 6x loading buffer at 95ºC for 10 min and 50 µl were loaded in the wells (except for IC in SDS-PAGE where only 5 µl were loaded). SDS-PAGE and western were made in 12.5% acrylamide gels. M – molecular weight marker; IC – Intracellular fraction; EC – extracellular fraction; TC – after tangential filtration; LDP – Low Density Precipitate, appeared during precipitation with high salt concentration; PrBF – after Precipitation with 90% ammonium sulphate, Before Filtration; PrAF – after Precipitation with 90% ammonium sulphate, After Filtration. a) Analysis by SDS-PAGE. b) Western blot analysis with anti His tag antibody

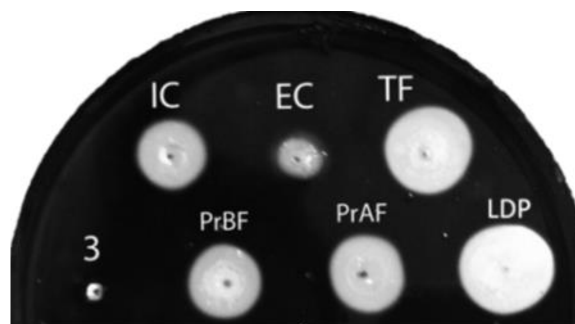

Figure 14 - **Amylase activity assay**. Agar plates with all the samples from the purification steps Solid medium is composed of 50mM Hepes Buffer, 5mM $CaCl_2$ pH 6.9, 2% starch and agar (20g/L). IC – Intracellular fraction; EC – extracellular fraction; TC – after tangential filtration; LDP – Low Density Precipitate, appeared during precipitation with high salt concentration; PrBF – after Precipitation with 90% ammonium sulphate, Before Filtration; PrAF – after Precipitation with 90% ammonium sulphate, After Filtration; 3- 20mM phosphate pH 7.5 buffer with 100mM NaCl.

From figure 14 we can easily see the difference between intracellular fraction (IC) and supernatant (EC), but the increase in activity after concentration with tangential filtration (TF) indicated that some amount of protein is still being secreted, and purification carried on. As discussed before, a low density precipitate (LDP) formed in the high salt concentrated solution, and contained higher activity levels

compared with the other precipitate (PrAF). Avoiding LDP formation could simply be performed by filtration, through 0.2 μM membrane, of TF sample before high amounts of salt is added. Looking ate the results, and because protein purification present in the supernatant was our main objective, LDP was prepared to be injected in size exclusion column Superdex 200 26/60 prep grade.

LDP sample was sequentially filtered through 0.8μ, 0.45μ, and 0.2μ membranes, 9 mL were applied to size exclusion column, previously equilibrated with 20mM phosphate pH 7.5 buffer with 100mM NaCl and 10 ml fractions were collected and analysed by SDS-PAGE and western blot. The chromatografic profile is shown in Figure 15.

The medium composition has a charecteristic brown pigment, and as purification steps advance, the intensity of the brown solution incresaes. This pigment then is separated in the superdex 200 26/60 prep grade to separate amylase from the pigment (starts to exit column after 230 ml in chromatogram). This pigment saturates the UV detector, thus only initial chromatogram profile is shown here. The fractions colected are represented above the peaks and were analysed in agar plates for activity, and in SDS-Page/Western blot.

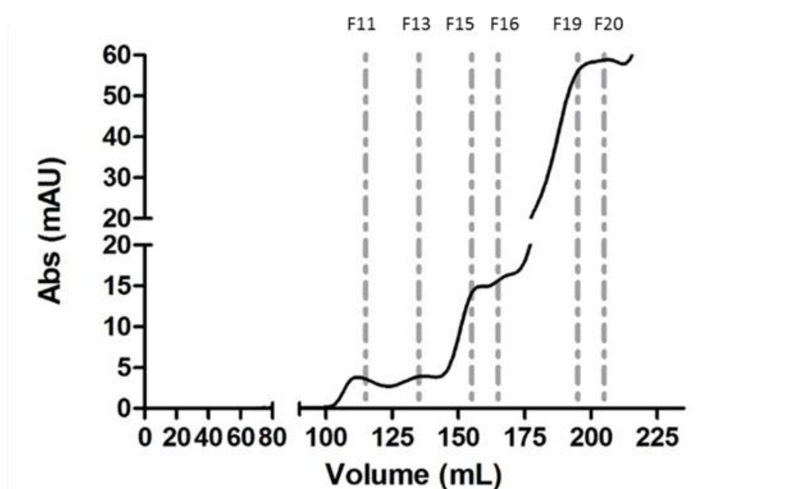

Figure 15 - Purification of LPD α-Amylase in size exclusion chromatography. Superdex 200 26/60 prep grade column equilibrated in 20mM phosphate pH 7.5 buffer with 100mM NaCl and 10 mL samples were collected. The elution occurred at a flow rate of 2 mL/min (at room temp) and A280nm was used to monitor protein elution. Outlined by dotted lines in the chromatogram are elution volume of collected fractions analysed by SDS-PAGE and Western blot. Note that Sample volume comprises the 5ml before and after dotted line.

From previous handling with the column in the lab, amylase should be eluted around 150 mL, but the correspondent peaks, have low amount of protein according to Absorbance units. The activity of LDP is not lost through the filtering process (Figure 16), but after passing through the size exclusion column, is completely lost or residual. Fractions 19 and 20 have some residual activity but is considered insignificant compared to levels seen before separation by chromatography. Western blot also shows signal for histidine tag fused in α-amylase C-terminal, in filtered samples, but no signal is seen in fractions collected from Superdex column. This western procedure was performed using.

This low density precipitate of unknown source and amylase embodied in it is active in solid medium assays, but disappears when injected in the column. Amylase could be forming protein aggregates, but this would be seen with a bigger peak in the chromatogram, as protein aggregates normally are eluted after 100 ml (corresponds to void volume).

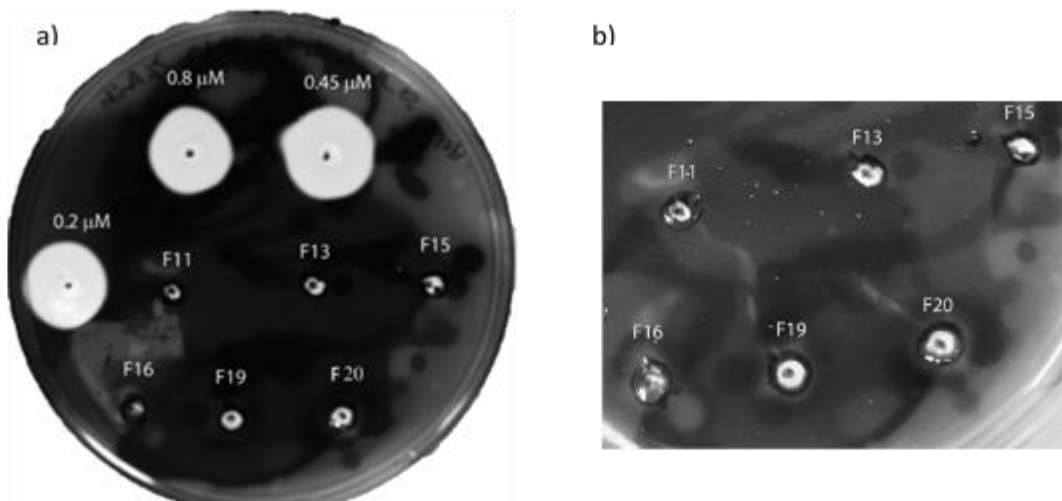

Figure 16 - **LDP Amylase activity assay**. Solid medium composition 50mM Hepes Buffer, 5mM CaCl2 pH 6.9, 2% starch. 0.8µM, 0.45µM and 0.2µM – represent the SFM sample filtered through membranes with respective pores sizes. Right image – zoom-in to visualize activity in samples separated in size exclusion chromatography. 20 µl were loaded in each well.

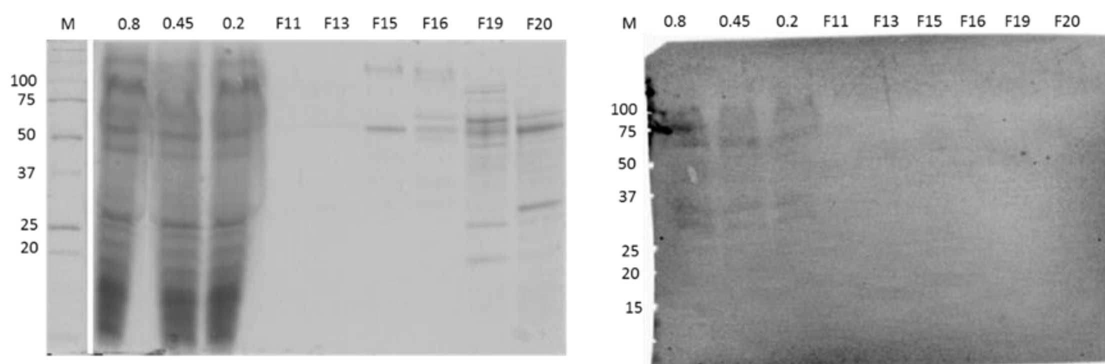

Figure 17 - **LDP α-amylase purification analysis through SDS-PAGE and Western blot**. The samples were denatured with 6x loading buffer at 95ºC for 10 min and 50 µl were loaded in the wells. SDS-PAGE and western were made in 12.5% acrylamide gels. M – molecular weight marker; 0.8 – LDP filtered with 0.8 µM membrane; 0.45 – LDP filtered with 0.45 µM membrane; 0.2 – LDP filtered with 0.2 µM membrane. Fn – fractions collected from Superdex 200 26/60 prep grade. 50µl were loaded in each well a) Analysis by SDS-PAGE. b) Western blot analysis with anti His tag antibody

Since amylase purification from LDP failed, we started procedures to purify amylase remaining in the other precipitate, correspondent to PrAF sample in figure 14. The PrAF samples was already filtered and 9ml were applied to a size exclusion chromatography Superdex 200 prep grade, previously equilibrated in 20mM phosphate pH 7.5 buffer with 100mM NaCl, and 10 mL fractions were colected and analyses by SDS-PAGE and western blot. Fractions colected are shown above the peaks and were analysed in agar plates for activity and in SDS-Page/Western

Peaks of this sample chromatogram (Figure 18 a) look more homogeneous, and have bigger absorbance units, but the activity remained residual (Figure 18 b), and not detectable by SDS-PAGE or western blot (Figure 18 c, d). The samples with residual activity in this chromatography were eluted in a different volume from first chromatogram (150 instead of 200ml). A different Superdex 200 26/60 prep grade was used this time and could be the reason for this difference. LDP sample had a higher viscosity and could retard amylase elution. Both way, activity remained reduced and this let could be due to a series of reasons.

The first reason could rely on amylase nature, and the matrix from the size exclusion column. Amylase (EC 3.2.1.1) hydrolyses α-1, 4-glucosidic bonds in polysaccharides, (such as starch), and could interact with column resin. Superdex, a registered trademark of GE Healthcare resin, is made of cross-linked dextran with highly cross-linked agarose. Dextran is a polymer of glucose units consisting of α-1, 6 glycosidic linkages between glucose molecules and α-1, 3 linkages braches and agarose is a linear polymer of D-galactose and 3, 6-anhydro-1-galactose. Amylase might interact with the column matrix and be eluted much later than expected.

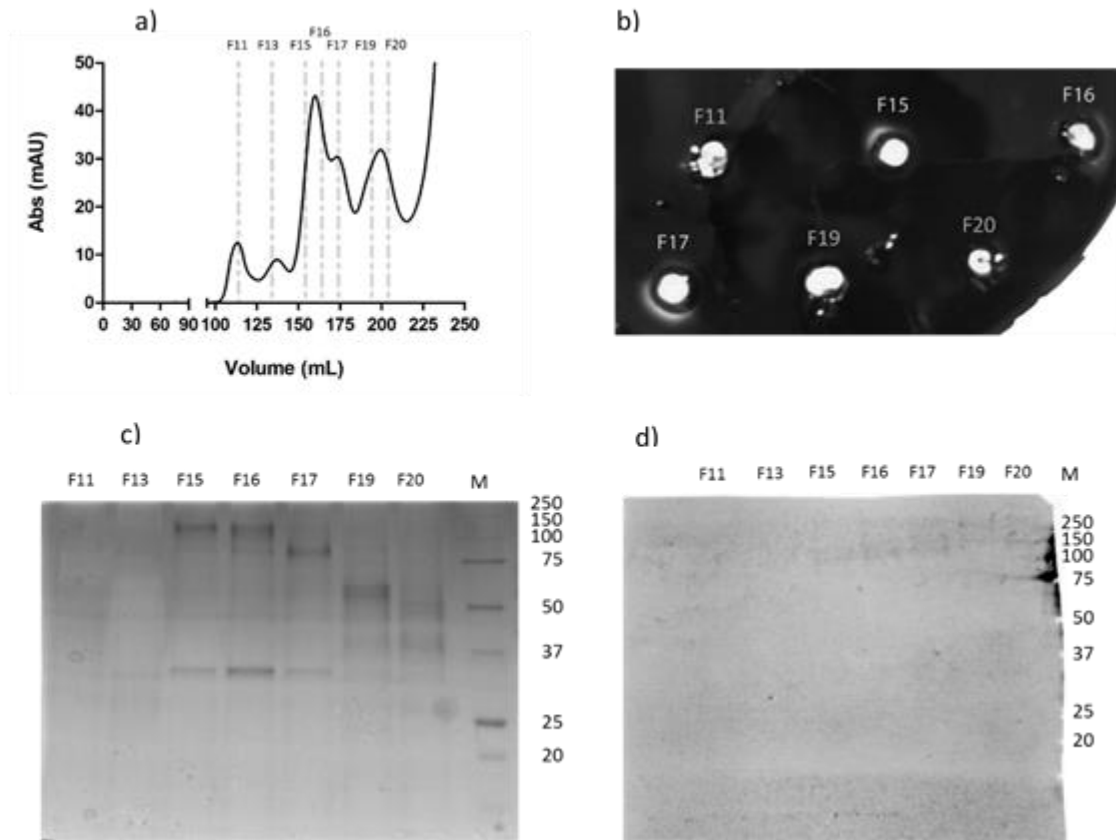

Figure 18 - **Purification and analysis of precipitated protein in High salt concentrations (PrAF sample).** a) Superdex 200 26/60 prep grade column equilibrated in 20mM phosphate pH 7.5 buffer with 100mM NaCl and 10 mL samples were collected. The elution occurred at a flow rate of 2 mL/min (at room temp) and A280nm was used to monitor protein elution. Outlined by dotted lines in the chromatogram are elution volume of collected fractions analysed by: b) activity assay in solid medium (50mM Hepes Buffer, 5mM CaCl2 pH 6.9, 2% starch) and 20ul were loaded in each well; c) SDS-PAGE and d) Western blot. 50 ul were loaded in each well

Samples were only collected until 230 ml and did not have any activity on solid medium (Data not shown). Literature does not provide an equal example to compare, there have been reports of glucosidades with same α-1, 4 linkage specificity, purified with Superdex matrix without complications, although no size exclusion chromatography was used for their purpose (Van Hove, Yang et al. 1997).

Another, more appealing causes for protein loss in the column would be aggregation and retention in the columns pre-filters. The buffer to resuspend, either the pellet of desalted protein (last chromatogram) or the LDP, is the same as the buffer used to equilibrate the column. So the protein does not suffer an abrupt environment change, caused by buffer switch. Maybe Amylase is unstable in this buffer and aggregates when resuspended, or aggregation is due to the previous treatment with

high salt concentration. As said before, amylase depends on Ca^{2+} to be active, and this particular detail could be influencing our experiments.

Stein et al, say that modifications of the secondary or tertiary structure in amylases might result in easy release of calcium ions, a conditions too such modifications could be, variation in pH, use of organic solvent or high salt concentration (Herberhold, Royer et al. 2004) Stein et, al. using EDTA as chelating agent, proved that denaturation induced by calcium release from active site of amylases was reversible simply by adding more calcium to the medium (Stein and Fischer 1958). High salt concentrations were used in our protein purification protocol, which could drive amylase to conformation and aggregate in chromatography columns. But amylase samples with high salt content also demonstrates activity in the agar plates. Agar plates contain calcium in their composition, and when amylase diffuses into the solid medium, conformation is restored and amylase active again.

In literature, reported cases of *Mus musculos* salivary α-Amylase expression and purification from yeasts included calcium in their buffers for chromatography (Thomsen 1983, Kato, Ishibashi et al. 2001). Thus adding calcium to resuspension buffer could recover all the protein loss, allowing for further activity assays and to set a threshold for comparison with a potential candidate from the Yeast wine library.

3.5. 96 microtiter well transformation protocol - Optimization

The transformation protocol needed to be validated before proceeding to high throughput, so a reduced number of yeast strains was used in this initial stage. In the first assay, 10^6 – 10^7 cells per well were used, 600 ng of plasmid DNA, and incubation at 42ºC during one hour were tested. The cell concentration was maintained through all procedures, plasmid DNA concentration and heat shock times were varied until transformation protocol was successful.

The first transformation in 96 microtiter multiwell, using lab strains W303, CEN.PK, and two random *S. cerevisiae* strains from the library resulted only in lab strains transformation (Figure 19 c). Controls were performed as normal transformation procedures, but with no plasmid DNA, all the other reagents were added.

In subsequent trials, and with more wine yeast strains the results were always the same. Lab strains were transfromed but not the wine strains. Wine yeasts are different from lab yeasts. In their natural environment they are exposed to severe conditions, particularly in fermenter tanks. So their exterior wall must somehow be more rigid, and thus less permissive for plasmid incorporation

The original protocol suggests a set of conditions to be tested with different concentrations or incubation times. Due to time constrains the plasmid DNA concentration was raised to 1.7 µg per well, incubation time at 42 ºC was performed during 3h, and only strains from the library were used.

From 16 random S. strains selected from the library, almost all of the yeasts were transformed and formed halos around the colonies (Figure 20). The transformation rate with this conditions was above 90 %, so the protocol was optimized and high throughput procedures were carried on

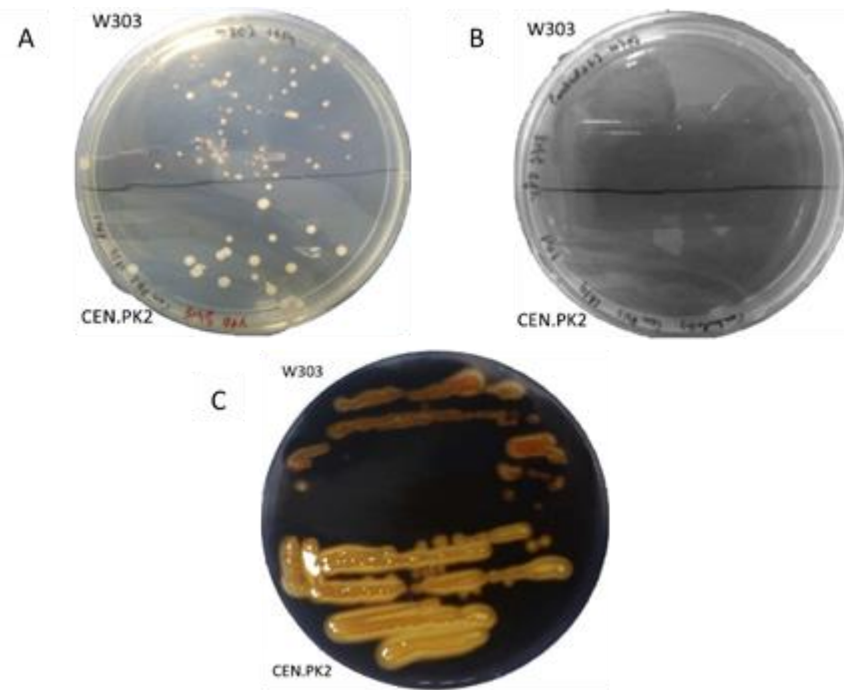

Figure 19 - **Positive clones from the first round of 96 microtiter multiwell transformation protocol.** A – Positive clones plated on selective medium containing 200 µg/ml G418. B – Controls of plated yeasts transformed without plasmid DNA. C – Amylase activity confirmation of positive clones plated on selective medium containing 200 µg/ml G418 and 2% starch and stained with lugol solution.

Figure 20 - **Positive clones transformed with the final protocol modifications**. 16 random strains from the library were transformed and plated on YPD medium supplemented with 200 µg/ml G418 and 2% starch. Plates were stained with lugol solution

3.6. High throughput Screening – Controls

Approximately 400 *S.cerevisiae* strains from the library were transformed as described in materials and methods. Two controls were designed for this purpose. One with antibiotic (agar plates supplemented with 200 µg/mL G418), to assess whether yeast strains were naturally resistant to G418 antibiotic. And another control on YPD agar plates supplemented with 2% starch (without antibiotic), to discriminate the strains capable of secreting their own amylases or other enzymes capable of degrading starch.

In total, 6 deepwell plates were used for the screening. Each well contains 80 strains of *S.cerevisiae* (16 wells serve as control). Figure 21 is only a representative example of the controls. In total, none of the 400 native strains were capable of forming halos, and only 1-3 % of the strains had natural resistance to 200 µg/mL G418. Several antibiotic concentrations were tested to eliminate this natural resistance, and 200 µg/mL was found to be the best concentration to minimize natural capability of yeasts strains to grow in selective conditions. Concentrations at 60 µg/mL were insufficient, as resistant yeasts strains increased to 11 %, and 300 µg G418 was no better than the 200 concentration.

In the starch control, none of all strains formed halos, so all halo formation in transformed yeast are assured to result from our plasmid DNA.

This screening was performed in a high throughput Sciclone ALH 3000 Workstation which transfered cells from deep well plates, to single wells plates with the specifc solid medium, and plates were grown O.N..

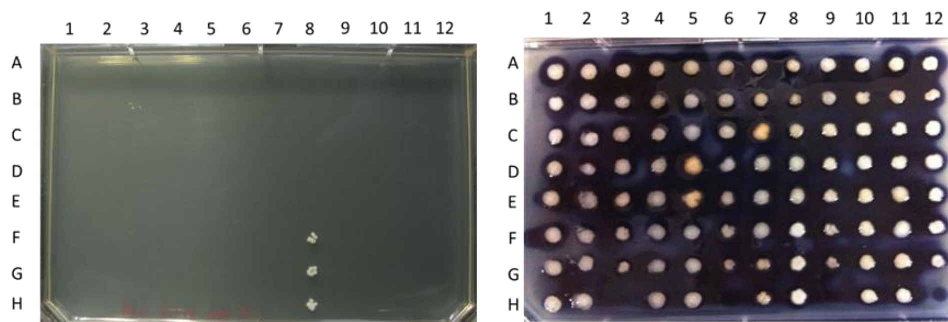

Figure 21 - **High throughput Controls of S. cerevisiae library**. Strains were picked from the deep well plate with the help of the Sciclone ALH 3000 Workstation-Caliper robot and the 96 pin tool, and plated on solid YPD. Left – single well plate with solid YPD + 200 µg/mL G418. Right – single well plate with solid YPD supplemented with 2% starch stained with lugol solution. Wells H(3,6,9,12) don't have cells and serve as controls. E (3,6,9,12) are comercial wine yeast Lalvin EC118; F (3,6,9,12) have comercial wine yeast QA23; G (3,6,9,12) have the lab strain s228.;

3.7. High throughput transformation

As explained above, after transformation protocol optimization, high throughput transformation was carried on. A total of 392 yeast strains, organized in 6 deep well plates were successfully transformed and halos around colonies formed. These strains would be isolated and further characterized to validate them as powerful recombinant protein expression system to use in the lab. Other laboratories with their own yeast libraries, upon screening, obtained 3 or 4 suitable candidates from the library suitable for heterologous protein expression, and this was our threshold as well. Figure 22 shows the transformation of one 96 deep well plate results after strains are grown in solid selective medium with starch after staining

To discriminate between yeasts with high, medium or low levels of protein secretion during halo formation, we measured the ratio of halo/colony size of all individuals. The transformed strains in this project represent only a small fraction of the whole library, but a significant number of strains was transformed and yielded good results. Overall, 82 % transformation efficiencies prevailed using this protocol, which can in the future be employed to screen the whole library.

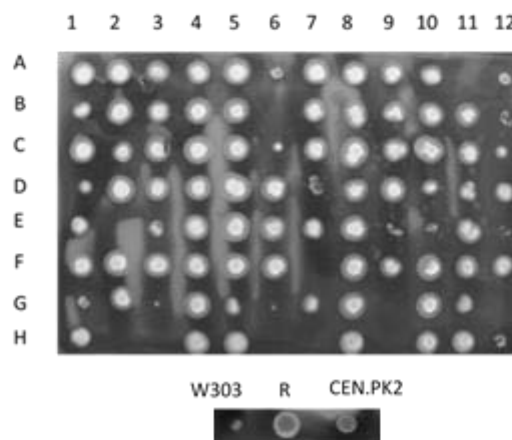

Figure 22 - **High throughput activity assay**. Example of a single well plate with transformed colonies plated on YPD solid medium supplemented with 200 µg/ml G418 and 2% starch. The image fragment contains lab strains transformed and grown in the same conditions. R stands for random yeast strain belonging to the library

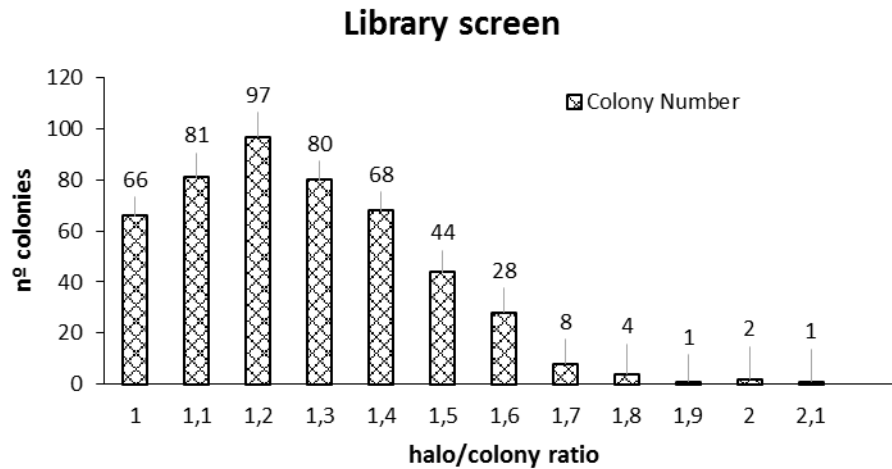

Figure 23 **Activity…**. X axis represents the interval of values, ploted against the number of colonies with rations between this gab. There are 82 strains with ratios between 1 and 1.0999, 83 strains with ratios between 1.1 and 1.1999 etc. Bars represent number of colonies, and black line represents the same data in percentage.

Figure 23 holds the graphic in with ratios were plotted against number of colonies. Basically it represents the number of strains with in a given interval of ratios. So colonies with ratios between 1 and 1.099 were 82 exemplars and so on. Most of the exemplars screened have small or medium values for ratios, and the number of candidates decreases for bigger ratios. From these analysis, we can see that there are some exemplars with ratios near 2. That represents a halo with double size of the colony. These strains are strong candidates to proceed to further characterization and establishment of an expression system. Overall, form about 392 strains, 3 good strains chowed higher protein capacity evidence.

If more candidates are screened, some more and maybe better strains will arise. Biodiversity of such libraries have proven invaluable and can outweigh genetic engineering performed in strains, and increase the understanding of the mechanisms behind the new features discovered. These strains have evolved over millions of years and can have some enhanced genetic traits for use in the laboratory, or even in industry. The reports about other yeasts (*P. pastoris*) outrunning *S. cerevisiae* in recombinant protein production is, in most times compared to lab strains and not, for instance, to industry strains used for protein production. These industry strains are highly protected and only rare reports are found in literature. They can probably perform 10 x (or more) the production of the best lab strains known of *saccharomyces cerevisiae*.

Some improvements can still be made to fine tune this screening. The possible problem with the secretion leader can be hiding the true potential of some strains in producing and secreting

recombinant protein. Another issue not mentioned until now in our screen is the percentage of agar used in solid medium plates. Agar plates are constituted of polymers with form an intrinsic net of pores with variable sizes which can influence the diffusion of macromolecules such as proteins. The bigger the percentage of agar used, the more restrictive becomes the intra space of the gel, and adding to this, 2 % starch are added to mixture in our experiments and makes the 3D structure of the gel even more rigid. These factors can influence Alfa amylase diffusion through the gel, and even if starch is degraded in the process, less concentration of both could lead to bigger halo formation.

Not much is available about the influence of protein size and diffusion of proteins in agar. (Ackers and Steere 1962) tested the restriction diffusion coefficient of several proteins through agar membranes ranging from 1-8 % concentration. They used commercial Haemoglobin (67 kDa), γ-globulin (160 kDa), serum albumin (69kDa) and Ribonucleases (13 kDa) and saw clear differences in restriction diffusion measurements according to the size of the protein, and more important, they also saw how agar gel % influences. In their case, agar membranes were used which is different from agar plates, but still it indicates that the pores sizes on agar significantly affects diffusion of macromolecules such as proteins. A simple test in laboratory can be made to assess whether α-amylase is influenced in our experiments, and if it is significant or not. This could lead to better and more accurate measurements of the halos helping to differentiate a good strain form an excellent one in our context.

CHAPTER 4

Conclusions and Future Remarks

4.1. Conclusions and future remarks.

The major goal of the developed work was find stains of S. cerevisiae with enhanced proteins production capacity. First, a high throughput protocol for yeast transformation, to screen a S. cerevisiae library for potential protein expression systems needed to be assembled.

The first task consisted in the assembling of an expression vector essential for the later screening procedure. In order to overcome Ligation-digestion cloning extensive time a labour drawbacks, CPEC Gibson and In-fusion were used to assemble the vector in a short period of time, with less labour required and to set-up the seamless ligation techniques in the laboratory. But no positive results came from using this techniques and the problem in vector assembling was resolved with digestion-ligation cloning.

With not much information in literature to compare, the prepro sequence of the mating factor leader from *K. Lactis* tested in a *S. cerevisiae* host seems to be responsible for intracellular accumulation of recombinant α-amylase. In literature when glycosylation or sequence of the prepro peptide of αMF are altered, intracellular accumulation of protein occurs, which seems to be the case, due to low levels of protein yielded in our large-scale experiments. Changing *K. lactis* for S. *cerevisiae* mating factor secretion leader can yield higher protein levels, and allowing to establish a threshold for protein levels using lab strains, and later to evaluate the protein production limits of the relevant candidates form the yeast library.

Another issue which happened during our experiments was the lack of activity in protein samples after purified in size exclusion chromatography. The high salt conditions could be responsible for removing Ca^{2+} from the α-Amylase globular structure leading to inactivation and aggregation, which could result in aggregates and are trapped in the column pre-filters. Protein samples in high salt concentrations that were not injected in the column do present activity, but this could be related to the another feature of α-amylase, which is their ability to regain back its activity in calcium rich environments, which in our case correspond to the agar plates with buffer and calcium. Another hypothesis is the fact that as a glycosidase, it could somehow interact with the column matrix, composed of highly cross linked dextran, which itself is a polymer of glucoses and be eluted in later, corresponding to fractions not even collected. Amylase has preference for α-1.4 glycosidic bonds, and dextran is made of α-1.6 for linear and α-1.3 for branched bonds nevertheless amylase might still interact with this sugars. Other groups have purified glucoamylases in Superdex resins with no problems..

Finally, the protocol establishment went successful and high throughput transformations performed with efficiency. Some candidates showed up with promising performances from the yeast strain library and can be further characterized. Yet some conditions can still be improved to fine tune this process. Along with the secretion leader change, the agar concentration in solid medium can also be varied to improve the ratios halo/colony to help differentiate good strains from outstanding ones.

REFERENCES

Ackers, G. and R. Steere (1962). "Restricted diffusion of macromolecules through agar-gel membranes." Biochimica et biophysica acta **59**(1): 137-149.

Agaphonov, M., N. Romanova, E. S. Choi and M. Ter-Avanesyan (2010). "A novel kanamycin/G418 resistance marker for direct selection of transformants in Escherichia coli and different yeast species." Yeast **27**(4): 189-195.

Agathos, S. N. (1991). "Production scale insect cell culture." Biotechnol Adv **9**(1): 51-68.

Almeida, J. R., T. Modig, A. Petersson, B. Hähn-Hägerdal, G. Lidén and M. F. Gorwa-Grauslund (2007). "Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae." Journal of chemical technology and biotechnology **82**(4): 340-349.

Bachmann, B. J. (1972). "Pedigrees of some mutant strains of Escherichia coli K-12." Bacteriological reviews **36**(4): 525.

Baker, S., S. Johnston, J. Hopper and J. Jaehning (1987). "Transcription of multiple copies of the yeast GAL7 gene is limited by specific factors in addition to GAL4." Molecular and General Genetics MGG **208**(1-2): 127-134.

Becker, D. M. and L. Guarente (1991). "High-efficiency transformation of yeast by electroporation." Methods Enzymol **194**: 182-187.

Benítez, T., J. M. Gasent-Ramírez, F. Castrejón and A. C. Codón (1996). "Development of new strains for the food industry." Biotechnology progress **12**(2): 149-163.

Bisbee, C. (1993). "Current perspectives on manufacturing and scaleup of biopharmaceuticals." Gen Eng News **13**: 8-10.

Bitter, G. A. and K. M. Egan (1984). "Expression of heterologous genes in< i> Saccharomyces cerevisiae</i> from vectors utilizing the glyceraldehyde-3-phosphate dehydrogenase gene promoter." Gene **32**(3): 263-274.

Bongers, R. S., J. W. Veening, M. Van Wieringen, O. P. Kuipers and M. Kleerebezem (2005). "Development and characterization of a subtilin-regulated expression system in Bacillus subtilis: strict control of gene expression by addition of subtilin." Appl Environ Microbiol **71**(12): 8818-8824.

Breiteneder, H., M. Krebitz, U. Wiedermann, B. Wagner, D. Essl, H. Steinkellner, T. Turpen, C. Ebner, D. Buck and B. Niggemann (2012). "Rapid production of recombinant allergens in Nicotiana benthamiana

and their impact on diagnosis and therapy." International archives of allergy and immunology **124**(1-3): 48-50.

Buckholz, R. G. and M. A. Gleeson (1991). "Yeast systems for the commercial production of heterologous proteins." Biotechnology (N Y) **9**(11): 1067-1072.

Bulter, T., M. Alcalde, V. Sieber, P. Meinhold, C. Schlachtbauer and F. H. Arnold (2003). "Functional expression of a fungal laccase in Saccharomyces cerevisiae by directed evolution." Appl Environ Microbiol **69**(2): 987-995.

Burgers, P. M. and K. J. Percival (1987). "Transformation of yeast spheroplasts without cell fusion." Analytical biochemistry **163**(2): 391-397.

Caplan, S., R. Green, J. Rocco and J. Kurjan (1991). "Glycosylation and structure of the yeast MF alpha 1 alpha-factor precursor is important for efficient transport through the secretory pathway." Journal of bacteriology **173**(2): 627-635.

Casey, G. P. and W. M. Ingledew (1986). "Ethanol tolerance in yeasts." Crit Rev Microbiol **13**(3): 219-280.

Casper, S. J. and C. A. Holt (1996). "Expression of the green fluorescent protein-encoding gene from a tobacco mosaic virus-based vector." Gene **173**(1 Spec No): 69-73.

Cereghino, J. L. and J. M. Cregg (2000). "Heterologous protein expression in the methylotrophic yeast Pichia pastoris." FEMS Microbiol Rev **24**(1): 45-66.

Chiba, Y., M. Suzuki, S. Yoshida, A. Yoshida, H. Ikenaga, M. Takeuchi, Y. Jigami and E. Ichishima (1998). "Production of Human Compatible High Mannose-type (Man5GlcNAc2) Sugar Chains inSaccharomyces cerevisiae." Journal of Biological Chemistry **273**(41): 26298-26304.

Clark, A. J. (1998). "The mammary gland as a bioreactor: expression, processing, and production of recombinant proteins." Journal of mammary gland biology and neoplasia **3**(3): 337-350.

Coca, M., C. Bortolotti, M. Rufat, G. Penas, R. Eritja, D. Tharreau, A. M. del Pozo, J. Messeguer and B. San Segundo (2004). "Transgenic rice plants expressing the antifungal AFP protein from Aspergillus giganteus show enhanced resistance to the rice blast fungus Magnaporthe grisea." Plant molecular biology **54**(2): 245-259.

Costanzo, M. C. and T. D. Fox (1988). "Transformation of yeast by agitation with glass beads." Genetics **120**(3): 667-670.

Courtois, S., C. M. Cappellano, M. Ball, F. X. Francou, P. Normand, G. Helynck, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osburne, P. R. August, R. Nalin, M. Guerineau, P. Jeannin, P. Simonet and J. L. Pernodet (2003). "Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products." Appl Environ Microbiol **69**(1): 49-55.

Cregg, J. (1999). Expression in the methylotrophic yeast Pichia pastoris, Academic Press, San Diego, CA**:** 157-191.

Cregg, J. M., K. Barringer, A. Hessler and K. Madden (1985). "Pichia pastoris as a host system for transformations." Molecular and Cellular Biology **5**(12): 3376-3385.

Cregg, J. M., K. R. Madden, K. J. Barringer, G. P. Thill and C. A. Stillman (1989). "Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris." Mol Cell Biol **9**(3): 1316-1323.

Cregg, J. M., T. S. Vedvick and W. C. Raschke (1993). "Recent advances in the expression of foreign genes in Pichia pastoris." Biotechnology (N Y) **11**(8): 905-910.

De Bernardez Clark, E. (1998). "Refolding of recombinant proteins." Current Opinion in Biotechnology **9**(2): 157-163.

De Ruyter, P., O. P. Kuipers, M. M. Beerthuyzen, I. van Alen-Boerrigter and W. De Vos (1996). "Functional analysis of promoters in the nisin gene cluster of Lactococcus lactis." Journal of bacteriology **178**(12): 3434-3439.

de Winde, H. (2003). Functional genetics of industrial yeasts; of ancient skills and modern applications, Springer.

Demain AL, V. P. (2009). "Production of recombinant proteins by microbes and higher organisms." Biotechnol Adv **27**: 297-306.

Dequin, S. (2001). "The potential of genetic engineering for improving brewing, wine-making and baking yeasts." Appl Microbiol Biotechnol **56**(5-6): 577-588.

Dower, W. J., J. F. Miller and C. W. Ragsdale (1988). "High efficiency transformation of E. coli by high voltage electroporation." Nucleic Acids Res **16**(13): 6127-6145.

Echelard, Y. (1996). "Recombinant protein production in transgenic animals." Curr Opin Biotechnol **7**(5): 536-540.

Eddy, A. and D. Williamson (1957). "A method of isolating protoplasts from yeast."

Edwards, C. P. and A. Aruffo (1993). "Current applications of COS cell based transient expression systems." Curr Opin Biotechnol **4**(5): 558-563.

Ellis, S. B., P. F. Brust, P. J. Koutz, A. F. Waters, M. M. Harpold and T. R. Gingeras (1985). "Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast Pichia pastoris." Mol Cell Biol **5**(5): 1111-1121.

Endo, Y. and T. Sawasaki (2003). "High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system." Biotechnol Adv **21**(8): 695-713.

Fan, J. and T. Watanabe (2003). "Transgenic rabbits as therapeutic protein bioreactors and human disease models." Pharmacology & therapeutics **99**(3): 261-282.

Fisher, B., I. Sumner and P. Goodenough (1993). "Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in Escherichia coli as inclusion bodies." Biotechnol. Bioeng **41**: 3-13.

Gellissen, G. (2000). "Heterologous protein production in methylotrophic yeasts." Appl Microbiol Biotechnol **54**(6): 741-750.

Gellissen, G. (2006). Production of recombinant proteins, Wiley. com.

Gibson, D. G., L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, 3rd and H. O. Smith (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." Nat Methods **6**(5): 343-345.

Gietz, R. D. and R. H. Schiestl (2007). "Microtiter plate transformation using the LiAc/SS carrier DNA/PEG method." Nat Protoc **2**(1): 5-8.

Gietz, R. D., R. H. Schiestl, A. R. Willems and R. A. Woods (1995). "Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure." Yeast **11**(4): 355-360.

Goffeau A, B. B., Bussey H, Davis RW, Dujon B, Feldmann H, (1996). "Life with 6000 genes. Science." Science **274**: 546-567.

Gouka, R. J., P. J. Punt and C. A. van den Hondel (1997). "Efficient production of secreted proteins by Aspergillus: progress, limitations and prospects." Appl Microbiol Biotechnol **47**(1): 1-11.

Gritz, L. and J. Davies (1983). "Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in Escherichia coli and Saccharomyces cerevisiae." Gene **25**(2-3): 179-188.

Guijo, S., J. Mauricio, J. Salmon and J. Ortega (1997). "Determination of the Relative Ploidy in Different Saccharomyces cerevisiae Strains used for Fermentation and 'Flor'Film Ageing of Dry Sherry-type Wines." Yeast **13**(2): 101-117.

Hadfield, C., A. M. Cashmore and P. A. Meacock (1986). "An efficient chloramphenicol-resistance marker for Saccharomyces cerevisiae and Escherichia coli." Gene **45**(2): 149-158.

Hammond, J. R. (1995). "Genetically-modified brewing yeasts for the 21st century. Progress to date." Yeast **11**(16): 1613-1627.

Hanai, N., K. Nagata, A. Kawajiri, T. Shiromizu, N. Saitoh, Y. Hasegawa, S. Murakami and M. Inagaki (2004). "Biochemical and cell biological characterization of a mammalian septin, Sept11." FEBS Lett **568**(1-3): 83-88.

Hayhurst, A. and W. J. Harris (1999). "< i> Escherichia coli</i> Skp Chaperone Coexpression Improves Solubility and Phage Display of Single-Chain Antibody Fragments." Protein expression and purification **15**(3): 336-343.

He, X.-S., R. Brückner and R. Doi (1991). "The protease genes of< i> Bacillus subtilis</i>." Research in microbiology **142**(7): 797-803.

Herberhold, H., C. A. Royer and R. Winter (2004). "Effects of chaotropic and kosmotropic cosolvents on the pressure-induced unfolding and denaturation of proteins: an FT-IR study on staphylococcal nuclease." Biochemistry **43**(12): 3336-3345.

Higgins, D. R. and J. M. Cregg (1998). Pichia protocols, Springer.

Hinnen, A., F. Buxton, B. Chaudhuri, J. Heim, T. Hottiger, B. Meyhack and G. Pohlig (1995). "Gene expression in recombinant yeast." Bioprocess Technol **22**: 121-193.

Hinnen, A., J. B. Hicks and G. R. Fink (1978). "Transformation of yeast." Proc Natl Acad Sci U S A **75**(4): 1929-1933.

Hinnen, A., B. Meyhack and J. Heim (1989). "Heterologous gene expression in yeast." Biotechnology **13**: 193-213.

Hitzeman, R. A., F. E. Hagie, H. L. Levine, D. V. Goeddel, G. Ammerer and B. D. Hall (1981). "Expression of a human gene for interferon in yeast." Nature **293**(5835): 717-722.

Hitzeman, R. A., D. W. Leung, L. J. Perry, W. J. Kohr, H. L. Levine and D. V. Goeddel (1983). "Secretion of human interferons by yeast." Science **219**(4585): 620-625.

Holmberg, N., G. Lilius, J. E. Bailey and L. Buelow (1997). "Transgenic tobacco expressing Vitreoscilla hemoglobin exhibits enhanced growth and altered metabolite production." Nature biotechnology **15**(3): 244-247.

Huang CJ, L. A., Batt CA (2010). "Recombinant immunotherapeutics: current state and perspectives regarding the feasibility and market." Appl Microbiol Biotechnol **87**: 401-410.

Ideker, T., T. Galitski and L. Hood (2001). "A new approach to decoding life: systems biology." Annu Rev Genomics Hum Genet **2**(1): 343-372.

Iha, H. and K. Tsurugi (1998). "Shuttle-vector system for Saccharomyces cerevisiae designed to produce C-terminal-Myc-tagged fusion proteins." Biotechniques **25**(6): 936-938.

Ito, H., Y. Fukuda, K. Murata and A. Kimura (1983). "Transformation of intact yeast cells treated with alkali cations." J Bacteriol **153**(1): 163-168.

Iwashita, K. (2002). "Recent studies of protein secretion by filamentous fungi." J Biosci Bioeng **94**(6): 530-535.

Jackson, A. M., J. Boutell, N. Cooley and M. He (2004). "Cell-free protein synthesis for proteomics." Brief Funct Genomic Proteomic **2**(4): 308-319.

Janiyani, K., T. Bordelon, G. L. Waldrop and J. E. Cronan (2001). "Function of Escherichia coli biotin carboxylase requires catalytic activity of both subunits of the homodimer." Journal of Biological Chemistry **276**(32): 29864-29870.

Jermutus, L., L. A. Ryabova and A. Plückthun (1998). "Recent advances in producing and selecting functional proteins by using cell-free translation." Current opinion in biotechnology **9**(5): 534-548.

Jiang, X., Y. Ookubo, I. Fujii, H. Nakano and T. Yamane (2002). "Expression of Fab fragment of catalytic antibody 6D9 in an< i> Escherichia coli</i> in vitro coupled transcription/translation system." FEBS letters **514**(2): 290-294.

Johnston, S. A. and M. J. DeVit (1996). Biolistic transformation of yeasts. Yeast Protocols, Springer**:** 147-153.

Kato, S., M. Ishibashi, D. Tatsuda, H. Tokunaga and M. Tokunaga (2001). "Efficient expression, purification and characterization of mouse salivary alpha-amylase secreted from methylotrophic yeast, Pichia pastoris." Yeast **18**(7): 643-655.

Katzen, F., G. Chang and W. Kudlicki (2005). "The past, present and future of cell-free protein synthesis." Trends Biotechnol **23**(3): 150-156.

Kigawa, T., T. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, T. Shibata and S. Yokoyama (1999). "Cell-free production and stable-isotope labeling of milligram quantities of proteins." FEBS Lett **442**(1): 15-19.

Kigawa, T., E. Yamaguchi-Nunokawa, K. Kodama, T. Matsuda, T. Yabuki, N. Matsuda, R. Ishitani, O. Nureki and S. Yokoyama (2002). "Selenomethionine incorporation into a protein by cell-free synthesis." Journal of structural and functional genomics **2**(1): 29-35.

King, L. A. and R. D. Possee (1992). The baculovirus expression system: a laboratory guide, Chapman & Hall London.

Kjeldsen, T. (2000). "Yeast secretory expression of insulin precursors." Appl Microbiol Biotechnol **54**(3): 277-286.

Klammt, C., F. Lohr, B. Schafer, W. Haase, V. Dotsch, H. Ruterjans, C. Glaubitz and F. Bernhard (2004). "High level cell-free expression and specific labeling of integral membrane proteins." Eur J Biochem **271**(3): 568-580.

Ko, J. H., M. S. Hahm, H. A. Kang, S. W. Nam and B. H. Chung (2002). "Secretory expression and purification of Aspergillus niger glucose oxidase in Saccharomyces cerevisiae mutant deficient in PMR1 gene." Protein Expr Purif **25**(3): 488-493.

Koichi Ogata, H. N., Masahiro Ohsugi (1969). "A Yeast Capable of Utilizing Methanol." Agricultural and Biological Chemistry **33**(10): 1519-1520.

Kost, T. A., J. P. Condreay and D. L. Jarvis (2005). "Baculovirus as versatile vectors for protein expression in insect and mammalian cells." Nature biotechnology **23**(5): 567-575.

Kuipers, O. P., P. G. de Ruyter, M. Kleerebezem and W. M. de Vos (1997). "Controlled overproduction of proteins by lactic acid bacteria." Trends Biotechnol **15**(4): 135-140.

Kuipers, O. P., P. G. de Ruyter, M. Kleerebezem and W. M. de Vos (1998). "Quorum sensing-controlled gene expression in lactic acid bacteria." Journal of Biotechnology **64**(1): 15-21.

Lacey, L. A. and T. R. Unruh (2005). "Biological control of codling moth (Cydia pomonella, Lepidoptera: Tortricidae) and its role in integrated pest management, with emphasis on entomopathogens." Vedalia **12**(1): 33-60.

LaVallie, E. R. and J. M. McCoy (1995). "Gene fusion expression systems in< i> Escherichia coli</i>." Current opinion in biotechnology **6**(5): 501-506.

Le Loir, Y., S. Nouaille, J. Commissaire, L. Bretigny, A. Gruss and P. Langella (2001). "Signal peptide and propeptide optimization for heterologous protein secretion in Lactococcus lactis." Appl Environ Microbiol **67**(9): 4119-4127.

Li, J., H. Xu, W. E. Bentley and G. Rao (2002). "Impediments to secretion of green fluorescent protein and its fusion from Saccharomyces cerevisiae." Biotechnol Prog **18**(4): 831-838.

Magliery, T. J., C. G. Wilson, W. Pan, D. Mishler, I. Ghosh, A. D. Hamilton and L. Regan (2005). "Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism." J Am Chem Soc **127**(1): 146-157.

Maiorella, B., D. Inlow, A. Shauger and D. Harano (1988). "Large-scale insect cell-culture for recombinant protein production." Nature Biotechnology **6**(12): 1406-1410.

Mao, X., Y. Hu, C. Liang and C. Lu (2002). "MET3 promoter: a tightly regulated promoter and its application in construction of conditional lethal strain." Current microbiology **45**(1): 37-40.

Mascorro-Gallardo, J., A. Covarrubias and R. Gaxiola (1996). "Construction of a< i> CUP1</i> promoter-based vector to modulate gene expression in< i> Saccharomyces cerevisiae</i>." Gene **172**(1): 169-170.

Matsuura, Y., R. D. Possee, H. A. Overton and D. H. Bishop (1987). "Baculovirus expression vectors: the requirements for high level expression of proteins, including glycoproteins." J Gen Virol **68 (Pt 5)**: 1233-1250.

Maurer, K. H. (2004). "Detergent proteases." Curr Opin Biotechnol **15**(4): 330-334.

McAleer WJ, B. E., Maigetter RZ, Wampler DE, Miller WJ, Hilleman MR. (1984). "Human hepatitis B vaccine from recombinant yeast." Nature **307**: 178-180.

McGovern, P. E., U. Hartung, V. R. Badler, D. L. Glusker and L. J. Exner (1997). "The beginnings of winemaking and viniculture in the ancient Near East and Egypt." Expedition **39**(1): 3-21.

McGovern, P. E., J. Zhang, J. Tang, Z. Zhang, G. R. Hall, R. A. Moreau, A. Nunez, E. D. Butrym, M. P. Richards, C. S. Wang, G. Cheng, Z. Zhao and C. Wang (2004). "Fermented beverages of pre- and proto-historic China." Proc Natl Acad Sci U S A **101**(51): 17593-17598.

Mentesana, P. E., M. Dosil and J. B. Konopka (2002). "Functional assays for mammalian G-protein-coupled receptors in yeast." Methods Enzymol **344**: 92-111.

Monfort, A., S. Finger, P. Sanz and J. Prieto (1999). "Evaluation of different promoters for the efficient production of heterologous proteins in baker's yeast." Biotechnology letters **21**(3): 225-229.

Moscardi, F. (1999). "Assessment of the application of baculoviruses for control of Lepidoptera." Annu Rev Entomol **44**(1): 257-289.

Murakami, Y., M. Tsuyama, Y. Kobayashi, H. Kodama and K. Iba (2000). "Trienoic fatty acids and plant tolerance of high temperature." Science **287**(5452): 476-479.

Neidhart, F. (1996). Escherichia coli and Salmonella, Blackwell Science Ltd.

Nevoigt, E. (2008). "Progress in metabolic engineering of Saccharomyces cerevisiae." Microbiol Mol Biol Rev **72**(3): 379-412.

Niedenthal, R., L. Riles, M. Johnston and J. Hegemann (1996). "Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast." Yeast **12**(8): 773-786.

Nirenberg, M. W. and J. H. Matthaei (1961). "The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides." Proceedings of the National Academy of Sciences of the United States of America **47**(10): 1588.

O'Reilly, D. R., L. K. Miller and V. A. Luckow (1994). The baculovirus expression vectors: a laboratory manual, Oxford University Press.

Oppenoorth, F. J. and A. van (1960). "Allelic genes in the housefly producing modified enzymes that cause organophosphate resistance." Science **132**(3422): 298-299.

Partow, S., V. Siewers, S. Bjorn, J. Nielsen and J. Maury (2010). "Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae." Yeast **27**(11): 955-964.

Porro, D., M. Sauer, P. Branduardi and D. Mattanovich (2005). "Recombinant protein production in yeasts." Mol Biotechnol **31**(3): 245-259.

Pretorius, I. S. (2000). "Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking." Yeast **16**(8): 675-729.

Prezant, T. R., E. William jr and N. Fischel-Ghodsian (1996). "Identification of an overexpressed yeast gene which prevents aminoglycoside toxicity." Microbiology **142**(12): 3407-3414.

Price, V. L., W. E. Taylor, W. Clevenger, M. Worthington and E. T. Young (1990). "Expression of heterologous proteins in Saccharomyces cerevisiae using the ADH2 promoter." Methods Enzymol **185**: 308-318.

Punt, P. J., N. van Biezen, A. Conesa, A. Albers, J. Mangnus and C. van den Hondel (2002). "Filamentous fungi as cell factories for heterologous protein production." Trends in biotechnology **20**(5): 200-206.

Quintero, M. J., D. Maya, M. Arevalo-Rodriguez, A. Cebolla and S. Chavez (2007). "An improved system for estradiol-dependent regulation of gene expression in yeast." Microb Cell Fact **6**: 10.

Raivio, T. L. and T. J. Silhavy (2001). "Periplasmic stress and ECF sigma factors." Annual Reviews in Microbiology **55**(1): 591-624.

Romanos, M. A., C. A. Scorer and J. J. Clare (1992). "Foreign gene expression in yeast: a review." Yeast **8**(6): 423-488.

Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman and R. M. Goodman (2000). "Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms." Appl Environ Microbiol **66**(6): 2541-2547.

Rudolph, R. and H. Lilie (1996). "In vitro folding of inclusion body proteins." FASEB J **10**(1): 49-56.

Ruohonen, L., M. K. Aalto and S. Keränen (1995). "Modifications to the< i> ADH1</i> promoter of< i> Saccharomyces cerevisiae</i> for efficient production of heterologous proteins." Journal of biotechnology **39**(3): 193-203.

Ryabova, L. A., D. Desplancq, A. S. Spirin and A. Plückthun (1997). "Functional antibody production using cell-free translation: effects of protein disulfide isomerase and chaperones." Nature biotechnology **15**(1): 79-84.

Sanchez, M., N. Prim, F. Randez-Gil, F. I. Pastor and P. Diaz (2002). "Engineering of baker's yeasts, E. coli and Bacillus hosts for the production of Bacillus subtilis Lipase A." Biotechnol Bioeng **78**(3): 339-345.

Sanford, J. C. (1988). "The biolistic process." Trends in Biotechnology **6**(12): 299-302.

Schauzu, M. (2004). "[Genetically modified plants and food safety. State of the art and discussion in the European Union]." Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz **47**(9): 826-833.

Schuster, E., N. Dunn-Coleman, J. Frisvad and P. Van Dijck (2002). "On the safety of Aspergillus niger– a review." Applied microbiology and biotechnology **59**(4-5): 426-435.

Short, J. M. (1997). "Recombinant approaches for accessing biodiversity." Nat Biotechnol **15**(13): 1322-1323.

Soler, E., A. Le Saux, F. Guinut, B. Passet, R. Cohen, C. Merle, A. Charpilienne, C. Fourgeux, V. Sorel, A. Piriou, I. Schwartz-Cornil, J. Cohen and L. M. Houdebine (2005). "Production of two vaccinating recombinant rotavirus proteins in the milk of transgenic rabbits." Transgenic Res **14**(6): 833-844.

Sørensen, H. P. and K. K. Mortensen (2005). "Advanced genetic strategies for recombinant protein expression in< i> Escherichia coli</i>." Journal of biotechnology **115**(2): 113-128.

Srivatsan, A., Y. Han, J. Peng, A. K. Tehranchi, R. Gibbs, J. D. Wang and R. Chen (2008). "High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies." PLoS Genet **4**(8): e1000139.

Stein, E. A. and E. H. Fischer (1958). "The resistance of alpha-amylases towards proteolytic attack." J Biol Chem **232**(2): 867-879.

Struhl, K. (1989). "Molecular mechanisms of transcriptional regulation in yeast." Annu Rev Biochem **58**(1): 1051-1077.

Sun, X.-l. and H.-y. Peng (2007). "Recent advances in biological control of pest insects by using viruses in China." Virologica Sinica **22**(2): 158-162.

Thomsen, K. K. (1983). "Mouse α-amylase synthesized by Saccharomyces cerevisiae is released into the culture medium." Carlsberg research communications **48**(6): 545-555.

Tschopp, J., G. Sverlow, R. Kosson, W. Craig and L. Grinna (1987). "High-level secretion of glycosylated invertase in the methylotrophic yeast, Pichia pastoris." Nature Biotechnology **5**(12): 1305-1308.

Tschopp, J. F., P. F. Brust, J. M. Cregg, C. A. Stillman and T. R. Gingeras (1987). "Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris." Nucleic Acids Res **15**(9): 3859-3876.

Valero, E., B. Cambon, D. Schuller, M. Casal and S. Dequin (2007). "Biodiversity of Saccharomyces yeast strains from grape berries of wine-producing areas using starter commercial yeasts." FEMS yeast research **7**(2): 317-329.

van Berkel, P. H., M. M. Welling, M. Geerts, H. A. van Veen, B. Ravensbergen, M. Salaheddine, E. K. Pauwels, F. Pieper, J. H. Nuijens and P. H. Nibbering (2002). "Large scale production of recombinant human lactoferrin in the milk of transgenic cows." Nat Biotechnol **20**(5): 484-487.

Van Hove, J. L., H. W. Yang, L. M. Oliver, M. F. Pennybacker and Y. T. Chen (1997). "Purification of recombinant human precursor acid alpha-glucosidase." Biochem Mol Biol Int **43**(3): 613-623.

Vaughn, J., R. Goodwin, G. Tompkins and P. McCawley (1977). "The establishment of two cell lines from the insectSpodoptera frugiperda (Lepidoptera; Noctuidae)." In vitro **13**(4): 213-217.

Vedadi, M., C. H. Arrowsmith, A. Allali-Hassani, G. Senisterra and G. A. Wasney (2010). "Biophysical characterization of recombinant proteins: a key to higher structural genomics success." J Struct Biol **172**(1): 107-119.

Verdoes, J. C., P. Punt and C. Van Den Hondel (1995). "Molecular genetic strain improvement for the overproduction of fungal proteins by filamentous fungi." Applied microbiology and biotechnology **43**(2): 195-205.

Vermelho, A. B., M. N. Meirelles, A. Lopes, S. D. Petinate, A. A. Chaia and M. H. Branquinha (1996). "Detection of extracellular proteases from microorganisms on agar plates." Mem Inst Oswaldo Cruz **91**(6): 755-760.

Walter, P. and G. Blobel (1982). "Preparation of microsomal membranes for cotranslational protein translocation." Methods in enzymology **96**: 84-93.

Wang, L., D. Ridgway, T. Gu and M. Moo-Young (2005). "Bioprocessing strategies to improve heterologous protein production in filamentous fungal fermentations." Biotechnol Adv **23**(2): 115-129.

Waters, M. G., E. A. Evans and G. Blobel (1988). "Prepro-alpha-factor has a cleavable signal sequence." Journal of Biological Chemistry **263**(13): 6209-6214.

Wu, C. Y., Y. C. Yeh, Y. C. Yang, C. Chou, M. T. Liu, H. S. Wu, J. T. Chan and P. W. Hsiao (2010). "Mammalian expression of virus-like particles for advanced mimicry of authentic influenza virus." PLoS One **5**(3): e9784.

Wurm, F. M. (2004). "Production of recombinant protein therapeutics in cultivated mammalian cells." Nat Biotechnol **22**(11): 1393-1398.

Yoshihara, T., F. Takaiwa and F. Goto (2005). "Improving rice nutrition: challenges and practical approaches for iron fortification." Food Nutr Bull **26**(4): 416-418.

Zhang, Z., J. Gildersleeve, Y. Y. Yang, R. Xu, J. A. Loo, S. Uryu, C. H. Wong and P. G. Schultz (2004). "A new strategy for the synthesis of glycoproteins." Science **303**(5656): 371-373.