Universidade de Aveiro
Ano 2012
Departamento de Biologia

Mohammad Zahangir Alam

Quantificação de erros de incorporação em proteínas recombinantes

Quantification of amino acids misincorporation in recombinant proteins

Universidade de Aveiro Departamento de Biologia
Ano 2012

Mohammad Zahangir Alam

Quantificação de erros de incorporação em proteínas recombinantes

Quantification of amino acids misincorporation in recombinant proteins

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica do Doutor Manuel António da Silva Santos, Professor Associado do Departamento de Biologia da Universidade de Aveiro

o júri

presidente

Doutor Rui Miguel Pinheiro Vitorino
Investigador associado do Departamento de Química da Universidade de Aveiro

Doutor Maria do Céu Gomes dos Santos
Professor Auxiliar do Departamento de Biologia da Universidade de Aveiro

Doutor Manuel António da Silva Santos
Professor Associado do Departamento de Biologia da Universidade de Aveiro

Doutor Lapo Ragionieri
Investigador Associado do Departamento de Biologia da Universidade de Aveiro

Acknowledgements

At first, I would like to express my profound gratitude to my supervisor Prof. Manual A S Santos for accepting me in RNA biology lab, his excellent guidance and supervision throughout the thesis. His stimulating ideas, suggestions and availability during his busyness, helped me a lot towards the development of this thesis.

I am also really very grateful to my co supervisor Dr. Lapo Ragionieri for his tremendous support from the beginning to the end of this thesis preparation particularly during the whole experimental work and also carefully checking the drafts. His continuous in hand training was very helpful for me to develop skill in different molecular techniques.

I am indebted to all the amazing colleagues and members of RNA biology laboratory of University of Aveiro for their well cooperation since my first day on the lab.

I want to give my special thanks to European Commission (EACEA) for providing me financial support through Erasmus Mundus scholarship ("mobility for life" program) during my stay at University of Aveiro, Portugal.

I am grateful to my university authority where I worked as a faculty for granting me study leave.

I owe my deepest gratitude to my parents for their blessings, patience and understanding to allow me follow through on my career, to my loved ones for their faith, hope and support on this journey.

Especially, I would like to give my special thanks to my daughter Esha and my wife whose patient and love during staying abroad enabled me completing this work.

palavras-chave β-galactosidase, GST-tag, a solubilidade da proteína, misincorporation do aminoácido, proteína recombinante

resumo

A síntese de proteínas de acordo com o código genético é essencial para manter proteoma estável e para a homeostase celular. No entanto, os erros podem ocorrer naturalmente durante a síntese da proteína a partir do seu mRNA, variando entre 10^{-3} a 10^{-4} erros por codão. Estes erros ocorrem com mais frequncia em proteínas recombinantes sobre-expressas em hospedeiros heterólogos. Quantidades crescentes de proteína não-funcional estão geralmente relacionados à tradução em condições de stress. Neste estudo utilizou-se *Saccharomyces cerevisiae* como um organismo hospedeiro para expressar o gene lacZ-GST para quantificar o erro de tradução. A levedura foi tratada com diversos agentes de stress tais como o etanol, o crómio (CrO_3), e aminoglicósido antibiótico - geneticina (G418). A incorporação de erros foi estudada em proteína solúvel e insolúvel para determiner se os erros de tradução aumentam a agregação de proteína. Usando esta abordagem, verificou-se que o stress aumenta o erro de tradução para níveis de 5.6×10^{-3} a 8×10^{-3}, 60 - 80 vezes mais que o nível normal. Esta taxa de erro inesperadamente elevada tem implicações para a utilização terapêutica de proteínas recombinantes.

keywords

β- galactosidase, gst-tag, protein solubility, amino acids misincorporation, recombinant protein

abstract

The synthesis of protein according to genetic code of a gene determines the basis of life and a stable proteome is necessary for cell homeostatis. Faithful translation of protein give gurantee of cell survival. However, errors occur naturally during translation of protein from its mRNA, which varies from 10^{-3} to 10^{-4} per codon. These errors are more frequent in recombinant protein overexpressed in heterologous hosts and affect protein functionality. The increasing amount of nonfunctional protein is often related to mistranslation of a gene under stress. In the present study, we used *Saccharomyces cerevisiae* as a host organism to overexpress *E. coli* lacZ gene fusion with GST to quantify misincorporation of amino acid in GST-β galactosidase recombinant protein. The yeast was treated with various stressors such as ethanol, chromium (CrO_3), and aminoglycoside antibiotic - geneticin (G418) to induce protein aggregation. The misincorporation of amino acids was studied in both soluble and insoluble protein fractions by mass-spectrometry to determine how much misincorporation occur and whether this is associated with protein insolubility. We found that under experimental stress conditions the misincorporation of amino acids ranges from 5.6×10^{-3} to 8×10^{-3}, which represents 60-80 fold higher than reported level. The unexpectedly high error rate has implications for the therapeutic use of the heterologous host derived recombinant proteins.

Contents

1. Introduction 9

1.1 Protein synthesis mechanism in eukaryotes 10

1.1.1 Initiation 10

1.1.2 Elongation 12

1.1.3 Termination 13

1.2 Protein synthesis errors 15

1.3 Mistranslation and heterologous protein production 17

1.4 Heterologous protein expression in yeast 18

1.5 Protein production in *S. cerevisiae* under stress 21

1.5.1 Effects of ethanol 21

1.5.2 Effects of chromium 22

1.5.3 Effects of geneticin 22

1.7 A brief overview of the β -galactosidase protein 23

2. Objectives 25

3. Materials and methods 26

3.1 Strain, plasmid and growth condition 26

3.1.1 Yeast Strain 26

3.1.2 Plasmid 26

3.1.3 Growth medium and condition 26

3.2. Sequencing and transformation of β-galactosidase (lacZ gene) 27

3.2.1 Sequencing of lacZ gene 27

3.2.2 Whole DNA extraction 27

3.2.3 Quantification of DNA 28

3.2.4 Primer design and optimization 29

3.2.5 Polymerase chain reaction (PCR) 29

3.2.6 Agarose gel electrophoresis 30

3.2.7 PCR product purification 31

3.2.8 Plasmids transformation 32

3.3 Heterologous expression of the *E. coli* lacZ gene in yeast 33
3.3.1 Growth in presence of different stressors 33
3.3.2 β-galactosidase thermal stability assay 35
3.4 Protein extraction, purification and analysis 37
3.4.1 Protein extraction 37
3.4.2 Protein purification by affinity chromatography 38
3.4.3 Protein purification with immunoprecipitation technique 40
3.4.4 SDS-PAGE analysis 40
3.4.5 Western blot analysis 42
3.4.6 Sample preparation for mass spectrometry analysis 43
3.5 Mass spectrometry analysis 44

4. Results 45
4.1 Transformations of β-galactosidase gene into *Saccharomyces cerevisiae* 45
4.1.1 Effect of different stressors on the growth of *S. cerevisiae* 45
4.2 β -galactosidase assay 48
4.2.1 Western Blot of β-galactosidase 51
4.3 Purification of β -galactosidase protein from *Saccharomyces cerevisiae* 52
4.4 Mass spectrometry analysis 54

5. Discussion 58

6. Conclusion 61

References 62

Appendices 70

1. Introduction

All three kingdoms of life from unicellular bacteria, archaea to multi cellular eukaryotes are built with protein, the fascinating molecules that perform a wide variety of functions in the cell, but different molecules such as carbohydrates, lipids, nucleotides, DNA, RNA also reside in the cell. Proteins are a diverse and abundant class of biomolecules, which constitute more than 50 % of the dry weight of cells. The central role of proteins in virtually all aspects of cell structure and function is reflected in their abundance and diversity. A protein can be made from a single polypeptide or several polypeptides. The three dimensional structure is the key to protein function that performs many vital functions of cells. Protein serves as important structural elements within the cell, can catalyze reactions for life of an organism, and can span cell membranes and control the import and export of molecules into and out of cells [1].

The central dogma of molecular biology was enunciated by Francis Crick in 1953 [2] which stated that the genetic information is transferred from DNA via RNA to the synthesis of proteins in all living organism. DNA plays a vital role in living cells. It contains the instructions for the assembly of proteins, which in turn organize the synthesis and breakdown of other molecules that form parts of the cell and living organisms. DNA and RNA works jointly to string together a different set of monomers that comprise proteins. The building blocks of proteins are the twenty naturally occurring amino acids which are joined together via a peptide bond to form the polypeptide chain. The size of protein may vary from just a few amino acids (e.g. peptide hormones) to several thousands of amino acids e.g. giant human protein, Titin have ~ 34,350 amino acids [3]. Proteins are key components of the machinery that determines which genes will express and whether mRNAs are translated into proteins. Nowadays a gene can be artificially synthesized in laboratories, which is the main focus of a new branch of biology known as *synthetic biology*. Using genetic engineering approaches it is also possible to produce novel proteins or combination of protein elements that can serve a specific function. However, proteins which have commercial and medical applications are very difficult to purify from the organism or cell in which they are produced naturally or they cannot be produced in economically viable amounts. Many pharmaceutical

and biotechnology industries turned to the production of relevant proteins in more favorable cellular factories, and this process is known as heterologous gene expression. A variety of hosts such as bacteria, mammalian cells, insect cells and yeasts are used in many instances. The development of the technology for DNA manipulation from transferring genes between organisms lead to efficient production of protein ultimately contributing to increase production of many enzymes. So far, no single expression system is suitable for universal protein expression, because there are various factors that determine the selection of the expression host, such as whether the protein to be expressed is a membrane protein or required in soluble form [4]. One of the main problems in recombinant protein production is the inability of the foreign polypeptides to reach their native conformation in heterologous host cells. This usually results in their prevalence in the insoluble cell fraction or pellets [5]. It is observed that many heterologously expressed proteins are insoluble and often solubilization is a trial and error process with relatively low success rate [6].

1.1 Protein synthesis mechanism in eukaryotes (from gene to protein)

Transcription of genetic information is the first step in gene expression where DNA is transcribed into messenger ribonucleic acid (mRNA), and translation is the latter step which ensures that the mRNA is used as the source of information for protein synthesis. The mRNAs synthesized must be translated accurately into the corresponding proteins to ensure proper cell function. The fundamental events of translation in all kingdoms of life are divided into four stages namely initiation, elongation, termination and ribosome recycling [7].

1.1.1 Initiation:

Translation initiation (figure 1) in eukaryotes is a complex process which is assisted by more than 25 polypeptides, whereas the elongation and termination phases are assisted by a limited set of dedicated factors [8]. The translation initiation involves the recognition and recruitment of mRNAs by the translation-initiation machinery. The 80S ribosome recognizes the initiation codon (AUG) of mRNAs. Many proteins that are known as eukaryotic initiation

factors (eIFs) mediate this process. The small (40S) ribosomal subunit initially binds to the 5′ end of the mRNA and scans in a 5′→3′ direction until the initiation codon is identified. The large 60S ribosomal subunit joins with the 40S subunit to form the 80S ribosome. The small ribosomal subunit, together with factors (eIFs) 3, 1, 1A and 5 and the ternary complex assemble the 43S pre-initiation complex, which comprises the methionine -loaded initiator tRNA, which recognizes the AUG codon initiations, and eIF2 coupled with GTP [7].

In yeast, the eIF4G interacts with the poly (A) binding protein (PABP), and the simultaneous interaction of eIF4E and PABP with eIF4G seems to circularize the mRNA, which brings the 3′ UTR in close proximity to the 5′ end of the mRNA [11]. The 40S ribosomal subunit is bound to the mRNA, starts to scan mRNA in 5′→3′ direction until it finds the initiation codon (AUG) in a favorable sequence context, and joins with 60S ribosomal subunit then joins to form 80S initiation complex. The initiation factors are released after the formation of the 80S initiation complex and recycled for another round of initiation. The eIF5B is required for the joining of 60S subunit, at which point the polypeptide elongation step of translation commences [12, 13].

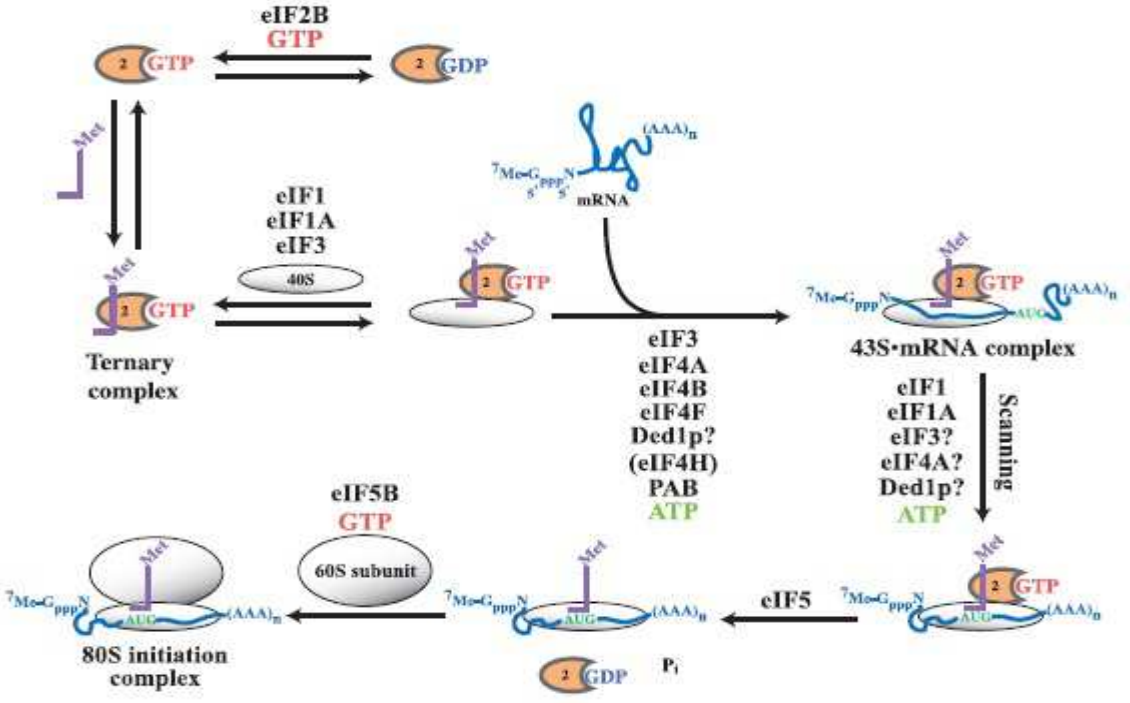

Figure 1: Current model of eukaryotic translation initiation and the roles of the initiation factors. For clarity, the 5′ and 3′ ends of the mRNA are not shown interacting (reference. 7)

1.1.2 Elongation:

The machinery used during the elongation step of translation is highly conserved across the three kingdoms of life in contrast to the initiation and termination steps [7]. During elongation three distinct steps are require to add one amino acid into growing polypeptide chains and 15-20 amino acids can be added per second [18]. Elongation of a peptide chain begins with selection of an aminoacyl -tRNA by the ribosomal A site (figure 2). Aminoacyl-tRNA binding is facilitated by elongation factor eEF1A, which forms a ternary complex with GTP. After binding of correct aa-tRNA to the A site, eEF1A GTPase activity is initiated and eEF1A-GDP releases the aa-tRNA into the A site for peptide bond formation. The ribosome catalytic center (peptidyl transferase) catalyses the formation of the peptide bond between the peptidyl tRNA and the incoming aa-tRNA in the A site [14]. The A-site aa-tRNA is then translocated to the P site and the P site tRNA is translocated to the E site of a ribosome, leaving the A site free to accept a new incoming aa-tRNA. A eukaryotic elongation factor (eEF2) hydrolyzes GTP during translocation [15] and this cycle repeats until a stop codon enters in the ribosomal A site and the termination process initiates. Accurate protein synthesis achieved in a two step process during elongation – first, it gets rid of ternary complexes bearing the wrong aminoacyl-tRNA before GTP hydrolysis occurs. Second, if this screening fails, it can still eliminate the incorrect aminoacyl-tRNA in a kinetic proofreading step before the wrong amino acid is incorporated into the growing polypeptide chain [16, 17]. A balance between speed and accuracy in the elongation step is necessary for successful protein synthesis.

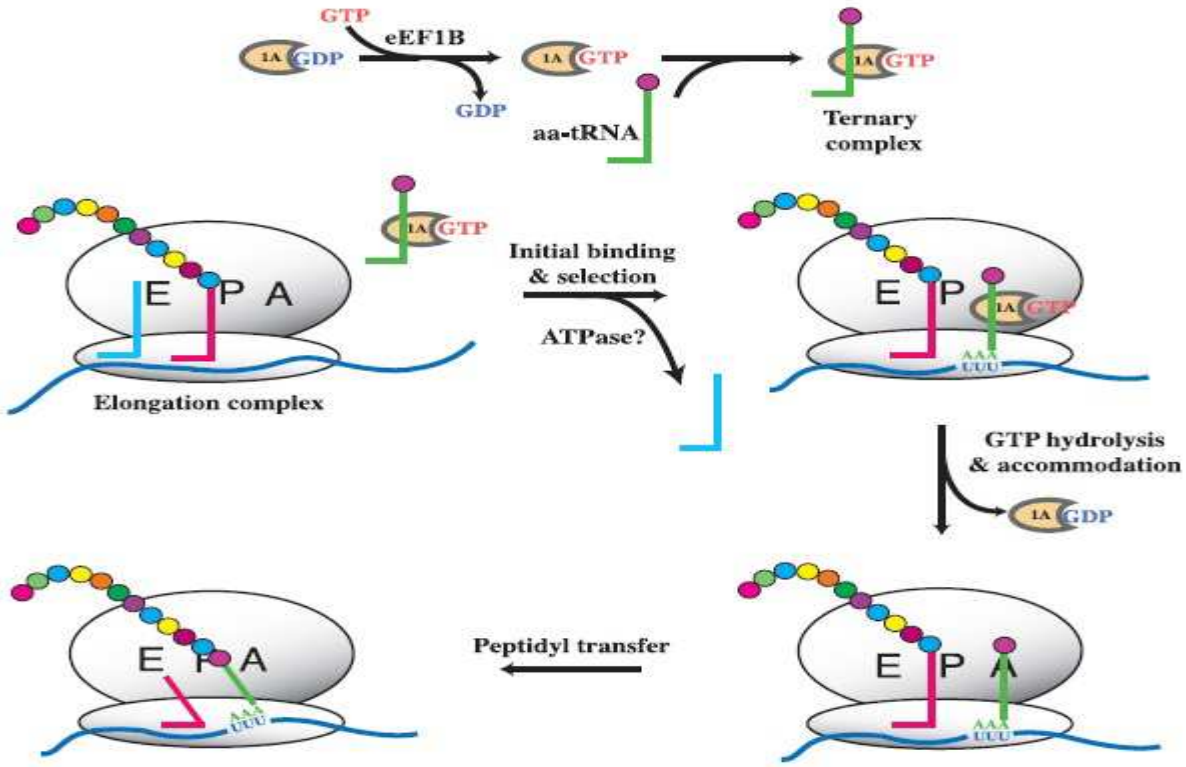

Figure 2: Model of the elongation step of protein translation in eukaryotes. The lines between the UUU and UGA codons are not meant to represent mRNA, but are merely for clarity (reference. 7)

1.1.3 Termination:

The step of termination of translation occurs when a stop codon reaches the ribosomal A site. This triggers dissociation of the polypeptide after the hydrolysis of the ester bond that links the polypeptide chain to the P site tRNA. The peptidyl transferase catalyzes the hydrolysis reaction in response by the release factor (RF1), which decode stop codons presented in the ribosomal A site. Another release factor RF2 is a GTPase that stimulate RF1 activity regardless of the nature of the stop codon engaged by the RF (Figure 3). Eukaryotic translation is mediated by two release factors – one class 1 eRF, which recognizes all three stop codons (UAA, UAG and UGA), and another release factor, eRF3, a ribosome dependent GTPase that helps eRF1 to release the finished polypeptide. In yeast, the eRF1 and eRF3 forms a complex in the absence of ribosomes which is required for optimal efficiency of termination. The ribosomal subunits are recycled after termination for another round of initiation.

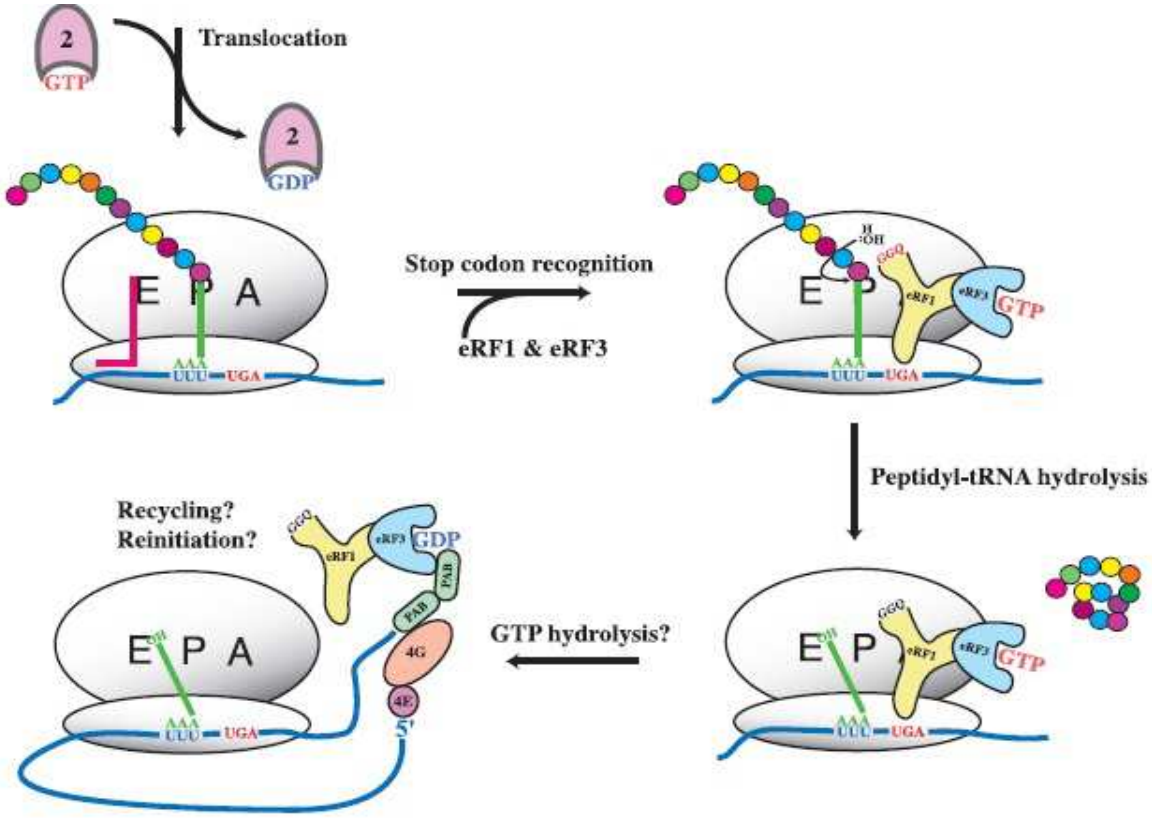

Figure 3: Model of termination step of protein translation in eukaryote. The lines between the UUU and UGA codons are not meant to represent mRNA, but are merely for clarity (reference. 7)

1.2 Protein synthesis errors:

Protein synthesis errors can be defined as any disruption in the conversion of a coding sequence into a functional protein. They arise at all steps of gene expression, from transcription to protein folding (Figure 4), and have many widespread phenotypic consequences. The fidelity of translation is an important problem in the production of pharmaceutical proteins because polypeptide chains containing incorrect amino acids are often difficult to remove from the product. Normally, translation is a remarkably accurate process. Cells have their own mechanism to control over translation. However, a variety of translational errors have been observed during overexpression of proteins in *Escherichia coli*, including frame shifts, premature truncation, read-through, and amino acid misincorporation [19-23]. Amino acid misincorporations during translation are estimated to occur once in every 10,000 codons translated [19]. However, amino acid misincorporation can occur at a much higher frequency in recombinant proteins that are overexpressed [24]. It was estimated that at this error rate, 18% of all average length protein (400 amino acids) molecules will contain at least one misincorporated amino acid [25]. For example, Bogosian et al., [26] found that overexpression of bovine somatotropin in *E. coli* caused up to 14 % of all the methionine residues to be substituted by norleucine.

Translation errors arise because tRNA is mistakenly acylated to the wrong amino acid or because a codon is improperly paired with its cognate anticodon in tRNA. The studies of norleucine incorporation in bovine somatotropin have shown that incorrect tRNA charging can be caused by limiting amounts of the respective amino acid in the cell and/or induction of pathways leading to the synthesis of unnatural amino acids [26]. Norleucine incorporation could be prevented by using a mutant strain that cannot synthesize norleucine and/or growing the cells in methionine-rich media. However, if there is reduced availability of the cognate aminoacyl-tRNA for a codon at the A site of a ribosome, other tRNAs whose anticodons differ by only one base are more likely to bind to the codon and cause errors by inserting their amino acids [28].

The other source of amino acid misincorporation, due to improper codon-anticodon pairing, depends upon the codons used by the organism. Evidence suggests that the fidelity of codon-anticodon pairing is higher for rare codons [29]. It was observed that the expression systems of yeast and *E. coli* are more or less similar in terms of fidelity but codon usage

patterns in the two species vary greatly for certain amino acids [30, 31]. In yeast, for instance, the preferred codon for Arginine is 5′-AGA-3′ with its corresponding tRNA accounting for over 50% of the total $tRNA^{Arg}$ available in the cell. On the other hand, in *E. coli*, the preferred codon for the Arginine is 5′-CGC-3′ or 5′-CGU-3′. According to tRNA availability, the 5′-AGA-3′ codon is the rarest codon used by *E. coli* which accounts only 4% of the total $tRNA^{Arg}$ population present in the cell [32]. Thus, expression of *E. coli* proteins in yeast might be constrained by codon usage.

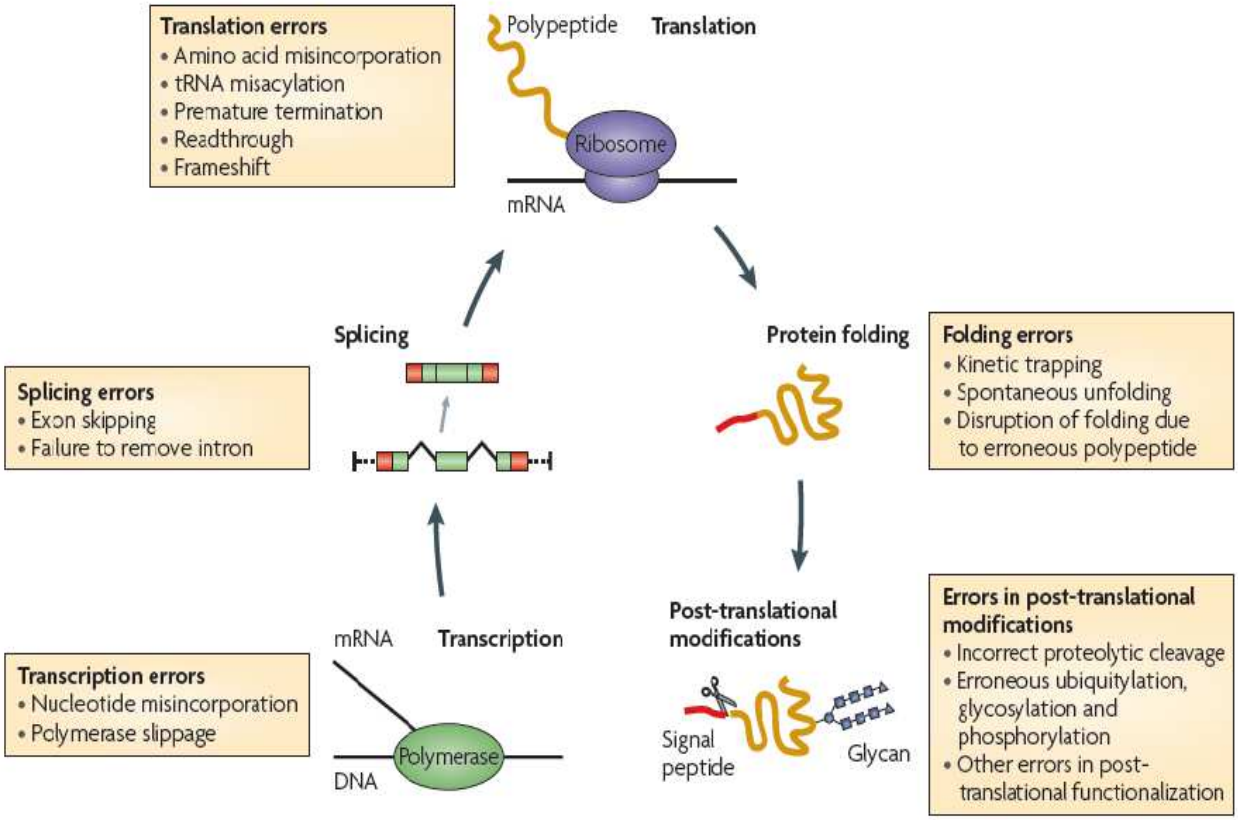

Figure 4: Sources of errors in eukaryotic protein synthesis (reference 3)

1.3 Mistranslation and heterologous protein production

Mistranslation can be described as the misincorporation of erroneous amino acid during protein synthesis which may prevent a protein from reaching its native conformation and functional state, producing - misfolded protein aggregates. A variety of factors determined the accurate protein synthesis during translational stage such as codon usage, tRNA and amino acid pool, ribosome concentration, mRNA structure, availability of GTP etc. One of the most important factors that should reduce the level of mistranslation is the aminoacylation reaction, which is highly due to the existence of both pre- and post-transfer editing mechanisms, whenever these reactions are not working properly the level of mistranslation rises dramatically. In general, it is well accepted that during protein synthesis there is an average frequency of around 1 misincorporated amino acid per 10^4 codons. Higher degrees of mistranslation are usually associated with mutagenicity and toxicity in bacteria.

In addition, heterologous overexpression of a foreign gene can increase cellular stress. This stress often leads to misincorporation of amino acids which is specially important during the production of proteins for clinical applications [85]. Although these proteins may be functional, misincorporated amino acids could potentially result in antigenic reactions, which can cause health problems. Current proteomic technologies are being used to identify and quantify amino acids misincorporated into proteins by analyzing theoretical and experimental mass spectrometry data.

1.4 Heterologous protein expression in yeast

Recombinant DNA technology has opened the door to a new world of biotechnological products for the production of heterologous proteins in microorganisms, like, for instance, pharmaceutical products and industrial enzymes. It is possible now to produce large quantities of proteins in industrial scale at low price and with constant high quality by means of recombinant DNA technology. In 1982, the first product of recombinant DNA technology was produced by Eli Lilly [33], namely insulin, a life saving drug for diabetic patients. Since then many proteins and enzymes have been produced worldwide for food and medical purposes. Moreover, using recombinant technology proteins can be easily engineered according to specific necessity. The most commonly used and well studied organism for the heterologous expression of proteins is *E. coli.* However, a variety of alternative hosts have been developed as the nature of the recombinant proteins being expressed became more complex.

In this context, yeasts are very useful hosts because they have several advantages over other microorganisms. *Saccharomyces cerevisiae* (figure 5) is a unicellular eukaryote which has many ideal properties for biological research, such as: rapid growth, nonpathogenic, discrete cells, ability for replica plating, an accommodating DNA transformation system [34], well defined genetics, cheap and commercially available mutants. Moreover, many human genes related to disease have orthologs in yeast [35], meaning that the same genes have been conserved through time in species, maintaining the same or similar function. Many scientific studies show that nearly 50% of human genes implicated in heritable diseases have yeast homologs [36] or at least 31% of proteins encoded by yeast genes have human homologs. Moreover, cell cycle mechanisms, signal transduction, and metabolic and regulatory mechanisms between yeast and other eukaryotic organisms are conserved. The close homology to higher order eukaryotes and simplicity of yeast genetics allows the introduction of mammalian genes into yeast and the analysis of their function [37]. *S. cerevisiae* was the first eukaryote with the genome completely sequenced, with the possibility of being transformed with plasmids and to have gene knockouts strains [35]. Yeast has shifted molecular genetics research to a systems level approach of functionally characterizing parts of the genome. *S. cerevisiae* has also been used for centuries in food production, especially in bakery and brewing industries, and is Generally Recognized As Safe (GRAS) by the

American Food and Drug Administration (FDA). By contrast, mammalian cells may contain oncogene or viral DNA, while other prokaryotic organisms may have toxic cell wall pyrogens (endotoxins). In addition, yeasts have the ability to perform eukaryotic processing steps during polypeptides synthesis. Post-translational processing and modification are required for the function of many proteins such as disulphide bond formation, N and O–linked glycosylation, proteolytic maturation of prohormones etc. Although yeast has many advantages, there are also some limitations. For example, protein product yield is normally low and the secretion system is often inefficient [35].

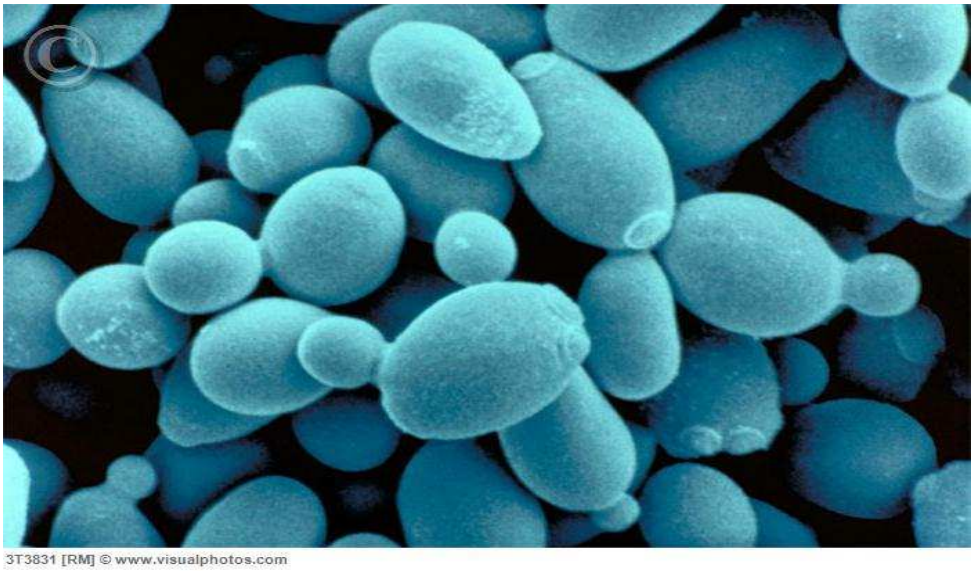

Figure 5: Picture of the budding yeast *Saccharomyces cerevisiae* (Retrieved from www.visualphotos.com)

Foreign proteins can be expressed either extracellularly or intracellularly, for recovery from the cytosol, nucleus, mitochondria, or peroxisomes. However, there are several factors affecting each step of efficient protein production, such as DNA copy number, promoter strength, mRNA stability, codon usage pattern etc. Post-translational modifications are necessary for correctly folding of protein. If the protein is not folded properly, it is usually degraded and the building blocks are recycled. Most of the recombinant proteins produced in *S. cerevisiae* are intracellular, i.e., they are produced and retained within cell. These proteins may be produced in a soluble, biologically active form, or they can be insoluble and precipitate in the cell as inclusion bodies [38]. The recovery of both forms of proteins

requires the disruption of the yeast cells as an initial step in the extraction and purification of proteins. Some of the examples in which direct intracellular expression have been successfully used to express heterologous gene products include the hepatitis B virus surface antigen (HBsAg) [39], human superoxide dismutase (hSOD) [40], fibroblast growth factor [41], human immunodeficiency virus type-1 (HIV1) env and gag polypeptides [42, 43], surface antigens of malaria parasites [44], rat cytochrome C [45], α1-antitrypsin proteinase inhibitor [50] and human viral surface glycoproteins [47].

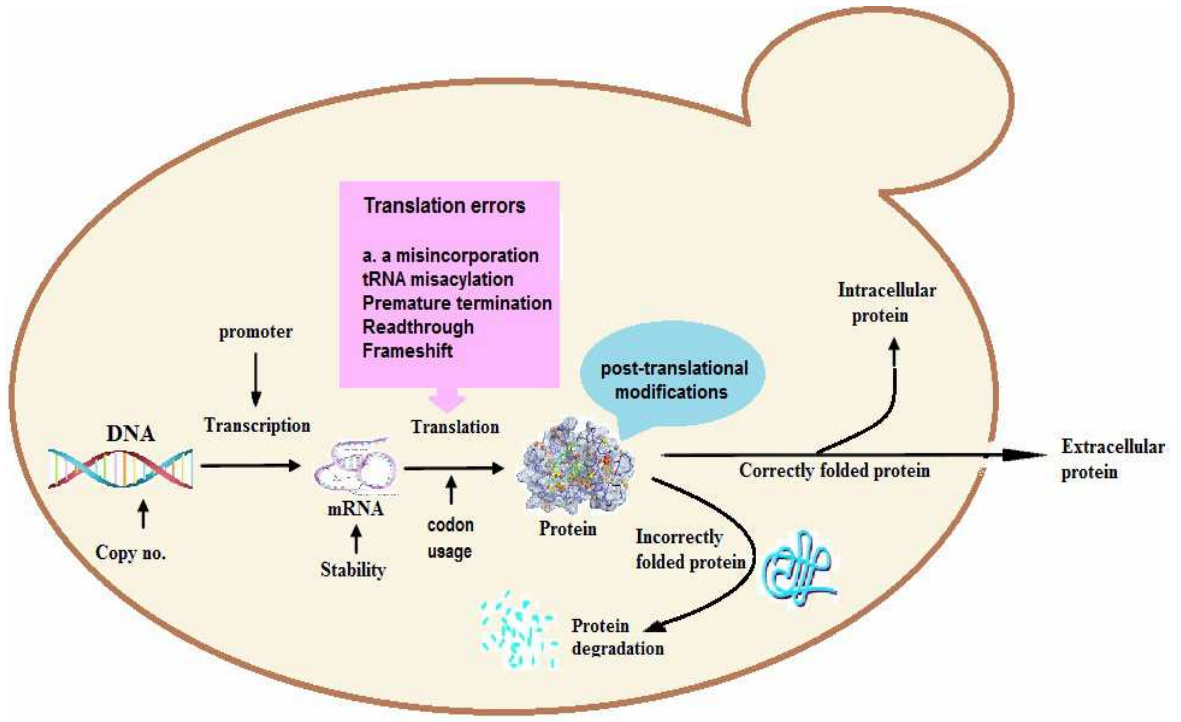

Figure 6: Schematic drawing of the most important steps from DNA to secreted protein in *Saccharomyces cerevisiae*.

1.5 Protein production in *S. cerevisiae* under stress

Protein translational accuracy depends on the growth conditions and energy availability. Under stress, cells induce heat shock proteins (Hsps)s and other molecules that confer stress protection. The molecular responses elicited by the cells dictate whether the organism adapts, survives, or, if injured beyond repair, undergoes death. In some instances, stress responses may eventually lead to stress tolerance as a long term defense mechanism toward damaging agents. The regulation of stress response includes transcriptional, posttranscriptional and posttranslational mechanisms.

Growth and metabolic activity of yeast are functions not only of the genetic background of the cell but also of the composition of the growth medium and other physical growth parameters such as temperature, pH, oxygen supply, osmotic conditions, the absence of toxic compounds, nutrient limitation, exposure of various drugs and so on. Any condition deviating from normal can be defined as a stress condition, which invokes particular stress responses to meet the hazardous consequences of cell damage. Yeast has served as a versatile model to dissect different types of these responses. Yeast exhibits a large degree of phenotypic variation and altered gene expression due to stress. The study of Kvitek and coworkers [71] also showed that yeast isolated from diverse niches have a wide range of stress sensitivity and over 70% of its genes display significant alteration in expression. However, effects of stress resulting from misincorporation of amino acids are not clear. It is assumed that if cells are exposed to stressful conditions, misincorporation of amino acids into de novo synthesized protein tends to rise [81]. In the present study three stressors, namely ethanol, chromium and geneticin (antibiotic) were chosen to examine the effect of stress on the solubility of β-galactosidase protein and on the misincorporation of amino acids.

1.5.1 Effects of ethanol

Ethyl alcohol frequently named as ethanol is well known as an inhibitor of growth and reproduction of many microorganisms including bacteria, fungi, protozoa and viruses. In yeast cells ethanol can damage mitochondrial DNA [48], and cause inactivation of some enzymes such as hexokinase and dehydrogenase. The yeast *Saccharomyces cerevisiae* is able to respire on ethanol and is considered one of the most ethanol resistant eukaryotes. For this reason it is widely used for the production of alcoholic beverages and leavened bread [49]. Ethanol influences cell metabolism, limits cells growth, denatures intracellular proteins and

glycolytic enzymes and reduces their activity [50], lowers the rate of mRNA and protein levels [53], reduces glycolytic enzyme activity and increases protein denaturation [54]. It also inhibits cell division and viability [52] and influences macromolecular biosynthesis by inducing the production of heat shock proteins [55]. Walker [51] mentioned that the major sites for ethanol action in yeast cells are the plasma membrane, hydrophobic proteins of cell and mitochondrial membranes, nuclear membrane, vacuolar membrane, endoplasmic reticulum and hydrophilic proteins of cytoplasm. Other works [83, 84] also suggesting that ethanol disruption of membrane function may affect the delivery of amino acids into the cell. Finally, ethanol has been observed to reduce the translation fidelity of poly (U) in rat brain extract [80]. Similarly a strong effect was observed in human cultured cells on read through of stop codon UGA than UAG [74].

1.5.2 Effects of chromium

Any metal at higher level of concentrations are toxic and harmful for an organism, although certain metals at moderate level are biologically essential [57]. Chromium has been linked to oncogenesis [56]. The underlying mechanisms of chromium toxicity remain unknown. In this regard, *Saccharomyces cerevisiae* provides an excellent model to elucidating biological questions related to metal toxicity [58]. Chromium (Cr) exists primarily in the Cr (III) and Cr (VI) oxidation states and Cr (VI) has been reported as being more toxic in the environment due to its higher solubility and mobility [59]. It was reported that chromium exposure tends to increase insoluble protein in yeast [60]. This study also suggested that mistranslation of mRNA is a primary cause of chromium toxicity, suggesting that protein aggregation may be related to chromium's effect on the translational machinery.

1.5.3 Effects of geneticin

Geneticin sulphate (G418) is an aminoglycoside antibiotic similar in structure to gentamicin that has originally been derived from *Micromonospora rhodorangea.* Aminoglycoside antibiotics, such as paromomycin, act on the eukaryotic 80S ribosome and induce errors in translation. Geneticin can cause a bypass of nonsense mutations during translation and increases the misincorporation of amino acids into protein [70]. At high dose G418 can also block polypeptide synthesis by inhibiting the elongation step in both prokaryotic and eukaryotic cells [61].

1.6 A brief overview of the β-galactosidase protein

β-galactosidase is a hydrolase enzyme (EC 3.2.1.23) produced by the lacZ gene in *E. coli* and some other enteric bacteria. This protein specified by the first structural gene of the lactose (*lac*) operon, which is activated in the presence of lactose when glucose level is low [62]. This enzyme hydrolyzes β 1→4 linkage of lactose to D-galactose and D-glucose [63, 64]. The *lac* operon system consists of three structural genes (lac-Z, lac-Y and lac-A), a promoter, a terminator, a regulator and an operator. The lac-Y encodes β-galactoside permease, a membrane - bound transport protein pumps lactose into the cell. Both of the lac-Z and lac-Y are necessary for lactose catabolism. The lac-A encodes thiogalactoside transacetylase which is an enzyme involved in detoxification. Another gene (lac-I) encodes a regulatory repressor protein, disaccharide allolactose which binds to the operator site and prevent the expression. Transcription is initiated when the repressor is removed from the operator by inducer [64, 65]. *E. coli* β-galactosidase was first sequenced in 1970 and its structure was determined in 1994. The β-galactosidase tetramer formed with four identical polypeptide chains (A to D) and each of the chain has 1023 amino acids. Each polypeptide (monomer) is made up of five domains and 50 residues at the N-terminal.

To complete four functional active sites, domain 2 of different monomers extends into the neighboring active sites called activating interface. Monomer A donates its Domain 2 loop to complete the active site of monomer D and the reciprocal monomer D donates its Domain 2 to complete the active site of monomer A (figure 7). Similar kind of contribution of domains happens between monomer B and C [63, 67]. The third domain contains the active site which is made up of elements from two subunits of the tetramer. The disassociation of the tetramer into dimmers removes the critical elements of the active site. The amino-terminal sequence of β-galactosidase, the α-peptide involved in α-complementation, participates in a subunit interface. Its residues 22-31 help to stabilize a four – helix bundle which forms the major part of that interface, and residue 13 and 15 also contributing to the activating interface. These structural features provide a rationale for the phenomenon of α-complementation, where the deletion of the amino-terminal segment results in the formation of an inactive dimer.

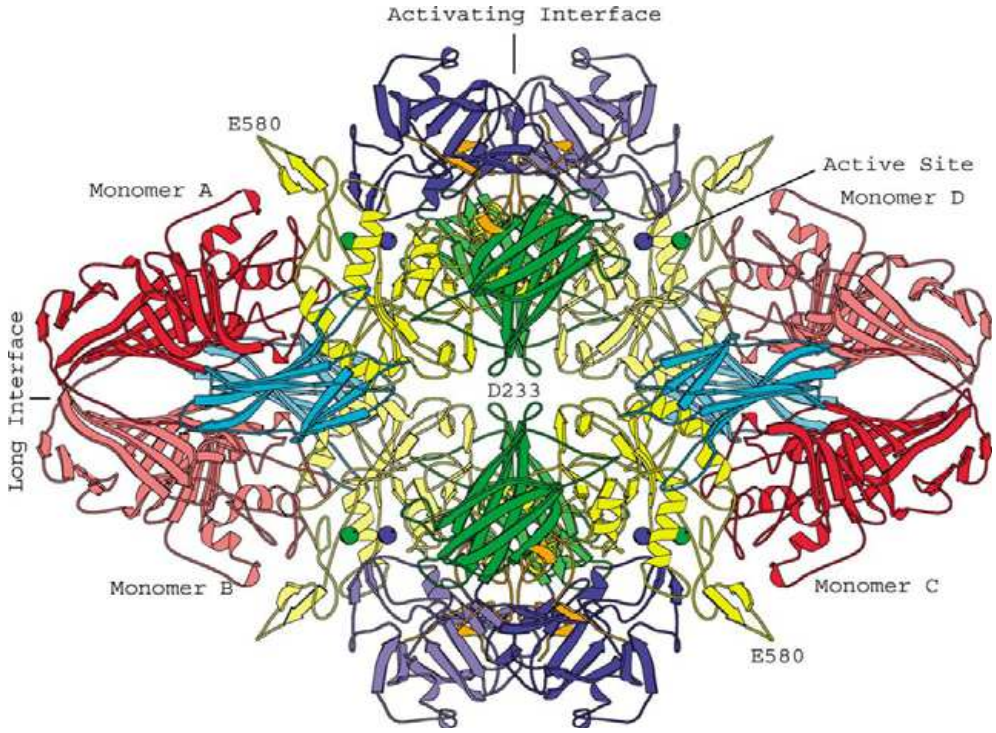

Figure 7: β-galactosidase tetramer looking down one of the two-fold axes. Colour indicates different domain, orange- domain 1, blue- domain 2, yellow- domain 3, cyan- domain 4 and red- domain 5. Light and dark colour shades is to separate the same domain in different subunits [63]

The β-galactosidase assay is frequently used in genetics, molecular biology, and other life sciences. Branscomb and Gallas [75] suggested that the thermostability of β-galactosidase affected due to sense codon misreading and the rate of thermoinactivation will give a relative measure of amino acid misincorporation. This assay, for instance, may be used as a reporter system for monitoring amino acid misincorporations in β-galactosidase during translation. This enzyme is also used for α-complementation detected using X-gal, which forms an intense blue product after cleavage by β-galactosidase, and is easy to identify and quantify in blue white screens [68]. Its production may be induced by a non-hydrolyzable analog of allolactose, IPTG, which binds and releases the lac repressor from the lac operator, thereby allowing the initiation of transcription to proceed. Another chromogenic substrate o-nitrophenyl-β-D-galactopyranoside (ONPG) is available to indicate the presence of β-galactosidase during cell expression. ONPG hydrolyzed by β-galactosidase produce the glycoside and o-nitrophenol. Glycoside has no optical density at visible wave length while free o-nitrophenol gives yellow colour at 420nm [63, 69].

2. Objectives

The objectives of this research work were the following:

1) To study the solubility of *E. coli* β-galactosidase protein in yeast and to optimize its purification using affinity chromatography.

2) To quantify amino acid misincorporation in a recombinant protein GST-β gal under environmental stress.

3) A specific aim of the study was to establish a new protocol for the extraction and purification of heterologous proteins from *Saccharomyces cerevisiae.*

3. Materials and methods

3.1 Strain, plasmid and growth condition

3.1.1 Yeast strains

The yeast *Saccharomyces cerevisiae* was used in this study for the heterologous over expression of β-galactosidase. Two different yeast strains were used: the diploid strain **CEN.PK2** (MATa/MATα, ura3-52/ura3-52, trp1-289/trp1-289, leu2-3_112/leu2-3_112, his3Δ1/his3Δ1, MAL2-8C/MAL2-8C, SUC2/SUC2) and the haploid strain **BY4742** (MATα, his3Δ1, leu2Δ0, lys2Δ0 and ura3Δ0).

3.1.2 Plasmids

The multicopy plasmid vector pGL-C1 carrying *E. coli* lacZ gene, producing β-galactosidase protein, was transformed in *S. cerevisiae* in the diploid **CEN.PK2** strain under the control of the GPD promoter as a fusion with glutathione S-transferase (GST- β gal). The pUKC815 plasmid containing the yeast phosphoglycerate kinase (PGK1) gene promoter and the N-terminal 33 amino acids fused in-frame to the *E. coli* lacZ gene coding for β-galactosidase was expressed in the haploid **BY4742** strain. The plasmid pUKC815 was constructed by cloning a 0.9 kb *ClaI-BamHI* fragment from pUKC350 [72] into the plasmid YCp50 [92] digested with *ClaI* and *BamHI,* thereby generating pUKC814. A 3.2-kb *BamHI* fragment from pUKC350 carrying the *lacZ* gene was then cloned into pUKC814 digested with *BamHI,* thereby generating pUKC815 [73].

3.1.3 Growth medium and growth conditions

All strains carrying plasmids for the heterologous expression of β-galactosidase were grown at 30° C in minimal medium (0.67% bacto yeast nitrogen without amino acids, 100 μg/ml of each of the required amino acids, 2% glucose and 2% bacto agar for solid medium). More specifically the diploid *S. cerevisiae* strain CEN.PK2 transformed with pGL-C1 plasmid was grown in minimal media lacking tryptophan, while haploid strain BY 4742 transformed with pUKC815 plasmid was grown in minimal medium lacking Uracil (Appendix A). All strains were preserved at -80° C in MM with 40% glycerol.

3.2 Sequencing and transformation of β-galactosidase (lacZ gene)

3.2.1 Sequencing of the lacZ gene

In order to confirm the correct sequence of the lacZ gene within the different constructs, two plasmids were extracted using a miniprep kit (Fermentas) and then sequenced using specific primers. Whole genome DNA of both strains carrying the two different plasmids was extracted and the lacZ gene was PCR amplified for sequencing using classic Sanger sequencing methods.

3.2.2 Whole DNA extraction

The whole DNA extraction was carried out from single colonies picked up from YPD agar plates. For this, one colony was inoculated into 10 ml of YPD broth and allowed to grow overnight (approximately 16h) at 30° C with agitation (180 rpm). The whole DNA was extracted according to instruction of the Wizard® Genomic DNA Purification Kit (Promega). Briefly, 1 mL of yeast cells was transferred to a 1.5 mL microcentrifuge tube and centrifugated at 13,000 X *g* for 2 minutes in order to pellet the cells. After removing the supernatant, the cells were resuspended thoroughly in 293 µl of 50 mM EDTA and 7.5 µl of 20 mg/ml lyticase was added and gently pipet 4 times to mix properly. The sample was incubated at 37° C for 30 minutes to digest the cell wall and, successively, cooled down at room temperature. The samples were centrifuged at 13,000 X *g* for 2 minutes and the supernatant was discarded. Then, 300 µl of Nuclei Lysis solution was added to the cell pellet and gently pipeted to mix. A 100µl of Protein Precipitation Solution was added and vortexed vigorously at high speed for 20 seconds. The sample was placed on ice for 5 minutes and centrifuged at 13,000 X *g* for 3 minutes. The supernatant containing the DNA was transferred to a clean 1.5 ml microcentrifuge tube containing 300 µl of room temperature isopropanol. The sample was gently mixed by inversion until the thread-like strands of DNA formed a visible mass and then centrifuged at 13,000 X *g* for 2 minutes. The supernatant was carefully decanted and tubes were drained on clean absorbent paper. A 300 µl of 70% ethanol at room temperature was added and gently inverted the tube several times to wash the DNA pellet. Then centrifuged at 13,000–16,000 × *g* for 2 minutes and carefully aspirated all of the ethanol. The tube was drained on clean absorbent paper and allowed the pellet to air-dry for

10–15 minutes. Then 30 µl of sterile Milli-Q water was added to the sample. Then 1.5 µl of RNase Solution was added to the purified DNA sample. The sample was vortexed for 1 second and Centrifuged briefly in a microcentrifuge for 5 seconds to collect the liquid and incubated at 37° C for 15 minutes in a thermomixer. Finally, the DNA was rehydrating by incubating at 65° C for 1 hour and store at 2–8°C. (See appendix A for protocol).

3.2.3 Quantification of DNA

After isolation of DNA, quantification and analysis of quality were necessary to ascertain the quantity of DNA obtained and the suitability of DNA sample for further analysis. The most commonly used methodologies for quantifying nucleic acids in a preparation are gel electrophoresis and spectrophotometric analysis.

Analysis of UV absorption by the nucleotides provides a simple and accurate estimation of the concentration of nucleic acids in a sample. Purines and pyrmidines in nucleic acid show maximum absorption around 260nm (e.g., dATP: 259nm; dCTP: 272nm; dTTP: 247nm) if the DNA sample is pure without significant contamination from proteins or organic solvents. The ratio of OD_{260}/OD_{280} is normally determined to assess the purity of the sample. This method is however limited by the quantity and purity of the DNA preparation. Accurate analysis of the DNA preparation may be impeded by the presence of impurities in the sample or if the amount of DNA is too little. In the estimation of total genomic DNA, for example, the presence of RNA, sheared DNA, could interfere with the accurate estimation of total high molecular weight genomic DNA. One disadvantage of the spectrophotometric analysis is that sheared or otherwise degraded DNA is indistinguishable from high quality DNA, since both have equal UV absorption. However, the amount of DNA can be quantified using the following formula:

$$\text{DNA concentration (µg/ml)} = \frac{OD_{260} \text{ x } 100 \text{ (dilution factor) x } 50 \text{ µg/ml}}{1000}$$

Spectrophotomteric Conversions for Nucleic Acids:

1 A 260 of ds DNA	= 50 µg/ml	= 50 ng/µl
1 A 260 of ss oligonucleotides	= 33 µg/ml	= 33 ng/µl
1 A 260 of ss RNA	= 40 µg/ml	= 40 ng/µl

Good quality DNA will have an A_{260}/A_{280} ratio of 1.7–2.0. An absorbance value of 1.6 does not render the DNA unsuitable for any application, but lower ratios indicate that more contaminants are present. A ratio lower than 1.8 indicates the presence of proteins and/or molecules that absorb UV. On the other hand, a ratio higher than 2.0 indicates that the samples may be contaminated with chloroform or phenol. In either case (<1.8 or >2.0) it is advisable to re-precipitate the DNA.

In the present study, DNA quantifications and quality were determined using a NanoDrop™ 1000 spectrophotometer (Thermo Scientific). This type of spectrophotometer allows the analysis of very small sample volumes (1-2 μl) and does not require cuvettes. Briefly, with the sampling arm open, 1 μL of milli-Q water was used as the blank sample. The same steps with each sample were repeated and the λ value and 260/280 ratios for purity analysis were recorded.

3.2.4 Primer design and optimization

The optimization is essential to maximize the specificity and efficiency of PCR. It includes primer sequences, appropriate primer and template DNA concentrations and annealing temperature. PCR primers are 15-30 bases long that do not contain bases complementary to each other to minimize formation of primer-dimers. A 40% - 60% G+C content is recommended for each primer, avoiding internal secondary structure and long stretches of repeated bases. Primers were designed using PerlPrimer v1.1.21 (http://perlprimer.sourceforge.net) and purchased from STAB VIDA (Portugal). For a complete list of primers see Appendix C.

3.2.5 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) is an *in vitro* method for the amplification of DNA that was introduced in 1985. Basic PCR consists of three steps: thermal denaturation of the target DNA, primer annealing of synthetic oligonucleotide primers, and extension of the annealed primers by a DNA polymerase. This three-step cycle is then repeated 30-40 cycles, each time doubling the number of product molecules. In the first step, the double-stranded DNA template is denatured by increasing the temperature to ~95º C. At this temperature, the hydrogen bonds between the complementary bases break up, yielding single DNA strands. Following this step, the annealing occurs at a temperature below the melting temperature

(Tm) of two oligonucleotide primers (usually 50–65° C, in this experiment 55° C works well), allowing them to bind to the complementary regions of the single-stranded DNA template. During the third step, extension or elongation, the *Taq* DNA polymerase recognizes the short double stranded sections of DNA created by the annealed primers and extends the primers in 5´ to 3´ direction by incorporating dNTPs. The optimal temperature for this enzymatic reaction is around 72º C. The newly created double stranded DNA molecules then serve as templates for the next cycle of denaturing, annealing and elongation.

For the master mix, 0.125 µl *Taq* DNA polymerase (1 U/µl), 2.5 µl of 10 X Buffer (500 mM KCl, 100 mM Tris-HCl (pH 9.0), 1.0% Triton X 100), 0.5 µl of dNTPs (5 mM) and 0.5 µl of each primer (10 mM) were combined and brought to a final volume of 25 µl per PCR reaction with sterile distilled water (milli-Q). The PCR protocol consisted of 30 cycles, of denaturing at 94º C for 60´, annealing at 55º C for 90´ and extension at 72º C for 60´ in a MyCycler^{TM} thermal cycler (BIO-RAD).

3.2.6 Agarose gel electrophoresis

Agarose gel electrophoresis is a method for determining the size of DNA and RNA by length and to assess their quality and yield. In this method, DNA is forced to migrate through a highly cross-linked agarose matrix in response to an electric current. In solution, the phosphates on the DNA are negatively charged, and the molecule will therefore migrate to the positive pole. At neutral pH, DNA is negatively charged and in an electric field migrates from the negative to the positive pole. There are three factors that affect migration rate through a gel; size of the DNA, conformation of the DNA and ionic strength of the running buffer.

The larger the fragments of DNA, the more easily will it become entangled in the matrix and, therefore, the more slowly it will migrate. Small fragments, therefore, run more quickly than large fragments at a rate proportional to their size. The relationship of size to migration rate is linear throughout most of the gel, except for the very largest fragments. Large fragments have a more difficult time penetrating the gel and their migration, therefore, does not have a linear relationship to size. The matrix can be adjusted by increasing the concentration of agarose (tighter matrix) or by decreasing it (looser matrix). A standard 1% agarose gel can resolve DNA from 0.2 - 30 kb in length. In the present study, 1.4% gel was

used successfully. The PCR products were separated for 30-35 minutes at a constant voltage of 100 V alongside a DNA ladder (GeneRuler^{TM} 50 bp DNA ladder Plus, Fermentas). For gel preparation 0.70 g of agarose was dissolved in 50 ml TAE 1X buffer with 0.5 µl of ethidium bromide (EtBr) mixed carefully to avoid inhalation.

In order to observe the DNA fractionated on gels, it must first stained with a dye fluoresces under ultraviolet light. Traditionally, ethidium bromide (EtBr) has been used for this purpose. Ethidium bromide is a fluorescent compound that binds strongly to double stranded DNA by intercalating between the bases. When intercalated, it strongly absorbs UV light and emits visible orange light. EtBr is mutagenic and must be handled as a hazardous waste. More recently, nontoxic dyes such as Gel Red have been introduced. The length of the DNA fragment of interest is then determined by comparison with a DNA marker of known fragment lengths.

3.2.7 PCR product purification

Enzyme contamination of DNA samples can interfere with subsequent downstream applications. In this study, QIAquick PCR Purification Kit from QIAGEN® to purify PCR amplicons were used. This kit is used for direct purification of double or single stranded PCR products (100 bp – 10 kb) from amplification reactions and DNA cleanup from other enzymatic reactions. The kit works on the basis of reversible adsorption of nucleic acids to silica-gel, in the presence of high concentrations of chaotropic salts and high pH. Chaotropic salts (e.g. urea, thiourea) disrupt hydrogen bond structures in water, affect the nucleic acids secondary structure and decrease the solubility of DNA in water.

The first step of the protocol consisted in adding the binding buffer (PB; pH 7.5 and containing the chaotropic salt guanidine hydrochloride) to the samples. Then, the DNA solution was passed through the silica-gel membrane by centrifugation to bind the DNA, followed by a washing step with an ethanol-containing buffer (PE buffer) to remove salts. Finally, the DNA was eluted in a low-salt solution (EB buffer; 10 mM Tris·Cl, pH 8.5) (Detailed protocol in Appendix A).

3.2.8 Plasmids transformation

Transformation of *S. cerevisiae* was done using the lithium acetate (LiAc) method [10]. Briefly, overnight cultures were diluted in 10mL (for 10 transformation reactions) of new and freshely prepared medium at an OD_{600} of 0.05. The cultures were grown at 30º C with 200rpm shaking until they reached to an OD_{600} of 0.4 – 0.5. Cells were then harvested by centrifugation for 5 minutes at 4000rpm. After washing with 5mL of sterile milli-Q water, the pellet was resuspended in 500µL of 0.1M LiAc solution. Around 50µL of cell suspension were transferred onto 1.5mL eppendorf tubes and cells were pelleted by centrifugation at maximum speed for 15 seconds. The supernatant was discarded and the transformation reagents were added to the pellet in the following order: 240 µL 50% (w/w) PEG, 36 µL 1.0M LiAc, 25µL single-stranded carrier DNA (2mg/mL) previously denatured and 50µL of an aqueous solution of the plasmid of interest (containing 0.1 – 1µg of plasmid). The tubes were vortexed until a homogeneous suspension was obtained. The cells were then heat-shocked in a water-bath at 42º C for 30 to 40 minutes. Cells were then harvested by centrifugation at 5000 rpm for 1 minute. The supernatant was discarded and the pellet was carefully resuspended in 100 µL of sterile milli-Q water. Each suspension was plated in selective media plates and incubated at 30º C, until transformant colonies were visible (about 3 – 4 days).

3.3 Heterologous expression of the *E. coli* lacZ gene in yeast

For the heterologous expression of the LacZ gene in *S. cerevisiae* under stress conditions two different constructs were transfected into strain CEN.PK2 and BY4742. The multi copy plasmid pGL-C1 under the control of the GPD promoter as a fusion with glutathione S-transferase (GST) was transformed into the diploid CEN.PK2 strain and the plasmid pUKC815 without the GST gene was transformed into the haploid BY4742 strain.

3.3.1 Growth in presence of different stressors

The above yeast strains transformed with the two different plasmids carrying the lacZ gene were grown at 30° C, 180 rpm until late exponential phase (between 0.6 and 0.8 OD_{600}). The growth curves were obtained by inoculating the same amount of yeast coming from the same pre-inoculum at exponential phase (0.5-0.7 OD) into media with stressor and without stressor (control) to an initial OD_{600} of 0.02 in 50 ml Erlenmeyer flasks (Figure 8). The cultures were allowed to grow for the first 8 hours and then every hour aliquots of cultures were collected and the OD_{600} was measured during the first day (around 20 hours). Since the stressors affect the yeast growth rate OD_{600} was measured during three days (around 96 hours) until the cultures reached stationary phase (OD_{600} around 2.5). Finally, the growth rate corresponding to the growth of yeast cells in exponential phase was calculated for cells growing with stressor relative to control cells without stressor. After several preliminary screens (data not shown) the sub lethal dose of each stressor was determined. The dose of the stressors tested in this experiment were ethanol (2% and 5%), CrO_3 (1 mM, 0.5 mM, 0.25 mM and 0.1 mM) and Geneticin 0.2 mg/L. The ethanol at 2% didn't showed sufficient change of growth rate and CrO_3 at higher concentration such as 1 mM, 0.5 mM, 0.25 mM were considered as lethal dose in contrast to control cells. The final concentrations of the three different stressors are represented in the following table:

Table 1: Experimental stressors

Stressors	Concentration
EtOH	5% (v/v)
CrO_3	0.1 mM
Geneticin® (G418)	0.2 mg/L

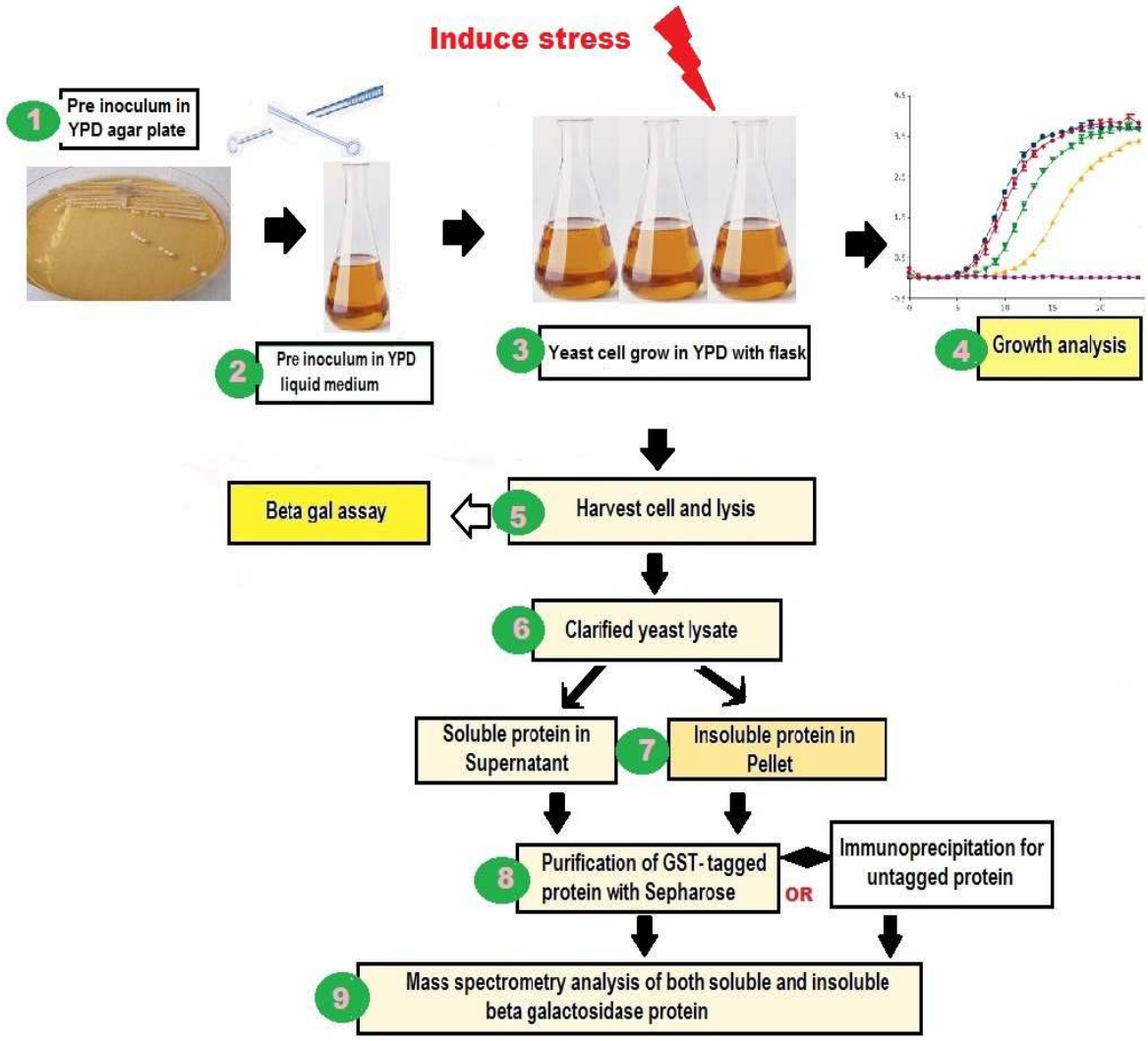

Figure 8: Schematic representation of the experimental design used in this study.

Similarly β-galactosidase was prepared from 20 ml pre-cultures that were inoculated with single colonies of transformed *S. cerevisae* and grown overnight (OD ~ 1.0-1.2) at 30º C with shaking at 180 rpm. On the following day, new cultures were started in triplicates by inoculating exactly the same amount of yeast cells coming from the same pre-inoculum into media with stressor (experimental) and without stressor (control) to an initial OD_{600} of 0.02 in 50 ml or 500 ml Erlenmeyer flasks. Cells were collected when the cultures reached an OD_{600} of ~ 1.5. Then 45 ml of culture were transferred to a 50 ml Falcon tube, pelleted at 4000 rpm in a refrigerated centrifuge (5810R, Eppendorf) at 4º C, the medium was decanted and the pellet was washed with 5 mL of 1X PBS and immediately frozen at -80º C to avoid protein degradation.

3.3.2 β –galactosidase thermal stability assay

Sense codon misreading affects the thermostability of β-galactosidase and the rate of thermoinactivation gives a relative measure of amino acid misincorporation [75]. For this reason, β-galactosidase was used as reporter protein to quantify mistranslation induced by amino acid misincorporation in the lacZ protein. The assay was adapted from [76] with some modifications.

β-galactosidase catalyzes the breakdown of the substrate lactose, a disaccharide sugar found in milk into two monosaccharide sugars, galactose and glucose. The oxygen bridge connecting the two sides of the lactose molecule is cleaved through the addition of a water molecule. The addition of the water molecule is known as hydrolysis. Since it is difficult to assay for the activity of β-galactosidase when lactose is the substrate, a lactose analog ONPG (o-nitrophenyl β-D-galactopyranoside) is used due to the similarity in the structure of ONPG and lactose (Figure 9).

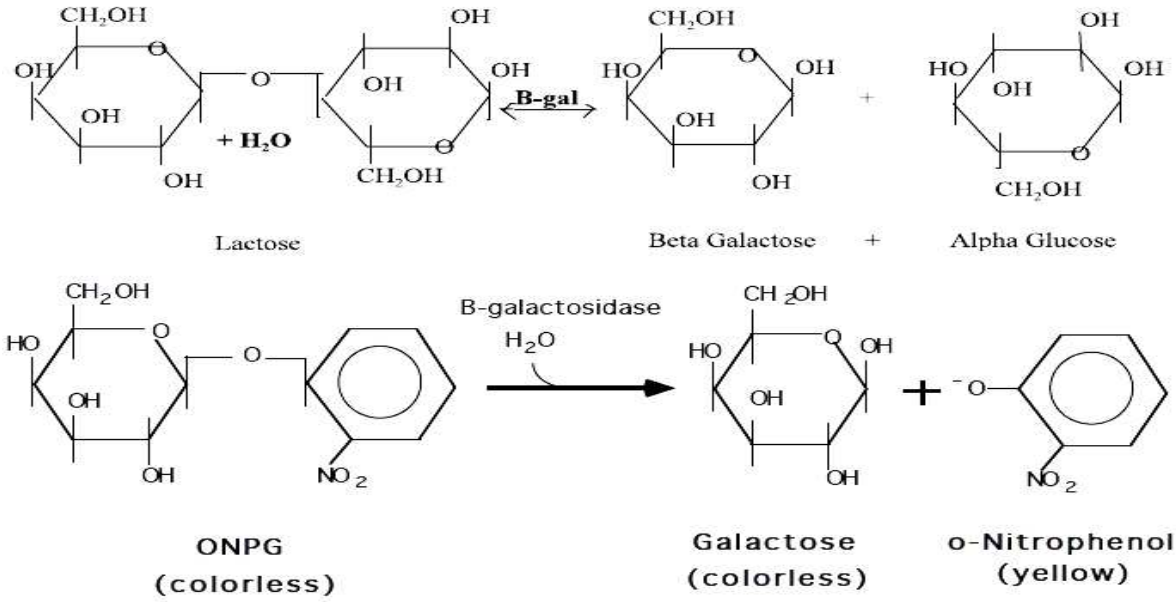

Figure 9: Similarities of structure and reaction of lactose and ONPG

The enzyme does not distinguish between lactose and ONPG, and cleaves the oxygen bridge between the two sides of the molecule, resulting in the products galactose and the o-nitrophenol. The advantage of using ONPG as the substrate is that the product ONP (o-nitrophenol) is yellow, and can be quantified by visible spectrophotometry (OD_{420}). By measuring the rate at which the color intensity increases we can calculate the activity of the

enzyme. The reaction is stopped by adding Na_2CO_3 which shifts the reaction mixture to pH 11. At this pH most of the o-nitrophenol is converted to the yellow colored anionic form and β-galactosidase is inactivated.

β-galactosidase activity was monitored as described by Sambrook and coworkers [66] with small modifications. Briefly, 10 ml of pre - culture were inoculated with a distinct yeast colony grown in solid agar and the cultures were grown over night at 30° C and 180 rpm. The cultures were then inoculated with different stressors into fresh 50 mL falcon tubes (3 replicate for each stressor plus positive control). Yeast cells expressing the plasmids were selected on minimal medium (lacking the appropriate amino acid) and inoculated (starting OD_{600} ~ 0.02) with stressor. Cells were then collected when the cultures reached OD_{600} ~ 1.5, chilled on ice and span down at 4000 rpm for 5 minutes. Then the cells were washed twice with 1 ml of PBS and finally resuspended in 250 µL of PBS and frozen immediately at -80°C and assayed later. Cell suspensions were thawed on ice and resuspended with 300 µl PBS, 1/3 volume of glass beads, 1mM PMSF and tablets of protease inhibitors (ROCHE) were added and vortexed 1 minute for 5 times using Precellys® 24 bead-beating homogenizer and chilled on ice for 1 minute between bursts. The supernatant was then removed carefully avoiding foam and transferred to a new previously refrigerated Eppendorf tube. The protein extract was clarified by centrifugation for 15 minutes at 13000 rpm at 4° C. A 2.5 µl extract was transferred to another microtube and adjusted the total volume to 0.5 ml with Z buffer (Appendix B). β -mercaptoethanol was added just before use 270 µl/100 ml. For thermal inactivation the mixture was incubated at 54° C for different test times (0, 2, 4, 6, 10 and 15 min) and then left on ice for 30 minute. The reaction tube was incubated at 37° C for 5 min and the reaction initiated by adding 0.1 mL of ONPG stock solution. The time lapse between two tubes was 10 seconds. Each reaction was then allowed to proceed for 5 minute. Finally the reaction was terminated by adding 0.25 ml of Na_2CO_3 stock solution (10 sec time lapse between tubes again). The optical density was measured at 420 nm using a Microplate Absorbance Reader (iMark, BIO-RAD). A 200 µL of reaction sample was loaded in 96 well plates and Z buffer was used as blank.

The protein concentration was measured using Pierce® BCA Protein Assay Kit (Thermo scientific) following Microplate procedure. The protein mixture, whenever necessary, was diluted in 1X PBS at 1: 5 ratios and the absorbance was measured at 575 nm

on a plate reader. The activity of the β-galactosidase was expressed according to the following formula:

$$\frac{\mathbf{OD_{420}\ X\ 0.85}}{\mathbf{(0.0045\ X\ Protein\ X\ Extract\ volume\ X\ Time)}}$$

OD_{420} is the value of optical density of o-nitrophenol measured at 420 nm. The factor 0.85 corrects for the reaction volume. The factor 0.0045 is the optical density of a 1 nmole/ml solution of o-nitrophenol. Protein concentration is expressed as mg/ml. Extract volume is the volume assayed in milliliters. Time is in minutes. Specific activity was expressed as nmoles/minute/mg protein.

3.4 Protein extraction, purification and analysis

Affinity chromatography is commonly used to purify a protein of interest from total protein preparations. For this, affinity tags are cloned into recombinant proteins for affinity purification and to improve protein solubility. A large number of tags have been developed for protein production. In the present study a GST tag was fused with the lacZ gene for expression in yeast strain CEN.PK2 (plasmid pGL-C1). This allowed for purification of β-galactosidase protein with glutathione sepharose beads. To exclude putative influences of the GST tag on the production and solubility of the recombinant β-gal, the strain BY4742 carrying the plasmid pUKC815 expressing the lacZ gene alone was also tested. Two different approaches were used for the purification of the β-galactosidase, namely immunoprecipitation and direct gel extraction of the bands corresponding to β-galactosidase.

3.4.1 Protein extraction

For the extraction of protein (soluble and insoluble) from *S. cerevisiae* CEN.PK2 carrying pGL-C1 plasmids, frozen cells were defrosted on ice and 5 mL of 1X PBS (recipe Appendix B) was added and the pellets re-suspended by vortexing. Then 1mM PMSF and a Roche Protease Inhibitor cocktail were added. Around ~ 1/3 volume of glass beads (0.1mm) were added to the mixture and cells were lysed by vortexing at high speed, 8 times for 30 seconds, with cooling on ice for 1 minute between each vortexing step. Then the cells were centrifuged at 3000 rpm for 5 minutes at 4° C. Supernatants (whole cell lysates) were

transferred to a new falcon tube and further fractioned by centrifugation at 11000 rpm for 20 min at 4° C. The supernatants were collected and filtered with a 0.45 µm filter into a new 15 mL falcon tube ("total soluble" protein) and a small fraction (100 µL) was saved for SDS-PAGE and western blot analysis. The pellets (insoluble under native condition) were re-suspended in 5 mL denaturing buffer (1X PBS with 8M urea) and centrifuged at 11000 rpm for 20 min at 4° C. The supernatants were collected and filtered with a 0.45 µm filter to a new 50 mL falcon tube ("total insoluble" protein) and a small fraction (100 µL) was saved for SDS-PAGE and western blot analysis.

3.4.2 Protein purification by affinity chromatography

Efficient transcription of a gene and translation of the expressed mRNA into protein is important when working with proteins. Once high yield of the protein has been obtained, it has to be purified from all the other components of the cell, including all the other proteins. Protein purification is a complex process requiring a high degree of skill and experience, and most protocols rely on finding particular properties of the protein in question, which can be used to gradually enrich the protein away from all the other proteins and components of the cell. For example, proteins can be separated by size exclusion on columns containing gels of the right sort, or by charge by binding them to a column and then gradually washing them off in various solutions with increasing salt concentrations. Often, several of these procedures have to be done one after the other until the protein obtained is highly pure. More recently, a new range of techniques have come into general use which can speed up the process by cutting down the number of steps necessary, and which are broadly applicable to a range of proteins with very different properties. All these techniques involve adding extra amino acids to the protein that give it a new properties, namely to bind very tightly to a particular substrate. This specific binding can be used to isolate the protein away from all the other materials in the cell which do not bind this substance. The method is known as affinity chromatography. A common example is the use glutathione S- transferase (GST) tags. GST is an abundantly expressed 26 kDa eukaryotic protein, and GST cloned from *Schistosoma japonicum* promotes solubility and permits purification of N-terminal fused proteins using glutathione sepharose [77](Figure 9). When applied to the affinity medium, fusion proteins bind to the ligand and impurities are removed by washing with binding buffer. Fusion proteins are then eluted from the Glutathione Sepharose under mild, non denaturing

conditions that preserve both protein antigenicity and function. In the present study the GST-β galactosidase protein was purified using Glutathione Sepharose 4B (GE Healthcare, Bio-Sciences AB, Sweden) medium. After the fusion protein is bound to the resin, it can be eluted under rather mild conditions using free reduced glutathione (between 10–40 mM) at neutral pH.

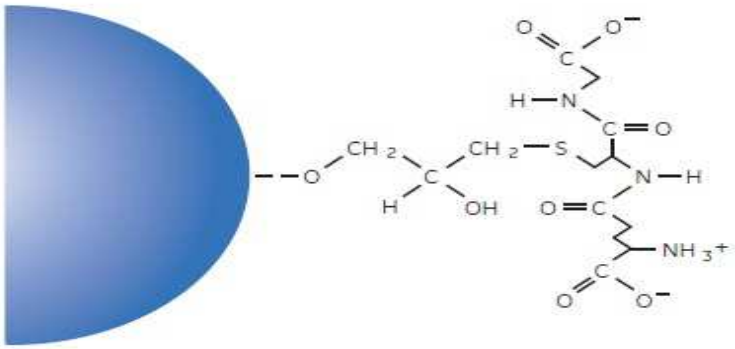

Figure 10: Terminal structure of Glutathione Sepharose. Glutathione is attached to Sepharose by coupling to the oxirane group using epoxy-activation. The structure of glutathione is complementary to the glutathione S -transferase binding site. (Source: GST Gene Fusion System Handbook, GE healthcare)

The Glutathione Sepharose medium was prepared according to protocol (see details in Appendix A). About 400 µL of prepared Sepharose 4B resin (50% slurry) was incubated overnight with total soluble protein with gentle slurry mixing. The day after a Poly-Prep® Chromatography Column (Bio-rad) was washed with one column volume of milli-Q water and then equilibrated with one column volume of binding buffer (1X PBS). Then, the proteins were mixed overnight with Sepharose beads and poured into the column and the flow through was collected in a 15 mL falcon tube. The column was then washed with one column volume of 1X PBS (washing buffer) and the flow through was collected in another 15 mL falcon tube. The bound protein was then eluted with 5 ml of elution buffer (see appendix B for recipe). The eluted protein fraction was collected into 5 Eppendorf tubes (1.5 ml each fraction) in order to identify the fraction which contained the higher amount of purified protein. The protein was then ready for SDS-PAGE and western blotting analysis.

3.4.3 Protein purification with Immunoprecipitation technique

Another important technique for separation of protein is Immunoprecipitation (IP), which precipitates a protein antigen out of solution using an antibody that specifically binds to that particular protein. This process can be used to isolate and concentrate a particular protein from a sample containing many thousands of different proteins. Immunoprecipitation requires that the antibody couples to a solid substrate at some point in the procedure. In the present study, β-galactosidase protein produced in the yeast strain BY4742 was purified using Dynabeads® (invitrogen), a magnetic bead-based separation technology. The technology is designed for immunoprecipitation of proteins, protein complexes, protein-nucleic acid complexes and other antigens. Anti β-galactosidase -IgG fraction A-11132 raised in rabbit, (Molecular probes) was added to Dynabeads® Protein G. During a short incubation, the Ab binds to the Dynabeads® via their Fc- (fragment crystallizable) region. The tube was then placed on a magnet, where the beads migrate to the side of the tube facing the magnet and allow for easy removal of the supernatant. The bead-bound Ab may now be used for immunoprecipitation. Bound material was easily collected utilizing the unique magnetic properties of the Dynabeads®. Magnetic separation facilitates washing, buffer changes and elution. The detail protocol of immunoprecipitation of β-galactosidase protein is reported in Appendix A.

3.4.4 SDS-PAGE analysis

The purpose of this method is to separate proteins according to their size, and no other physical features. In order to get the protein in a linear form secondary, tertiary or quaternary structures are eliminated using a detergent and reducing agents.

SDS (sodium dodecyl sulfate) is a detergent that can dissolve hydrophobic molecules, but also has a negative charge (sulfate) attached to it. Therefore, if a protein is incubated with SDS, it will be linearize by the detergent, moreover the proteins will be covered with negative charges. So a protein will migrate towards the positive pole when placed in an electric field. In addition the proteins should be put into an environment that will allow different sized proteins to move at different rates. The environment of choice is polyacrylamide, which is a mixture of two polymers, acrylamide and bisacrylamide. The first is a linear molecule whereas the second has a ‘T’ shape, and when mixed they form a matrix with different separation gradient, depending on the concentration of each polymer. When

this polymer is formed, it turns into a gel and electrical current pulls the proteins through the gel in a process called Polyacrylamide gel electrophoresis (PAGE). A polyacrylamide gel is not solid but is made of a labyrinth of tunnels through a meshwork of fibers.

Each gel is composed by a resolving gel with a percentage of acrylamide adapted to the size of the target protein. The smaller the molecular weight of the protein(s) of interest, the higher is the percentage of acrylamide to be used. To improve the resolution, proteins first pass through a stacking gel polymerized on the top of the resolving gel. The stacking gel has lower pH and acrylamide concentration (4%) as well as different ionic strength, which allows the proteins to be concentrated during the first minutes (~30 min) of electrophoresis, before entering the resolving portion of the gel. Since the GST- β galactosidase protein has a MW of 142 kDa and the β galactosidase alone has a MW of 116 kDa, 10% resolving gels were used. During gel preparation Tris-HCl was used as buffer to absorbs counter ions (H^+ and OH^-) keeping the solution in a stable pH level. For gel polymerization, APS (ammonium persulfate at 10 mg/100 ml), a source of free radicals was used as initiator of gel formation and TEMED (N, N, N', N'-tetramethylethylene diamine), a stabilizer of free radicals, was used to improve polymerization. The gel was prepared according to the recipe of Appendix B and Bio-Rad® SDS-PAGE gel apparatus were used. The Protein samples for SDS-PAGE were prepared by adding 6X loading buffer (appendix B) in a ratio of 1:6, followed by denaturing the samples for seven minute at 95º C. Around 15 µL of protein preparation were loaded into the gel pockets along with a pre-stained protein marker and the gels were run with 1X SDS running buffer (see recipe appendix B) at 80 V for the first 30 minutes until the samples reached the stacking gel. Then, the voltage was increased to 130 V and the gels were run for 2 hours until the dye front reached the bottom of the gel. The gel was then carefully removed and stained with Coomassie Brilliant Blue (see recipe appendix B) dye for 1-2 hours and destained overnight with destain solution (see recipe appendix B)

3.4.5 Western blot analysis

Western blotting is a widely used technique for the detection and analysis of proteins. It was first described in 1979 [46] and has since become one of the most commonly used methods in life science. It uses gel electrophoresis (SDS-PAGE) to separate native proteins or denatured proteins, which are then transferred to a membrane (typically nitrocellulose or PVDF), where they are probed (detected) using antibodies specific to the target protein. In the present study, after electrophoresis, proteins were electro blotted onto nitrocellulose membranes (Amersham Hybond ECL) prior to immunodetection. Soluble and insoluble protein fractions were analyzed under reducing conditions using 10% SDS-PAGE, as described above, and blotted onto a nitrocellulose membrane. For this, six sheets of 3MM paper (Whatman) and blotting membranes were cut to gel dimensions. Membranes were prehydrated in distilled water and 3 sheets of 3 MM paper and hydrated in TGM buffer (see Appendix B). Then papers were placed above pads on the anode of the transfer system, and a "sandwich" was assembled by laying down the gel on top of the paper sheets. The blot was run for 100 min at 100V at 4° C in TGM buffer (see recipe appendix B) using a BIO-RAD® wet transferring system (assembled according to manufacturer's instructions). The membranes were removed carefully and washed 2 times with 1X TBS and blocked in 1X TBS (Appendix B) with 3% non-fat milk for 30 minutes at room temperature with gentle agitation. After removing blocking agent, the membranes were washed two times for 5 minutes with 1X TBS. Following the blocking and washing, the primary antibody (anti β-galactosidase IgG fraction A-11132 raised in rabbit, Molecular probes) was added (1:5000) to the membranes in a solution of 3% non-fat dry milk in 1X TBS. For this, the membranes were placed in heat sealable plastic bags, and each membrane was incubated overnight at 4º C with agitation. After the overnight incubation, the membranes were washed twice for 5 minutes with 1X TBS to remove unbound primary antibody.

Then, the membranes were incubated with the secondary antibody, a mouse anti-rabbit antibody reactive against the primary antibody and coupled to a fluorochrome that allows subsequent visualization. A 1:10,000 dilution of mouse anti-rabbit antibody was prepared in 1X TBS and 3% non-fat dry milk. Again, incubations were performed in a plastic tray, which were then wrapped in aluminium foil (to avoid fluorochrome degradation by light) at room temperature for 2 hours with gentle agitation. Following, three washes were performed with 1X TBS-Tween (recipe Appendix B), for 15 minutes each time. The

membranes were then washed with distilled water and analyzed at 700 nm (anti rabbit) using an Odyssey Li-COR fluorescence imager (Bioscience).

3.4.6 Sample preparation for mass spectrometry analysis

The β-galactosidase protein was purified according to a protocol mentioned before and the sample was concentrated. The quality of the sample concentration is a critical step in protein biochemistry. In the present study, eluted GST-β galactosidase purified by glutathione sepharose were concentrated using an Amicon Ultra-15 mL filters (Millipore) of 3,000 molecular weight cut-off (MWCO) ultra filtration membranes and sent for mass spectrometry analysis in the form of in-solution. For in-solution digestion the protein sample were mixed with an aqueous buffer (NH_4HCO_3) containing trypsin (5%).

β-galactosidase expressed in yeast strain BY4742 was separated by SDS-PAGE, in gel digested and analysed directly by mass spectrometry. Protein bands were excised from stained polyacrylamide gels. Each gel band was cut into pieces (~1 mm^2) using a scalpel, and placed into a 1.5 mL microcentrifuge tubes containing milli-Q water.

3.5 Mass spectrometry analysis

Mass spectrometry is a high-throughput technology for protein identification and quantification with high-accuracy [78]. The misincorporation of specific amino acids can be identified by peptide mapping and LC-MS/MS analysis on proteolyzed peptides of the protein. The peptide is fragmented by collision-induced dissociation during MS/MS sequencing analysis, which breaks backbone amide bonds generating a series of fragment ions, mainly C-terminal y ions and N-terminal b ions in this case [79]. The identity of an amino acid at each position in a peptide can be determined by the difference in the m/z values of two adjacent y or b ions. For example, a difference of 71 atomic mass units between adjacent y ions (or b ions) indicates an Alanine at that position, and a difference of 57 atomic mass units indicates a Glycine.

Another MS technique known as Matrix Assisted Laser Desorption Ionization (MALDI) –Time of Flight (ToF) has become a major ionization method for analysis of large biological samples and synthetic polymers [91]. Various matrices are used in this technique that assists in desorption and ionization of the analytes having molecular weight of up to 500 kDa. The basic concept of MALDI-TOF mass spectrometer is that the ions are separated based on the time it takes for the ion to drift down the flight tube to the detector. Lighter ions have higher velocities than heavier ions and reach the detector first. Peptides with a high mass need longer flight time in the tube before they reach the detector at the end of flight tube. Flight times of ions are inversely proportional to the square root of their molecular mass. The m/z ratio of a peptide is calculated based on the energy equation $E=½\ mv^2$ (where E = kinetic energy of the ion, m = mass, and v = velocity) [91]. Therefore, the peptides are sorted due to the size of the individual peptides.

4. Results

4.1 Transformations of β-galactosidase gene into *Saccharomyces cerevisiae*

Following transformation of pGL-C1 and pUKC815 plasmid into the yeast strains CEN.PK2 and BY4742 the presence of the lacZ gene in the plasmid was confirmed by amplification of the gene by polymerase chain reaction. The primers successfully amplified the GST-lacZ gene (1254 aa) mentioned in appendix D. In this sequence AA position from 5-218 (214 AA) matched with the GST sequence (UniProt identifier accession no. P08515 and name GST26_SCHJA) isolated from *Schistosoma japonicum.* The lacZ sequence matched with the sequence of the *E. coli* lacZ gene (UniProt identifier accession no. Q8VNN2 and entry name BGAL_ECOLX) from AA position 240 to 1254 (1015AA). Moreover the Beta-gal assay confirmed the presence of β-galactosidase protein in the yeast cell lysate.

4.1.1 Effect of different stressors on the growth of *Saccharomyces cerevisiae*

The effect of different stressors on the growth of *S. cerevisiae* strains CEN.PK2 and BY4742, transformed with plasmid pGL-C1 and pUKC815 respectively, was monitored by measuring the optical density at different time points at wavelength 600 nm. After several trials the right concentration of stressors were found (data not shown). Different stressors were applied with varying concentration such as ethyl alcohol 2% and 5% (v/v), chromium (CrO_3) with 1 mM, 0.5 mM, 0.25 mM and 0.1 mM concentration and Geneticin (G418) at 0.2 mg/L. The strain CEN.PK2 did not show any marked difference in growth with 2% (v/v) ethanol as compared to the control without stressor. While the chromium concentrations of 1 mM, 0.5 mM and 0.25 mM were lethal.

The presence of ethanol at 5% (v/v), chromium 0.1 mM and Geneticin 0.2 mg/L strongly reduced the growth rate of the CEN.PK2 strain relative to the untreated control cells (figure 10 A and B). The Geneticin had the strongest effect on the growth rate (OD_{600} 0.018 h^{-1}) compared to other treatments and, at that concentration, appeared to be a limiting factor for the growth rate. Similarly, the cells treated with chromium were significantly affected.

Even ethanol appeared to affect the growth rate significantly (OD_{600} 0.192 h^{-1} vs. 0.299 h^{-1}) but after an initial phase the cells were able to reach the same OD values of the untreated cells.

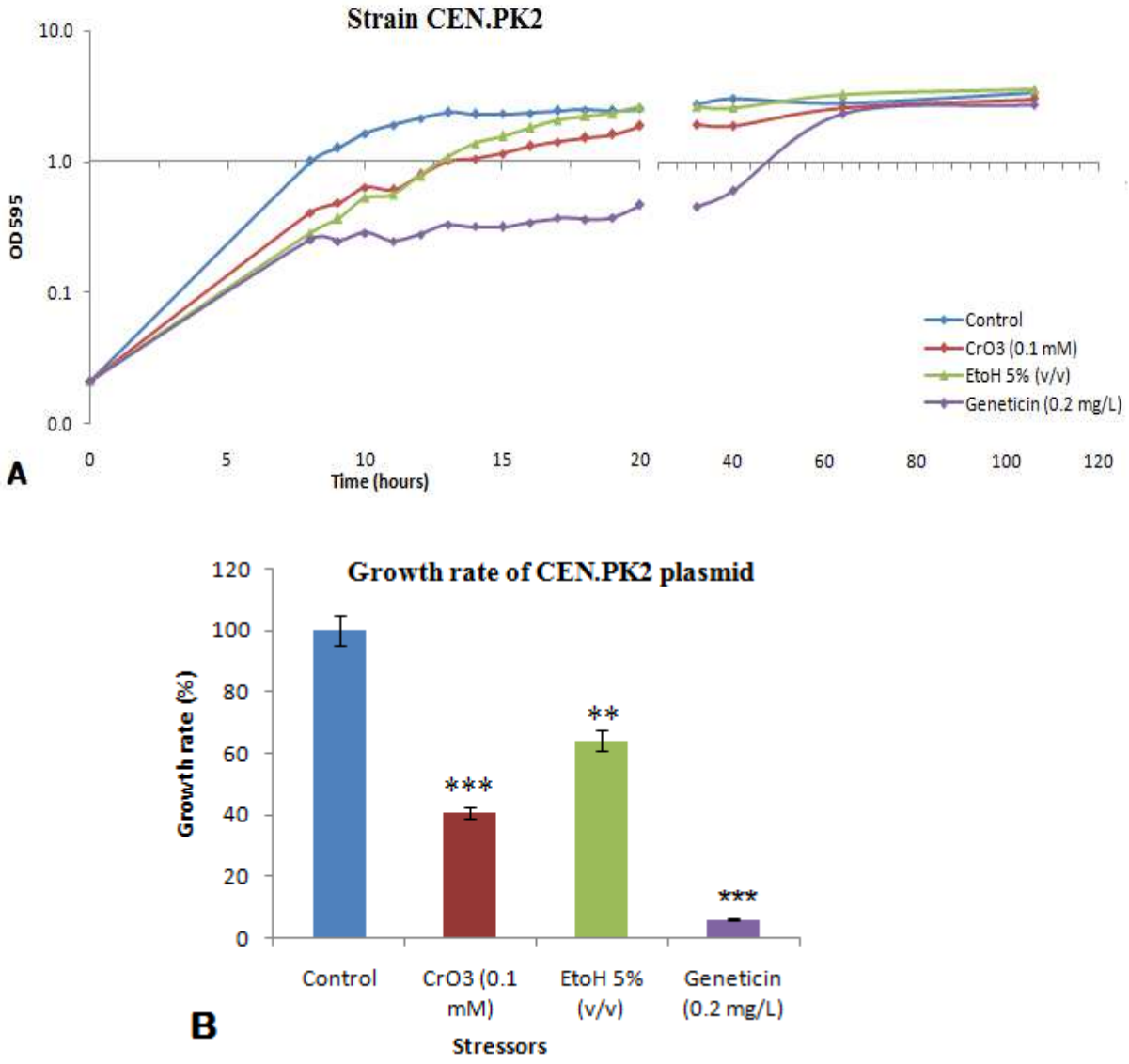

Figure 11: Growth curves (A) and rate (B) of *Saccharomyces cerevisiae* strain CEN.PK2 carrying pGL-C1 plasmid (student t-test: *** P > 0.05, ** P < 0.01)

The yeast strain BY4742 carrying the plasmid pUKC815 (figure 12) treated with identical concentration of stressors as the strain CEN.PK2 carrying pGL-C1 plasmid showed different growth sensitivity. Nevertheless the 3 stressors clearly affected the growth rate of this strain, chromium affected strongly (OD_{600} 0.056 h^{-1} vs. 0.303 h^{-1} in control cells) growth rate of this strain. Other treatments produced a slow growth phenotype (OD_{600} 0.182 h^{-1} in

ethanol vs. 0.222 h^{-1} in geneticin treated cells) in the exponential phase, but cells were able to reach the stationary phase like the positive control.

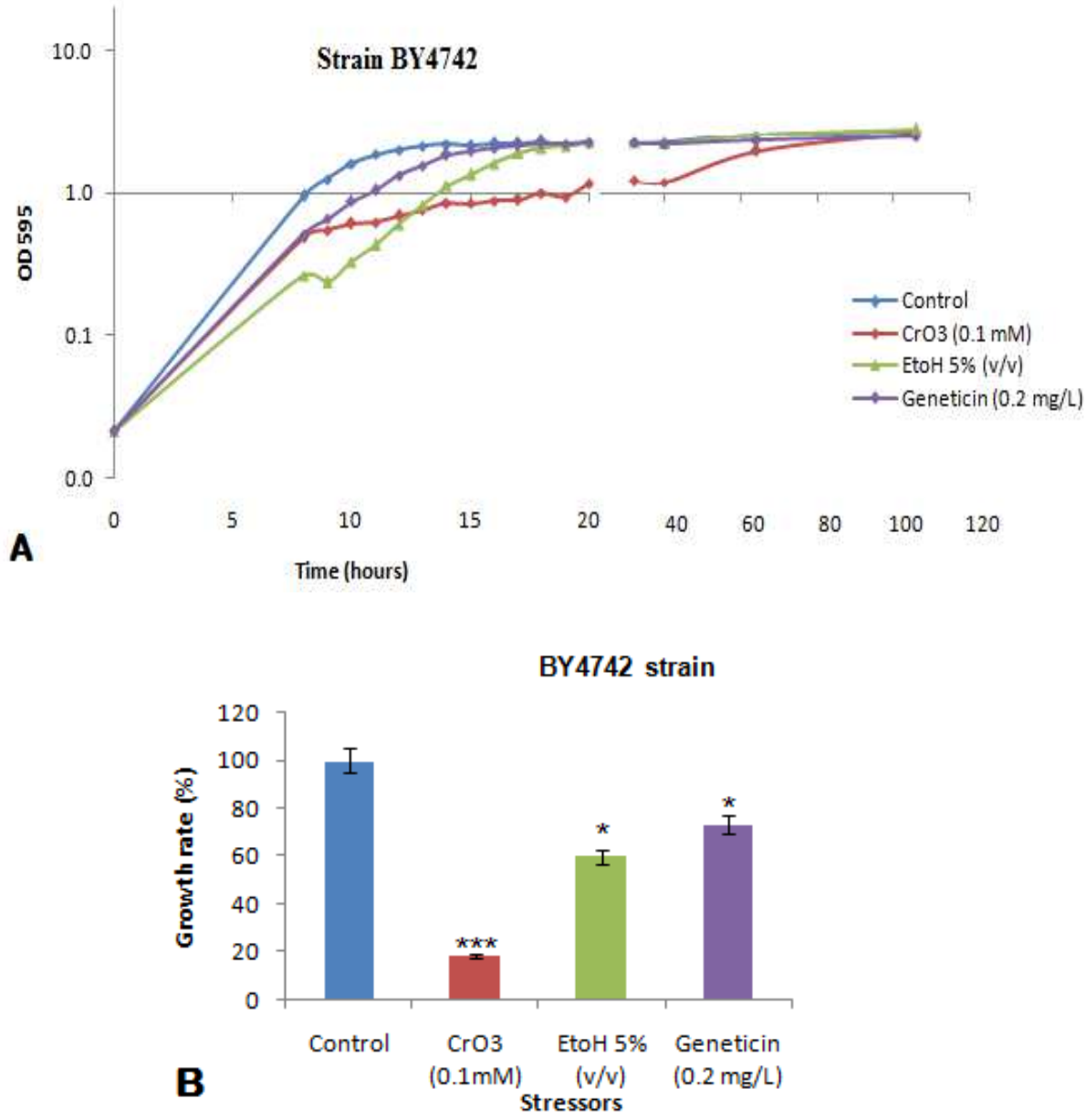

Figure 12: Growth curve (A) and rate (B) of *Saccharomyces cerevisiae* strain BY4742 carrying pUKC815 plasmid (student t-test: *** $P < 0.05$, * $P < 0.01$)

4.2 β-galactosidase assay

The β-galactosidase (β-gal) reporter system was used to test the effects of the different stressors on the fidelity of protein synthesis. This test is based on the fact that if misincorporation of amino acids occurs in the β-gal protein then its thermostability will decrease. In the present study we observed a decrease in activity of the β-galactosidase over heat inactivation time, the strongest decrease of enzymatic activity was observed after 6 minutes of heat inactivation. However, relative differences in thermal stability in the β-galactosidase protein were observed in presence of different stressors. Geneticin affected the most β-galactosidase thermal stability. The effect of chromium (0.1mM) and ethanol (5%) was similar and not distinguishable from untreated control cells (Figure 13A). Similarly, in yeast strain BY4742 the effect of geneticin was stronger than that of the other stressors (Figure 13B).

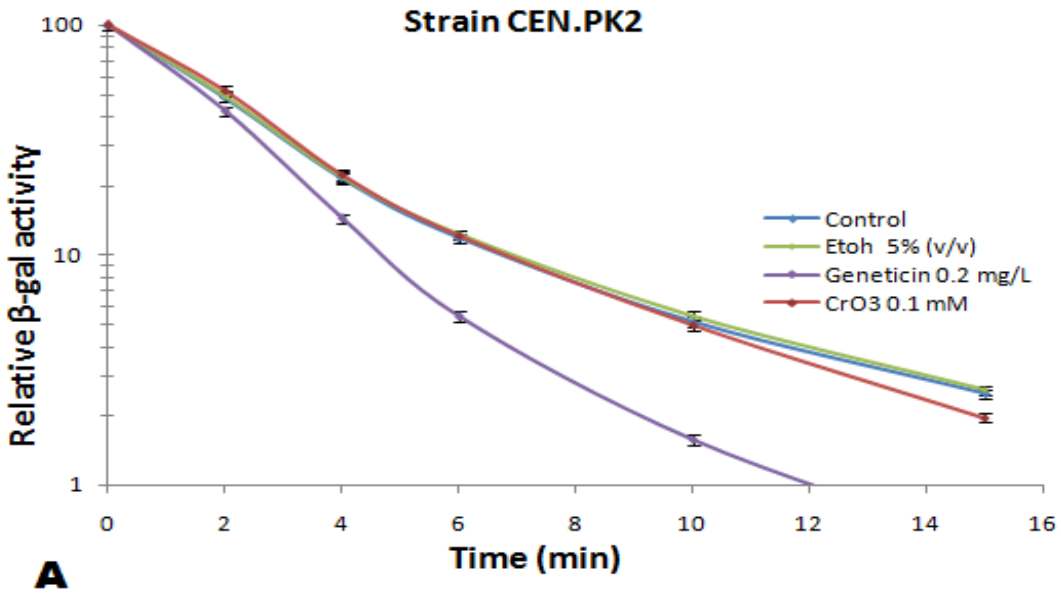

Figure 13 (A): β-galactosidase activity of yeast strain CEN.PK2 carrying pGL-C1 plasmid. β-galactosidase thermal inactivation profiles were determined by measuring β-galactosidase activity directly in *S. cerevisiae* transformed with pGL-Cl plasmid according to protocol described in materials and methods.

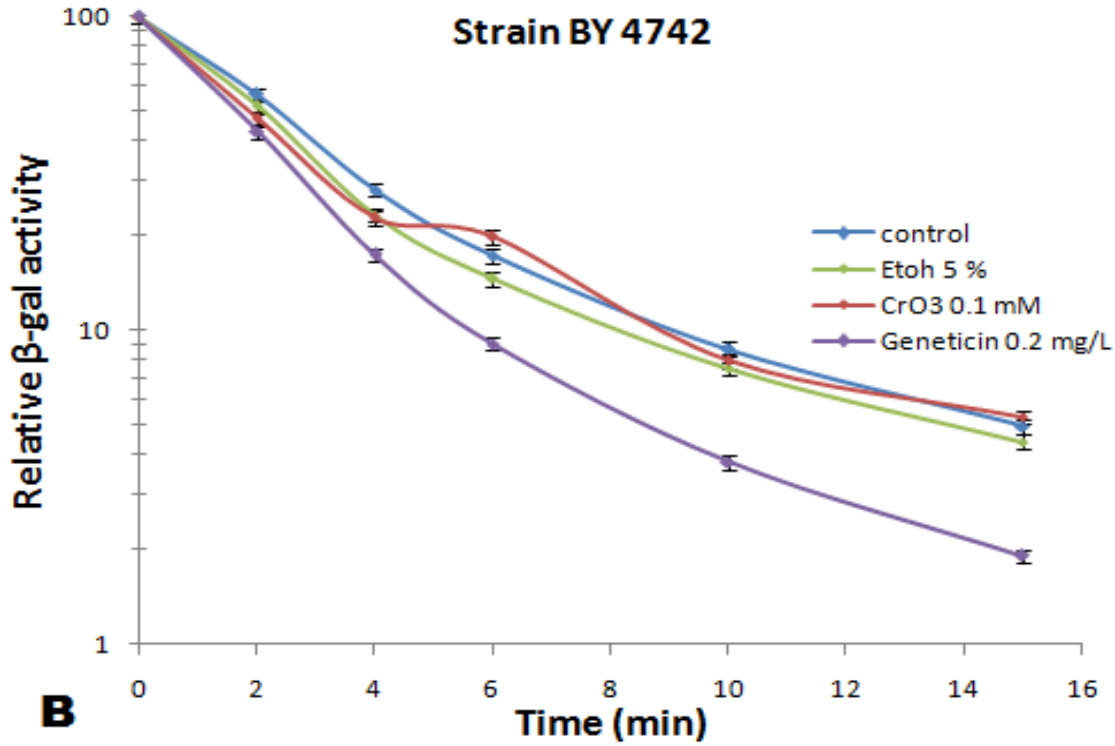

Figure 13(B): β -galactosidase activity of yeast strain BY4742 carrying plasmid pUKC815. β-galactosidase thermal inactivation profiles were determined by measuring β-galactosidase activity directly in *S. cerevisiae* transformed with pUKC815 plasmid according to protocol described in materials and methods.

4.2.1 Western Blot of β-galactosidase

The β-galactosidase abundance was also quantified by western blot. After SDS-PAGE (figure 14A) analysis, intense bands in soluble protein were observed in the three different treatments as well as in the untreated cells. The intensity of bands in insoluble fraction was much weaker.

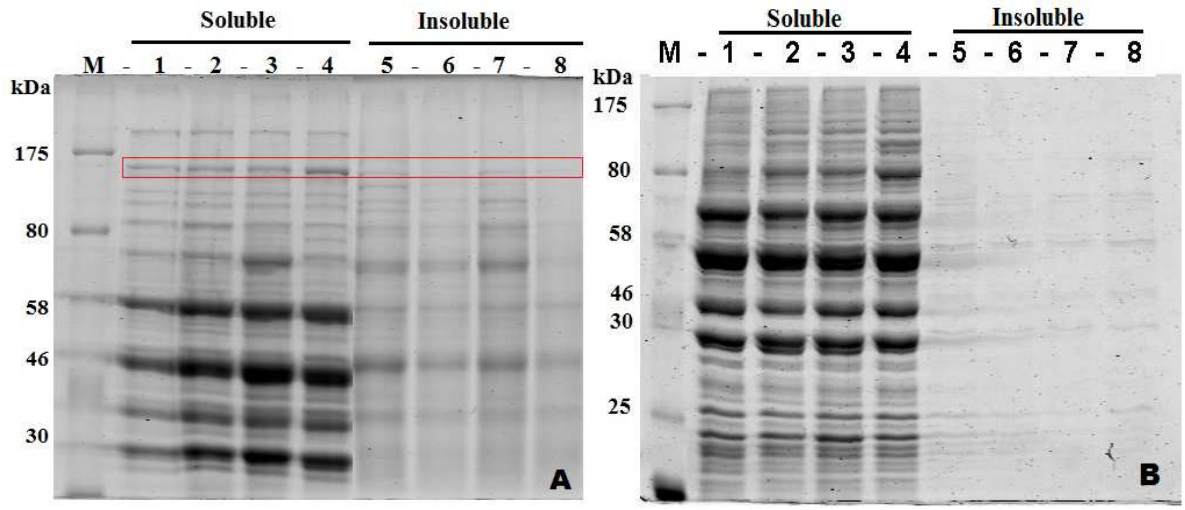

Figure 14: SDS-PAGE analysis showing of whole protein extract from yeast strain CEN.PK2 carrying plasmid pGL-C1 and BY4742 strain carrying plasmid pUKC815. **[A]** CEN.PK2: lane M- marker, lane 1, 2, 3 and 4 indicates soluble protein extract from untreated control, Ethanol, CrO_3 and Geneticin respectively. Lane 5, 6, 7, and 8 indicates insoluble protein extract from control, Ethanol, CrO_3 and Geneticin respectively. The red rectangle indicates the position of GST-β galactosidase protein confirmed by western blotting **[B]** BY4742: lane M- Protein marker, lane 1, 2, 3 and 4 indicates soluble protein extract from untreated control, Ethanol, CrO_3 and Geneticin respectively. Lane 5, 6, 7, and 8 indicates insoluble protein extract from control, Ethanol, CrO_3 and Geneticin respectively.

The presence of GST- β galactosidase protein was confirmed by western blotting and immuno detection (figure 15) with a rabbit monoclonal antibody raised against *E. coli* β-galactosidase. The protein was detected both in the soluble and insoluble fractions (figures 14, 15), the intensity of bands in insoluble fraction was very weak. The β-galactosidase protein was also detected in the yeast BY4742 carrying plasmid pUKC815 (figure 14B), which was also confirmed by western blotting (figure 15B). After western blotting the bands were only observed in the soluble fraction and were not detected in the insoluble whole

extract. The predicted MW of mistranslated GST- β galactosidase protein (Appendix D) was also calculated using the BioEdit (version 7.0.9) sequence alignment software, which predicted that expected mistranslated protein varies from 119 Daltons (Da) to maximum 475 Da due to the amino acid misincorporations. This difference is not sufficient to produce distinct bands in SDS-PAGE, confirming our observation of a single band in the western blot.

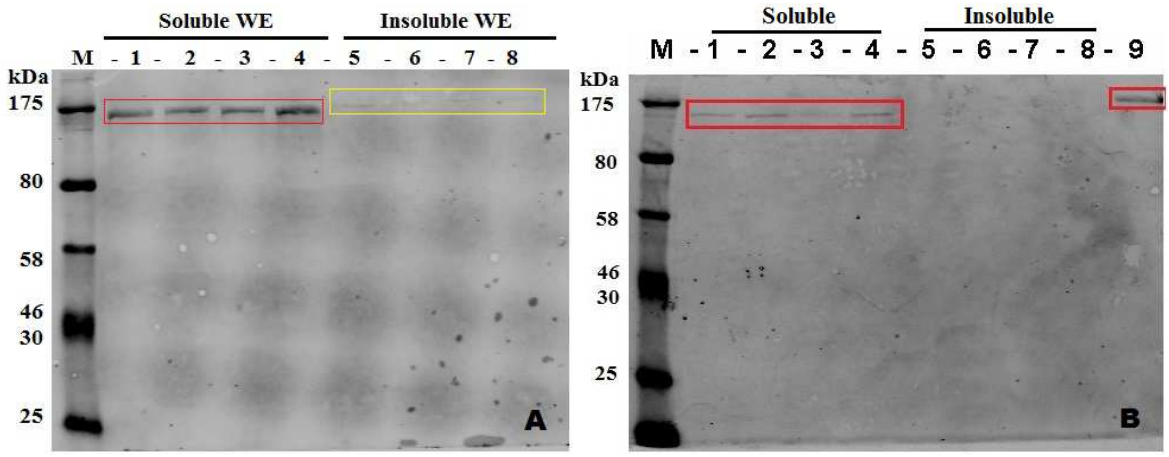

Figure 15: **[A]** Western blotting (WB) showing GST-β gal protein detected using an anti- β galactosidase antibody. Whole protein extract of yeast strain CEN.PK2 grows in presence of different stressors, where Lane M- Protein marker, Lane 1 - control, 2 - Ethanol, 3 - CrO_3 and 4 - Geneticin stressed soluble protein extract and from Lane 5 - control, 6 - Ethanol, 7- CrO_3, and 8- Geneticin stressed insoluble protein extract. Red and yellow rectangle shows the GST- β-galactosidase bands in soluble and insoluble extract. **[B]** WB shows the β galactosidase present in the whole protein extract prepared from yeast strain BY4742. Lane M- Protein marker, Lane 1 - control, 2 - Ethanol, 3 - CrO_3 and 4 - Geneticin stressed soluble protein extract and from Lane 5 - control, 6 - Ethanol, 7- CrO_3, and 8- Geneticin stressed insoluble protein extract. Lane 9 indicates control whole soluble protein extract from CEN.PK2 strain loaded as a reference. Red rectangles indicate the presence of β-galactosidase protein.

4.3 Purification of β-galactosidase protein from *Saccharomyces cerevisiae*

The β-galactosidase protein produced from lacZ gene has a molecular weight of ~116 kDa and the fusion with GST increase the molecular weight to ~ 142 kDa. The GST binds to resin immobilized glutathione, and this property was successfully used for affinity purification of GST tagged proteins. The fusion protein bound to the glutathione - sepharose beads, was eluted under mild conditions using free reduced glutathione (10 mM) at neutral pH. Most of the GST-β galactosidase protein was soluble (Figure 16A), the amount of protein extracted from the inclusion bodies was very little and could not be analyzed by Mass spectrometry. Different protocols have been applied for the solubilization of the inclusion bodies using high concentration of denaturant, such as urea or reducing agents such as Dithiothreitol (DTT) or detergent as Triton-X. Unfortunately none of these protocols produced the expected results.

The other yeast strain BY4742 used in this study, carrying the low copy plasmid pUKC815, which expresses non tagged β-galactosidase protein was purified by immunoprecipitation with Dynabeads® (invitrogen). Unfortunately, after several trials, the immunoprecipitation of β-galactosidase was not successful (Figure 16B) and the protocol was not used for further studies. Purification of insoluble protein expressed from both constructs as well as the soluble protein without the GST tag was also done by cutting bands directly from SDS-PAGE gels.

The purified GST-β galactosidase protein from the yeast strain CEN.PK2 was successively filtered and concentrated using Amicon® Ultra centrifugal (15 mL) 3000 MWCO devices (Millipore). A small fraction of the protein was loaded on a SDS-PAGE to check the protein purity. SDS-PAGE analysis (Figure 17) shows the presence of β galactosidase protein in solution. The concentration of GST-β galactosidase protein was quantified using NanoDrop™ spectrophotometer and was recorded 0.57, 0.38, 0.30 and 0.21 mg/ml in control, ethanol, chromium and geneticin treated yeast respectively.

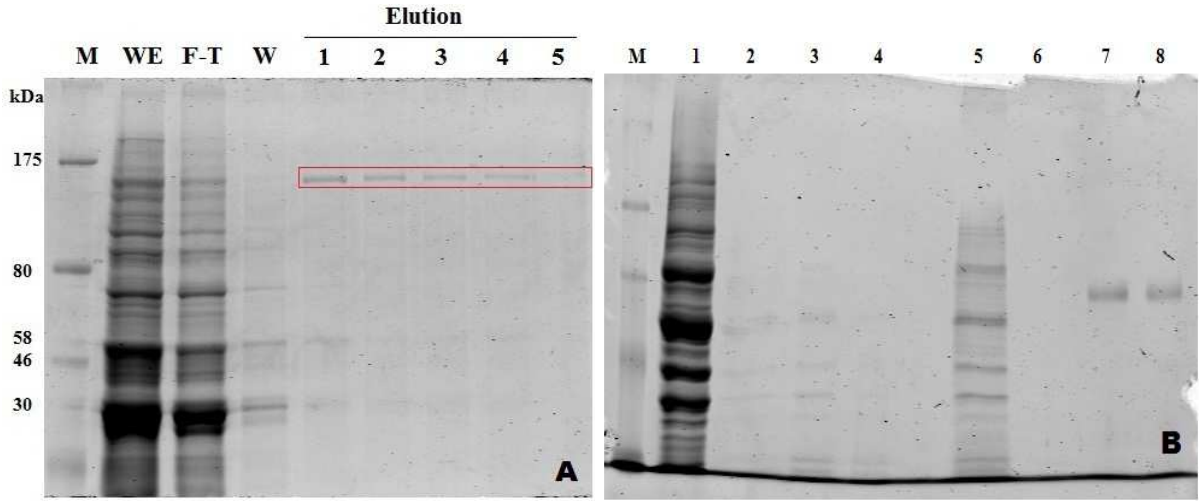

Figure 16: **A)** SDS-PAGE showing purified GST–β galactosidase protein from *Saccharomyces cerevisiae* strain CEN.PK2. Lanes (from left to right) showing : M- protein marker, WE – Whole extract, FT- Flow through, W- Wash with PBS and Elution fraction 1, 2, 3, 4 and 5. The red rectangle shows the band of GST- β galactosidase protein in the soluble eluted fraction. **B)** SDS-PAGE showing immuneprecipitated β-gal protein from the yeast strain BY 4742. Lane M- protein marker, Lane 1, 2, 3, 4 denotes control soluble wash 1, wash 2, elution and elution (the same of lane 3) respectively and Lane 5, 6, 7, 8 denotes control insoluble wash 1, wash 2, elution and elution (the same of lane 7) respectively.

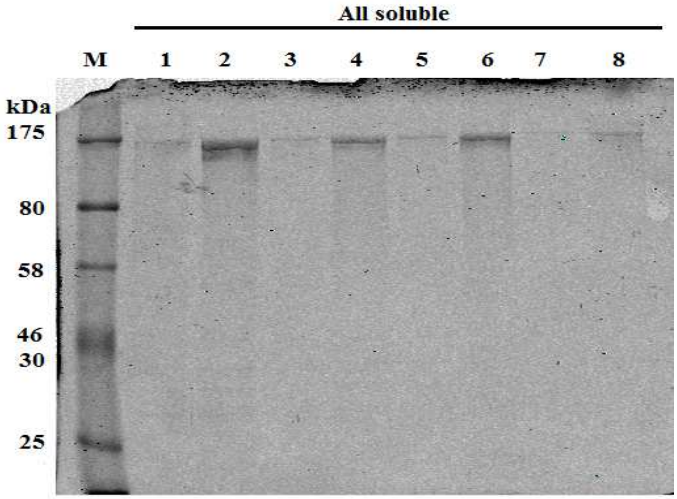

Figure 17: SDS-PAGE showing the concentrated protein centrifuged for 90 minutes (lane 1, 3, 5 and 7) and 150 minutes (lane 2, 4, 6 and 8 respectively). 15 µl of sample was loaded in 10% PAGE gel whereas lane 1& 2 indicate β-galactosidase protein from control without stressor, lane 3 & 4 with stressor 5% (v/v) Ethanol, lane 5 & 6 with stressor 0.1 mM CrO3 , and lane 7 & 8 indicate protein exposed with stressor 0.2 mg/L Geneticin.

4.4 Mass spectrometry analysis

Mass spectrometry data confirmed the amino acids misincorporations in GST-β galactosidase protein in (table 2). Eight out of the twenty amino acids namely Ala, His, Pro, Ser, Asp, Arg, Met and Val were misincorporated in different positions of GST-β galactosidase protein treated with different stressors. However, the misincorporation of amino acid is not random, but depends on position and amino acids specificity. There were two substitutions of amino acids that were position specific, namely Ser → Ala in position 423 (figure 18 B) and Arg → His in position 274 (figure 18A) and 1173 (figure 18C) in β-galactosidase treated with all stressors. The residues positions are mentioned here according to GST- β galactosidase protein (1-1254 aa) sequenced in this study (Appendix C). Alanine may misincorporated at highest level [figure 19B]. The order of other amino acids misincorporations (%) were Ala > His > Ser and Pro >Arg and Asp > Met and Val. Some of the amino acids were substituted with a specific amino acid, for example Arg → His, Leu → Ser, Lys → Arg, Ser → Asp, His → Met and Trp → Val. Alanine was substituted for Trp / Ser / Tyr and Proline was substituted with Cys / Gln / Ser / Trp. These substitutions occurred in specific positions of the GST- β protein.

The *E. coli* lacZ gene aligns with positions 240 to 1254 of the GST- β gal produced in CEN.PK2 strain. Therefore, the amino acids misincorporated in the position 242, 243, 247, 267, 274, 422, 423, 662, 767, 771, 878, 1173, 1193 and 1194 are considered true misincorporations in β gal protein. The rest of the misincorporations in the position 4 and 47 denotes GST protein. Further studies are needed to confirm these results.

Table 2: Amino acids misincorporation in GST –β galactosidase analyzed with MALDI-TOF mass spectrometry.

Treatments	Peptide fragments	AA → Misincorporated AA	Misincorporated AA	Position in GST-β gal sequence
Control (without Stressor)	NKKFELGLEFPNLPYYIDGDVK	Lys → Arg	NKRFELGLEFPNLPYYIDGDVK	47
	DWENPGVTQLNR	Trp → Ala	DAENPGVTQLNR	247
	NSEEARTDRPSQQLR	Arg → His	NSEEAHTDRPSQQLR	274
	WSDGSYLEDQDMWR	Ser → Ala	WADGSYLEDQDMWR	423
	LILCEYAHAMGNSLGGFAK	Cys → Pro	LILPEYAHAMGNSLGGFAK	767
	HSDNELLHWMVAL	Ser → Asp	HDDNELLHWMVAL	878
	CGTRELNYGPHQWR	Arg → His	CGTHELNYGPHQWR	1173
Ethanol 5% (v/v)	AKLPILGYWK	Leu → Ser	AKSPILGYWK	4
	NKKFELGLEFPNLPYYIDGDVK	Lys → Arg	NKRFELGLEFPNLPYYIDGDVK	47
	GSGHHLEWLVPDPVVLQR	Gln → Pro	GSGHHLEWLVPDPVVLPR	243
	DWENPGVTQLNR	Trp → Pro	DPENPGVTQLNR	247
	NSEEARTDRPSQQLR	Arg → His	NSEEAHTDRPSQQLR	274
	WSDGSYLEDQDMWR	Trp → Ala; Ser → Ala	AADGSYLEDQDMWR	422; 423
	LILCEYAHAMGNSLGGFAK	Cys → Asp	LILDEYAHAMGNSLGGFAK	767
	HSDNELLHWMVAL	Ser → Pro	HPDNELLHWMVAL	878
	CGTRELNYGPHQWR	Arg → His	CGTHELNYGPHQWR	1173
Chromium 0.1 mM	AKLPILGYWK	Leu → Ser	AKSPILGYWK	4
	GSGHHLEWLVPDPVVLQR	Leu → Ser	GSGHHLEWLVPDPVVSQR	242
	LAAHPPFASWR	Trp → Val	LAAHPPFASVR	267
	NSEEARTDRPSQQLR	Arg → His	NSEEAHTDRPSQQLR	274
	WSDGSYLEDQDMWR	Trp → Ala; Ser → Ala	AADGSYLEDQDMWR	422; 423
	CGTRELNYGPHQWR	Arg → His	CGTHELNYGPHQWR	1173
	YSQQQLMETSHR	Tyr → Ala; Ser → Ala	AAQQQLMETSHR	1193; 1194
Geneticin 0.2 mg/L	AKLPILGYWK	Leu → Ser	AKSPILGYWK	4
	DWENPGVTQLNR	Trp → Ala	DAENPGVTQLNR	247
	NSEEARTDRPSQQLR	Arg → His	NSEEAHTDRPSQQLR	274
	WSDGSYLEDQDMWR	Trp → Ala; Ser → Ala	AADGSYLEDQDMWR	422; 423
	LTDDPRWLPAMSER	Arg → His	LTDDPHWLPAMSER	662
	LILCEYAHAMGNSLGGFAK	His → Met	LILCEYAMAMGNSLGGFAK	771
	CGTRELNYGPHQWR	Arg → His	CGTHELNYGPHQWR	1173
	YSQQQLMETSHR	Tyr → Ala; Ser → Ala	AAQQQLMETSHR	1193; 1194

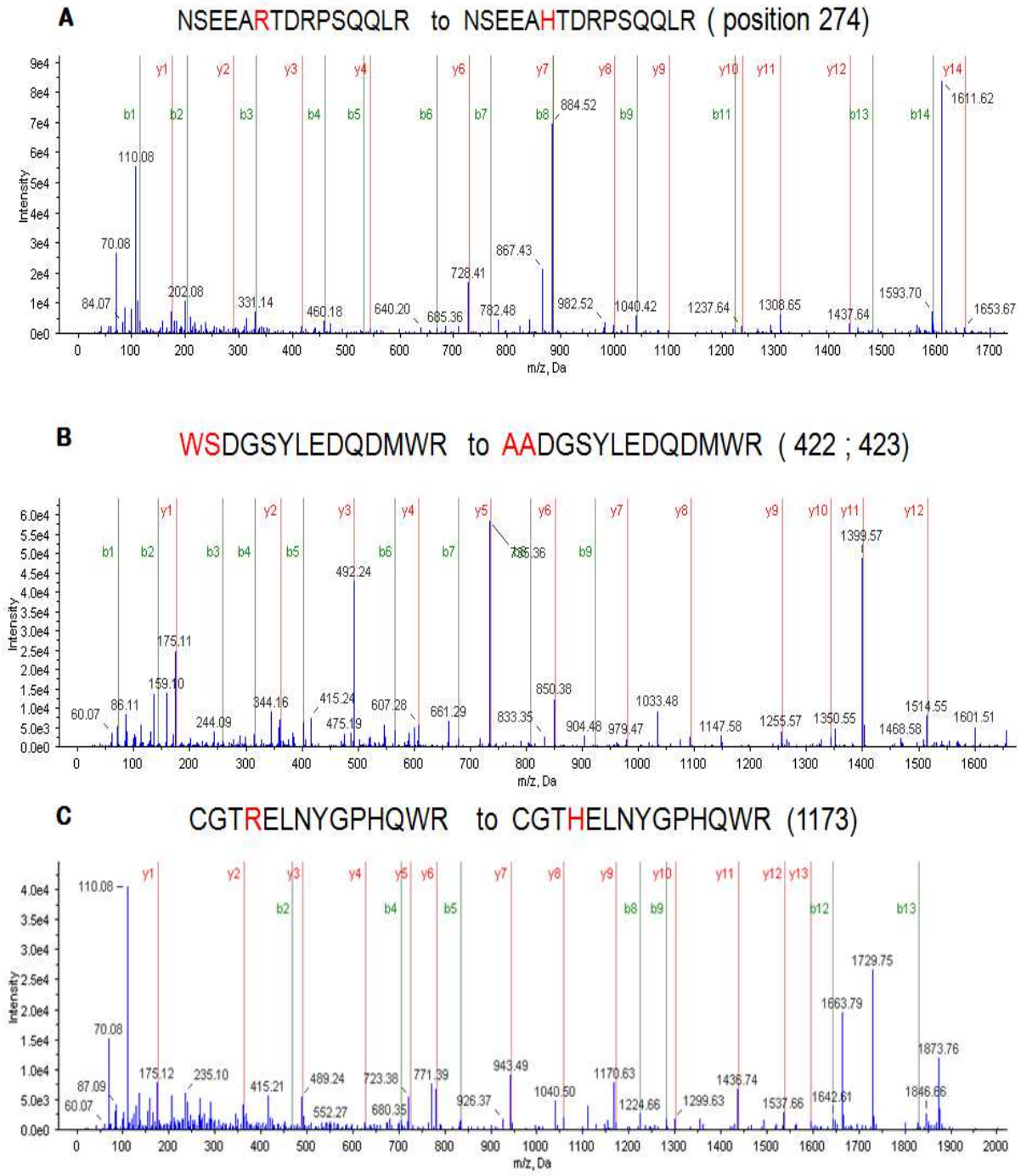

Figure 18: Misincorporation of amino acids in GST-β galactosidase protein (1254 aa) **A)** Histidine at arginine position, **B)** Alanine at tryptophan and serine position and **C)** Histidine at arginine position.

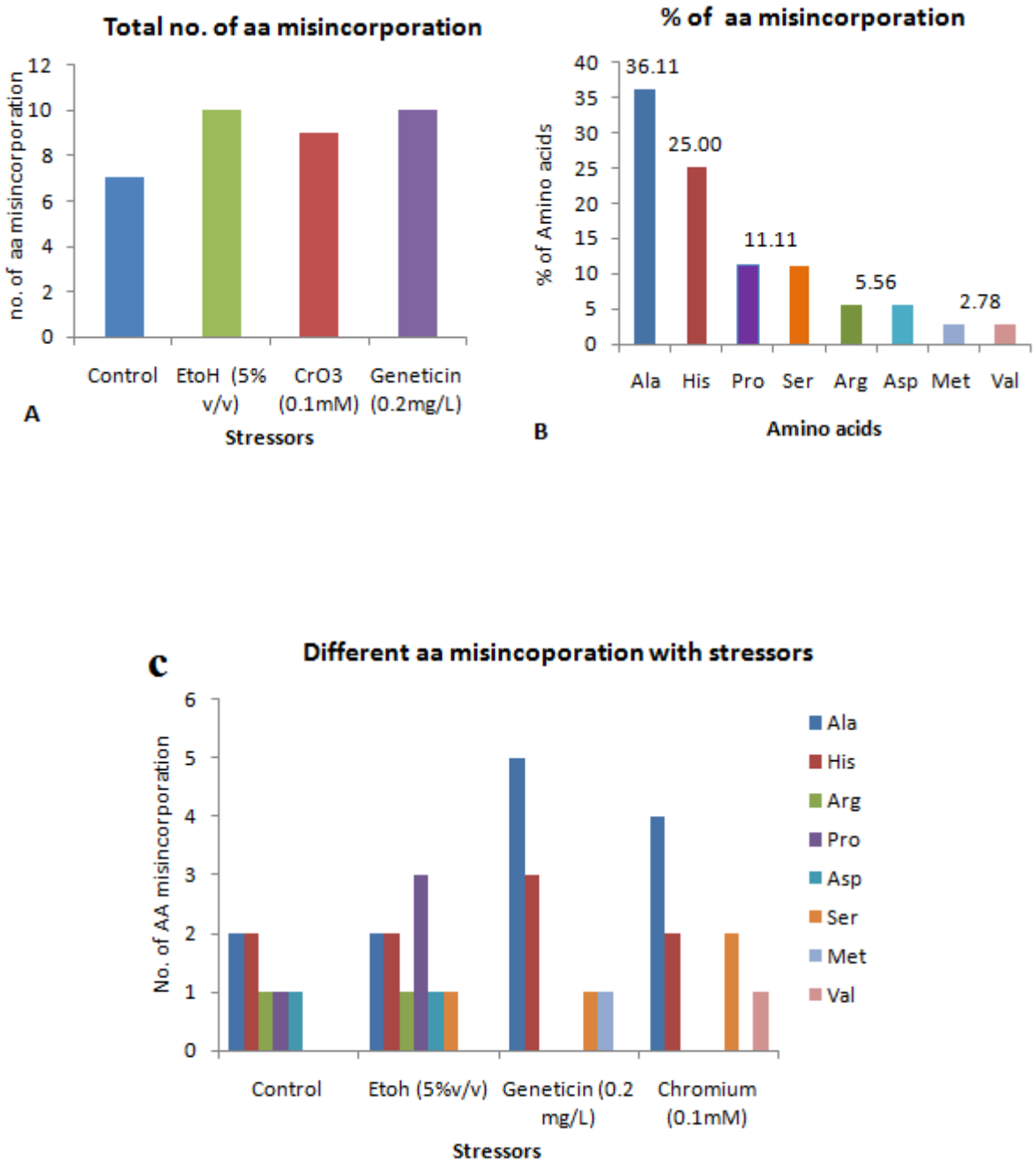

Figure 19: Bar charts showing **A)** Total number of amino acids misincorporation in GST-β galactosidase protein produced in presence of different stressors, **B)** Percentage of amino acids misincorporation in GST-β galactosidase protein and **C)** Number of different amino acids in GST-β galactosidase protein produced in presence of the stressors indicated.

5. Discussion

5.1 Effect of environmental stressors on the growth of *Saccharomyces cerevisiae*

Environmental stressors have a strong affect on yeast growth rate. Both the strains, CEN.PK2 to BY4742 strains were affected by geneticin as expected. A different pattern was observed in the case of chromium where the strain BY4742 showed reduced growth rate (18% vs. 40%) in contrast to CEN.PK2 strain. Both strains showed similar responses to ethanol. It is clear from many studies that yeast treated with ethanol initially struggle to maintain energy production, and various expression of genes related to energy-generating activities, namely glycolysis and mitochondrial function are up-regulated [82]. At the same time yeast down regulate expression of many genes associated with energy demanding processes, such as growth [82]. In the present study, 5% ethanol decreases the growth rate of the yeast strains CEN.PK2 and BY4742 by 40 %. After 64 hours both strains showed slight increase in OD, which could be explained by ethanol being metabolized. Regarding amino acid misincorporations, we are able to detect 10 amino acids misincorporated in GST-β gal protein produced in presence of 5% ethanol, which was higher than the levels of misincorporations in control cells (7 aa).

In the present study geneticin reduced the yield of GST-β galactosidase protein at least 3 times (0.21 mg/ml vs. 0.57 mg/ml) relative to that of the control cells, likely due to the strong negative effect of geneticin on the growth rate of the CEN.PK2 strain. The number of amino acids misincorporated in GST- β gal was also higher (10 vs. 7 aa) in geneticin treated cells than in control cells.

Chromium like other heavy metals also reduced the relative growth of both yeast strains studied and the concentration of GST-β galactosidase protein was also reduced to 50% (0.30 mg/ml vs. 0.57 mg/ml) in the chromium treated cells. Amino acids misincorporation into the GST-β gal increased (7 vs. 9 aa) relative to control cells.

5.2 β -gal assay

The enzyme β–galactosidase (β-gal) has an active quaternary structure and it is an efficient reporter protein for sense codon misreading due to the negative effects of the misincorporation on the thermostability of the enzyme [75]. However, in the present study the misincorporation of amino acids in the GST- β gal purified from control, ethanol and chromium treated cells did not show visible effects on the heat stability of the expressed enzyme. Only geneticin treated CEN.PK2 strain showed a relative decrease in β-gal activity (%) during first twelve minutes of thermal inactivation. Therefore, the number of amino acids misincorporations in GST-β gal was not sufficient to cause a pronounced heat stability difference between control, ethanol and chromium treated yeast cell. Conversely, the geneticin treated sample showed misincorporation of amino acids at the position Arg_{662} (to His_{662}) and His_{771} (to Met_{771}). These misincorporations were not observed in GST- β gal protein purified from cells treated with the other environmental stressors. Therefore, the histidine and methionine substitutions may explain the decreased thermostability of the β-galactosidase enzyme relative to the other samples. However, further tests are necessary to confirm these results.

5.3 Purification of β- galactosidase protein from yeast

One of the objectives of this study was to optimize the GST- β gal purification protocol. The 26-kD GST tag is short (218 aa) [27] and is frequently used as a fusion in molecular biology research. The study from Pawel and co-workers [86] defined soluble protein as the fraction that stayed in solution, did not oligomerize strongly and was stable (did not precipitate or aggregate), while insoluble proteins are defined as those that can not stay in solution without denaturing agents like urea. Our β-galactosidase protein was successfully purified from the soluble protein extracts of yeast strain CEN.PK2 lysates. However, the GST-β galactosidase protein purified from the insoluble fraction was not sufficient for SDS-PAGE gel and western blotting analysis. Several trials following different protocols [9, 47] to solubilize and purify the GST-β gal from inclusion bodies were carried out using lysis buffers containing DTT, Triton-X (1%), N-lauroylsarcosine sodium salt (sarkosyl) or NaCl (1M), but they were not successful. A likely reason for these results is that the GST tag increases solubility of β-galactosidase. Indeed, a study from Kim and Lee [87] showed that GST can act as a solubility tag. Another study [88] suggests that the GST tag has to be folded properly

in order to bind glutathione and should be purified under non denaturing conditions. In the present study a very faint band was observed in the insoluble fraction β-galactosidase treated with 8M urea. Therefore, the GST tag may solubilize β-galactosidase and urea may prevent its purification by denaturing GST. Additional experiments are therefore needed to clarify this issue.

5.4 Mass spectrometry analysis

E. coli β-galactosidase protein produced in *Saccharomyces cerevisiae* showed various amino acids misincorporations. Indeed, at least 7 amino acids were misincorporated in the GST-β galactosidase protein produced in control cells and 9-10 amino acids were misincorporated under stress. The frequency of errors was 5.6×10^{-3} to 8×10^{-3} which represents the 60-80 fold increase relative to the typical errors 1×10^{-4} [19], under normal growth conditions. Two amino acids substitution, namely Alanine for Serine (AGU) at position 423 and histidine for arginine (CGC) at position 274 and 1173 in our GST- β gal sequence are common and may not be related with stressor. The lacZ gene has 9, 5′-AGU-3′ and 37, 5′-CGC-3′ codons, but misincorporation occurred only at the above mentioned positions. It is believed that misincorporation can be caused by either an anticodon-codon mismatch, *i.e.* misreading or the use of an erroneously charged tRNA, *i.e.* mischarging or misacylation [19]. The histidine was misincorporated at position [CGC_{662}] in geneticin treated cells. The preferred codons for histidine residue in yeast cells are 5′-CAC-3′ (frequency 7.8/1000) or 5′-CAU -3′ (frequency 13.6/1000). So, in this case the codon for histidine differing by only one base at the second position might be misread. Many studies support that most misreadings occur at the third position of the codon due to a weaker interaction with its cognate base than do bases at the first and second positions [79]. Although misreading at the second codon position is very rare [19, 89], misincorporations at this position have been reported in several studies. For example, G•U mismatches that include serine to asparagine mistranslation (AGC to AAC) [89] and arginine to asparagine mistranslations (CGG to CAG) [90] have been demonstrated. Some of the misincorporation such as proline for glutamine codon CAA_{243} (ethanol), serine for leucine codon UUA_{242} (chromium), valine for tryptophan codon UGG_{267} (chromium) and methionine for histidine codon CAC_{771} (geneticin) occurred once in GST-β-galactosidase produced in presence of different stressors.

6. Conclusion

Fidelity in protein synthesis is essential whether protein is produced in native or heterologous hosts. However, gene translation into corresponding amino acid sequences is not always accurate. This is also true during recombinant protein production in heterologous hosts which often gives rise to aberrant proteins. Faithful translation from messenger RNA (mRNA) to protein depends on many factors and is a crucial step during gene expression. With the aim of quantifying misincorporation of amino acids in recombinant proteins, GST-β galactosidase was expressed in yeast. Technological innovations, in mass spectrometry now allow for detection of errors with high precisions. Indeed, our data show that environmental stressors increase the misincorporation at least 60-80 folds relative to that observed in control cells. The misincorporation of amino acids was only studied in soluble protein extract and we faced technical hurdles during the purification of primarily insoluble proteins under non-denaturing conditions. However, misincorporation in insoluble protein should also be determined in the future to determine whether such misincorporations increase protein aggregation.

References:

1. Chris Impey (2011) Molecules of life. Retrieved from *http://www.teachastronomy. com /astropedia/article/Molecules-of-Life*

2. Crick, F.H (1958) On protein synthesis. *Symposia of the society for experimental biology*. 1958, 12: p.138 -63.

3. Drummond D.A, Wilke C.O (2009) The evolutionary consequences of erroneous protein synthesis. *Nature Review Genetics* 10: 715-724.

4. Birkholtz, L.M., Blatch, G., Coetzer, T.L. et al. (2008a) Heterologous expression of plasmodial proteins for structural studies and functional annotation. *Malaria Journal* 7.

5. Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodriguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C, Porro D, Ferrer P, Tutino ML, Mattanovich D, Villaverde A (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Mirob Cell Fact* 2008, 7:11.

6. Chow, M.K et al. (2006) The REFOLD database: a tool for the optimization of protein expression and refolding. *Nucleic Acids Res.*, 34, D207-D212.

7. Lee D. Kapp and Jon R. Lorsch (2004) The molecular mechanics of eukaryotic translation. *Annu. Rev. Biochem*., 75: 657-704.

8. F. Gebauer and Matthias W. Hentze (2004) Molecular mechanisms of translational control. *Nature Reviews Molecular Cell Biology* 5, 827-835.

9. Dae-Wook Park, Sang-Soo Kim, Min-Kyung Nam, Goo-Young Kim, Jungho Kim & Hyangshuk Rhim (2011) Improved recovery of active GST-fusion proteins from insoluble aggregates: solubilization and purification conditions using PKM2 and HtrA2 as model proteins. *BMB reports*, 44(4): 279-284.

10. Gietz, R.D. and R.A. Woods. (2002) Transformation of yeast by the Liac/SS carrier DNA/PEG method. *Methods in Enzymology* 350: 87-96.

11. Wells, S.E., Hillner, P.E., Vale, R.D. & Sachs, A.B (1998) Circularisation of mRNA by eukaryotic translation initiation factors. *Mol. Cell* 2, 135-140.

12. Hershey, J.W.B & Merrick, W.C (2000) Translation control of gene expression (eds Sonenberg, N., Hershey, J. W. B & Mathews, M. B.) *Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory, New York,* 2000, 33-38.

13. Merrick, W.C (2004) Cap-dependent and cap-independent translation in eukaryotic systems. *Gene* 332, 1-11.

14. Green R, Noller H.F (1997) Ribosomes and translation. *Annu. Rev. Biochem*. 66: 679-716.

15. Wintermeyer W, Savelsbergh A, Semenkov Y. P, Katunin V. I, Rodnina M. V (2001) Mechanism of elongation factor G function in tRNA translocation on the ribosome. *Cold Spring Harbor Symp. Quant. Bio.* 66: 449-58.

16. Zaher H.S and Green R. (2009a) Fidelity at the molecular level: lessons from protein synthesis. *Cell*, 136, 746-762.

17. Zaher H.S and Green R (2009b) Quality control by the ribosome following peptide bond formation. *Nature*, 457, 161-166.

18. Denis L.J. Lafontaine & David Tollervey (2001). The function and synthesis of ribosomes. *Nature Reviews Molecular Cell Biology* 2, 514-520.

19. Parker, J (1989) Errors and alternatives in reading the universal genetic code. *Microbiol. Rev*. 53, 273-298.

20. Spanjaard, R.A., Chen, K., Walker, J.R., and Van Duin, J. (1990) Frame shift suppression at tandem AGA and AGG codons by cloned tRNA gene: assigning a codon to argU tRNA and T4 tRNA(Arg). *Nucleic Acids Res.* 18, 5031-5036.

21. Brinkmann, U., Mattes, R. E., and Buckel, P. (1989) High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product. *Gene* 85, 109-114.

22. Seetharam, R., Heeren, R. A., Wong, E. Y., Braford, S. R., Klein, B. K., Aykent, S., Kotts, C. E., Mathis, K. J., Bishop, B. F., Jennings, M. J., Smith, C. E., and Siegel, N. R. (1988) Mistranslation in IGF-1 during over-expression of the protein in *Escherichia coli* using a synthetic gene containing low frequency codons. *Biochem. Biophys. Res. Commun.* 155, 518-523.

23. Calderone, T.L., Stevens, R D., and Oas, T.G. (1996) High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*. *Journal of Molecular Biology*, 262, 407-412.

24. Bogosian, G., et al. in : The ribosome : Structure, Function and Evolution, Hill, W. E., Dahlberg, A., Garrett, R. A., Moore, P. B., Schlessinger, D. and Warner, J. R., Eds. *American Society for Microbiology*, Washington. DC 1990, p. 546.

25. Powers, E.T., Balch W.E. (2008) Costly mistakes: translation infidelity and protein homeostasis. *Cell* 134:204-206.

26. Bogosian, G., Violand, B.N., Dorward-King, E.J., Workman, W.E., Jung, P.E., and Kane, J.F. (1989) Biosynthesis and incorporation into protein of norleucine by *Escherichia coli*. *J. Biol. Chem*., 264, 531.

27. Jordan J. Lichty, Joshua L. Malecki, Heather D. Agnew, Daniel J. Michelson-Horowitz, Song Tan (2005) Comparison of affinity tags for protein purification. *Protein Expression and Purification* 41, 98–105.

28. Carol A. Scorer, Martin J. Carrier and Robert F. Rosenberger (1991) Amino acid misincorporation during high-level expression of mouse epidermal growth factor in *Escherichia coli. Nucleic Acids Research*, Vol. 19, No. 13 3511-3516.

29. Dix, D.B., and Thompson, R.C (1989) Codon choice and gene expression: synonymous codons differ in translational accuracy. *Proc Natl Acad Sci* U. S. A. 1989, 86, 6888.

30. Sharp, P.M., Tuohy, T.M., Mosurski K.R (1986) Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* 14, 5125-5143.

31. Kane, J.F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol* 6: 494-500.

32. Wada, K., Wada, Y., Ishibashi, F., Gojobori, T., Ikemura. T. (1992) Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res* 20: 2111-2118.

33. Walsh, G. (2005) Biopharmaceuticals: recent approvals and likely directions. *Trends Biotechnol.* 23, 553-558.

34. Broach, J. R., Jones, E. W. and Pringle, J. R. (1991) Eds.: *The Molecular and Cellular Biology of the yeast Saccharomyces, Vol. 1. Genome Dynamics, Protein Synthesis, and Energetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

35. Forsburg, Susan, L (2001) "The Art and Design of Genetic Screens: Yeast." *Nature Reviews Genetics* 2, 659-668.

36. Hartwell, L.H (2004) "Yeast and Cancer." *Bioscience Reports* 24, 523-544.

37. Sherman, Fred (2002) "Getting started with yeast." *Methods in Enzymology* 3503-41.

38. F. Garrido, U.C. Banerjee, Y. Chisti and M. Moo-Young (1994) Disruption of a recombinant yeast for the release of β- galactosidase. *Bioseparation* 4: 319-328.

39. Valenzuela, P., Medina, A., Rutter, W.J., Ammerer, G. & Hall, B.D. (1982) Synthesis and assembly of hepatitis B virus surface antigen particles yeast. *Nature* 298, 347.

40. Hallewell, R. A., Mills, R., Tekamp Olson, P., Blacher, R., Rosenberg, S., Otting, F., Masiarz, F. R., Scandella, C. J. (1987) Amino terminal acetylation of authentic human Cu, Zn superoxide dismutase produced in yeast. *Biotechnology* 1, 5, 363-366.

41. Barr, P.J, Cousens, L S., Lee-Ng, C.T., Medina-Selby, A., Masiarz, F.R., Hallewell, R.A., Chamberlain, S.H., Bradley, J.D., Lee, D., Steimer, K.S., et al. (1988)

Expression and processing of biologically active fibroblast growth factors in the yeast *Saccharomyces cerevisiae*. *J Biol Chem*. 5; 263(31): 16471-16478.

42. Bathurst, I.C., Chester, N., Gibson, H.L., Dennis, A.F., Steimer, K.S. Barr, P.J. (1989) N myristylation of the human immunodeficiency virus type 1 gag polyprotein precursor in *Saccharomyces cerevisiae*. *J Virol.* ; 63 (7): 3176-3179.

43. Barr, P. J., Steimer, K. S., Sabin, E. A., Parkes, D., George-Nascimento, C., Stephans, J. C., Powers, M. A., Gyenes, A., Van Nest, G. A., Miller, E. T, et al. (1987) Antigenicity and immunogenicity of domains of the human immunodeficiency virus (HIV) envelope polypeptide expressed in the yeast *Saccharomyces cerevisiae*. *Vaccine* 5(2):90–101.

44. Barr, P.J., Gibson, H.L., Enea, V., Arnot, D.E., Hollingdale, M.R., and Nussenzweig, V. (1987) Expression in yeast of a *Plasmodium vivax* antigen of potential use in a human malaria vaccine. *J. Exp. Med.* 165, 1160.

45. Clements, J.M., O'Connell, L.I., Tsunasawa, S., and Sherman, F. (1989) Expression and activity of a gene encoding rat cytochrome c in the yeast *Saccharomyces cerevisiae*. *Gene* 83, 1.

46. Harry Towbin, Theophil Staehelint and Julian Gordon (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Biochemistry*, Vol. 76, no. 9, 4350-4354.

47. Ciplys, E., Samuel, D., Juozapaitis, M., Sasnauskas, K., Slibinskas, R., (2011) Overexpression of human virus surface glycoprotein precursors induces cytosolic unfolded protein response in *Saccharomyces cerevisiae*. *Microb Cell Fact*. 10:37.

48. Ibeas, J.I., and Jimenez, J. (1997) Mitochondrial DNA loss caused by ethanol in Saccharomyces flor yeast. *Appl. Environ. Microbiol*. 63:7-12.

49. Chamber, P. (2007) Yeasts: products and discovery (YPD) in Australia. *Microbiol Aust* 28.

50. Hu, X.H., Wang, M.H., Tan, T., Li, J.R., Yang, H., Leach, L., Zhang, R.M. and Luo, Z.W. (2007) Genetic dissection of ethanol tolerance in the budding yeast *Saccharomyces cerevisiae*. *Genetics* 175, 1479-1487.

51. Walker, G.M. (1998) Yeast physiology and biotechnology. England: John Wiley & Sons Ltd.

52. Stanley, G.A., Hobley, T.J. and Pamment, N.B. (1997) Effect of acetaldehyde on *Saccharomyces cerevisiae* and *Zymomonas mobilis* subjected to environmental shocks. *Biotechnol Bioeng*, 53, 71-78.

53. Chandler, M., Stanley, G.A., Rogers, P. and Chambers, P. (2004) A genomic approach to defining the ethanol stress response in the yeast *Saccharomyces cerevisiae*. *Ann Microbiol* 54, 427-454.

54. Hallsworth, J.E., Nomura, Y. and Iwahara, M. (1998) Ethanol induced water stress and fungal growth. *J Ferment Bioeng* 86, 451-456.

55. Plesset, J., Palm, C. and McLaughlin, C.S. (1982) Induction of heat shock proteins and thermotolerance by ethanol in *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun* 108, 1340-1345.

56. Dayan, A.D. and Paine, A.J. (2001) Mechanisms of chromium toxicity, carcinogenicity and allergenicity: review of the literature from 1985 to 2000. *Human Exp Toxicol* 20, 439-451.

57. Gadd, G.M. (1992) Metals and microorganisms – a problem of definition. *FEMS Microbiol Lett* 100, 197-203.

58. Avery, S.V. (2001) Metal toxicity in yeasts and the role of oxidative stress. *Adv Appl Microbiol* 49, 111-142.

59. Edward R. Sumner, Anupama Shanmuganathan, Theodora C. Sideri, Sylvia A. Willetts, John E. Houghton and Simon V. Avery (2005) Oxidative protein damage causes chromium toxicity in yeast. *Microbiology*, 151, 1939-1948.

60. Sara Holland, Emma Lodwig, Theodora Sideri, Tom Reader, Ian Clarke, Konstantinos Gkargkas, David C Hoyle, Daniela Delneri, Stephen G Oliver, and Simon V Avery (2007) Application of the comprehensive set of heterozygous yeast deletion mutants to elucidate the molecular basis of cellular chromium toxicity. *Genome biology*, Volume 8, Issue 12, Article R268.

61. Panchal C.J., Whitney G.K and Stewart G.G (1984) Susceptibility of *Saccharomyces spp.* and *Schwanniomyces spp*. to the aminoglycoside antibiotic G418. *Appl Environ. Microbiol*., 47(5):1164.

62. Fowler, A.V. & Zabin, I. (1970) The amino acid sequence of beta galactosidase. Isolation and composition of tryptic peptides. *J Biol Chem*, 245, 5032-41.

63. Matthews, B.W. (2005) The structure of *E. coli* beta-galactosidase. *Comptes Rendus Biologies*, 328, 549-56.

64. Berg, J.M., Tymoczko, J. L. & Stryer, L. (2001) Biochemistry, New York, Freeman.

65. Andrews, K.J. & Lin, E.C. (1976) Thiogalactoside transacetylase of the lactose operon as an enzyme for detoxification. J. Bacteriol., 128, 510-513.

66. Sambrook, J., Fritsch, E.F. & Maniatis, T. (2008). Molecular cloning: a laboratory manual. (2nd ed.) *Cold Spring Harbor Laboratory Press.*

67. Jacobson, R.H., Zhang, X.J., Dubose, R.F. & Matthews, B.W. (1994) Three-dimensional structure of beta-galactosidase from *E. coli*. *Nature,* 369, 761-6.

68. Langley, K.E., Villarejo, M.R., Fowler, A.V., Zamenhof, P.J. & Zabin, I. (1975) Molecular basis of beta-galactosidase alpha-complementation. *Proc Natl Acad Sci U S A*, 72, 1254-7.

69. Lederberg, J. (1950) The beta-d-galactosidase of *Escherichia coli*, strain K-12. *J Bacteriol*, 60, 381-92.

70. Chunmei Yang, Jinong Fen, Wenjia Song, Jicheng Wan, Becky Tsai, Yunwu Zhang, William A. Scaringe, Kathleen A. Hill, Paris Margaritis, Katherine A. High, and Steve S. Sommer (2007) A mouse model for nonsense mutation bypass therapy shows a dramatic multiday response to geneticin. *PNAS*, vol. 104, no. 39, 15394–15399

71. Kvitek D.J, Will J.L, Gasch A.P (2008) Variations in Stress Sensitivity and Genomic Expression in Diverse *S. cerevisiae* Isolates. *PLoS Genet* 4(10).

72. Firoozan M, Grant C.M, Duarte J.A.B, Tuite M.F (1991) Quantitation of read through of termination codons in yeast using a novel gene fusion assay. *Yeast* 7:173-184.

73. Stansfield I., Akhmaloka and Tuite. M.F (1995) Mutant allele of the SUP45 (SAL4) gene of *Saccharomyces cerevisiae* shows temperature-dependent allosuppressor and omnipotent suppressor phenotypes. *Current genetics* 27, 417-426

74. T.M.M. Sogaard, C.G. Jakobsen, and J. Justesen (1999) A Sensitive Assay of Translational Fidelity (Readthrough and Termination) in Eukaryotic Cells. *Biochemistry (Moscow),* Vol. 64, No. 12, pp. 1408-1417.

75. Branscomb. E.W. and Galas. D.J. (1975) Progressive decrease in protein synthesis accuracy induced by streptomycin in *Escherichia coli*. *Nature*, 254. 161-163.

76. Rose M and Botstein D (1983) Construction and use of gene fusions to lacZ (beta-galactosidase) that are expressed in yeast. *Methods Enzymology*. 101: 167-80.

77. Smith, D.B. and Johnson, K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* 67, 31-40.

78. Walther TC, Mann M (2010) Mass spectrometry based proteomics in cell biology. *J Cell Biol* 190: 491.

79. Dingyi Wen, Malgorzata M. Vecchi, Sheng Gu, Lihe Su, Jana Dolnikova, Yao-Ming Huang, Susan F. Foley, Ellen Garber, Nels Pederson, and Werner Meier (2009) Discovery and Investigation of Misincorporation of Serine at Asparagine Positions in Recombinant Proteins Expressed in Chinese Hamster Ovary Cells. *J Biol Chem.* 284(47): 32686–32694.

80. Laughrea M, Latulippe J, Filion AM (1984) Effect of ethanol, phenol, formamide, dimethyl sulfoxide, paromomycin, and deuterium oxide on the fidelity of translation in a brain cell-free system. *Biochemistry* 14; 23(4):753-8.

81. Silva, Filipa Isabel Serra e. (2011) Mistranslation and telomere stability. *Master thesis in biology department*, University of Aveiro, Portugal.

82. D. Stanley, A. Bandara, S. Fraser, P.J. Chambers and G.A. Stanley (2010) The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*. *Journal of Applied Microbiology* 109, 13–24.

83. Pham, T.K and Wright, P.C. (2008) The proteomic response of *Saccharomyces cerevisiae* in very high glucose conditions with amino acid supplementation. *J Proteome Res 7*, 4766-4774.

84. Yoshikawa, K., Tanaka, T., Furusawa, C., Nagahisa, K., Hirasawa, T. and Shimizu, H. (2009) Comprehensive phenotypic analysis for identification of genes affecting growth under ethanol stress in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 9, 32-44.

85. Robert P. Harris, Jo Mattocks, Philip S. Green, Frank Moffatt, Peter M. Kilby (2012) Determination and control of low-level amino acid misincorporation in human thioredoxin protein produced in a recombinant *Escherichia coli* production system. *Biotechnol. Bioeng.*109:1987–1995.

86. Pawel Smialowski, Antonio J. Martin-Galiano, Aleksandra Mikolajka, Tobias Girschick, Tad A. Holak and Dmitrij Frishman (2007) Protein solubility: sequence based prediction and experimental verification. *Bioinformatics*, Vol 23 no. 19: 2536–2542.

87. Kim, S., and Lee, S.B. (2008). Soluble expression of archaeal proteins in *Escherichia coli* by using fusion-partners. *Protein Expr. Purif*. 62, 116–119.

88. Malhotra, Arun (2011) Tagging for protein expression. *Methods in enzymology*. 463:239-58.

89. Freeland SJ, Hurst LD (1998) The genetic code is one in a million. *J Mol Evol* 47: 238-248.

90. McNulty DE, Claffee BA, Huddleston MJ, Porter ML, Cavnar KM, et al. (2003) Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli. Protein Expr Purif.*, 27: 365-374.

91. Karin Hjernø and Peter Højrup. Peptide Mass Mapping by MALDI-TOF MS. Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark. Retrieved from *http://www.protein.sdu.dk/bm131/ download/ProteinIdentification.pdf*

92. Rose MD, Novick R Thomas JH, Botstein D, Fink GR (1987) A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. *Gene* 60:237-243

APPENDIX A

YPD agar plates

1. YPD medium was dissolved at a concentration of 6.7 g/L, 2% Agar (20 g/L), 0.2% drop out mix (2 g/L) and 2% glucose (20 g/L) in distilled water.

2. All the components was mix in a big flask and water was added until 1 L. The flask for sterilization of the medium was autoclaved;

3. Once the medium reached to room temperature, it was kept stirred for spreading in the dishes before solidified.

Note: The procedure for YPD liquid medium is equal without adding Agar

- **Isolating Genomic DNA from Yeast**

Materials to Be Supplied by the User

• 1.5ml microcentrifuge tubes

• YPD broth

• 50mM EDTA (pH 8.0)

• 20mg/ml lyticase (Sigma Cat.# L2524)

• Water bath, 37°C

• Isopropanol, room temperature

• 70% ethanol, room temperature

• Water bath, 65°C (optional; for rapid DNA rehydration)

1. Add 1ml of a culture grown for 20 hours in YPD broth to a 1.5ml microcentrifuge tube.

2. Centrifuge at 13,000–16,000 × *g* for 2 minutes to pellet the cells. Remove the supernatant.

3. Resuspend the cells thoroughly in 293µl of 50mM EDTA.

4. Add 7.5µl of 20mg/ml lyticase and gently pipet 4 times to mix.

5. Incubate the sample at 37°C for 30–60 minutes to digest the cell wall. Cool to room temperature.

6. Centrifuge the sample at 13,000–16,000 × *g* for 2 minutes and then remove the supernatant.

7. Add 300µl of Nuclei Lysis Solution to the cell pellet and gently pipet to mix.

8. Add 100µl of Protein Precipitation Solution and vortex vigorously at high speed for 20 seconds.

9. Let the sample sit on ice for 5 minutes.

10. Centrifuge at 13,000–16,000 × *g* for 3 minutes.

11. Transfer the supernatant containing the DNA to a clean 1.5ml microcentrifuge tube containing 300µl of room temperature isopropanol.

Note: Some supernatant may remain in the original tube containing the protein pellet. Leave this residual liquid in the tube to avoid contaminating the DNA solution with the precipitated protein.

12. Gently mix by inversion until the thread-like strands of DNA form a visible mass.

13. Centrifuge at 13,000–16,000 × *g* for 2 minutes.

14. Carefully decant the supernatant and drain the tube on clean absorbent paper. Add 300µl of room temperature 70% ethanol and gently invert the tube several times to wash the DNA pellet.

15. Centrifuge at 13,000–16,000 × *g* for 2 minutes. Carefully aspirate all of the ethanol.

16. Drain the tube on clean absorbent paper and allow the pellet to air-dry for 10–15 minutes.

17. Add 50µl of DNA Rehydration Solution.

18. Add 1.5µl of RNase Solution to the purified DNA sample. Vortex the sample for 1 second. Centrifuge briefly in a microcentrifuge for 5 seconds to collect the liquid and incubate at 37°C for 15 minutes.

19. Rehydrate the DNA by incubating at 65°C for 1 hour. Periodically mix the solution by gently tapping the tube. Alternatively, rehydrate the DNA by incubating the solution overnight at room temperature or at 4°C.

20. Store the DNA at 2–8°C.

- **QIAquick PCR Purification Kit Protocol (using a microcentrifuge)**

This protocol is designed to purify single- or double-stranded DNA fragments from PCR and other enzymatic reactions. Fragments ranging from 100 bp to 10 kb are purified from primers, nucleotides, polymerases, and salts using QIAquick spin columns in a microcentrifuge.

Important points before starting

■ Add ethanol (96–100%) to Buffer PE before use (see bottle label for volume).

■ All centrifugation steps are carried out at 17,900 x g (13,000 rpm) in a conventional tabletop microcentrifuge at room temperature.

■ Add 1:250 volume pH indicator I to Buffer PB (i.e., add 120 µl pH indicator I to 30 ml Buffer PB or add 600 µl pH indicator I to 150 ml Buffer PB). The yellow color of Buffer PB with pH indicator I indicates a pH of ≤7.5.

■ Add pH indicator I to entire buffer contents. Do not add pH indicator I to buffer aliquots.

Procedure

1. Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix. It is not necessary to remove mineral oil or kerosene. For example, add 500 µl of Buffer PB to 100 µl PCR samples (not including oil).

2. If pH indicator I has been added to Buffer PB, check that the color of the mixture isyellow. If the color of the mixture is orange or violet, add 10 µl of 3 M sodium acetate, pH 5.0, and mix. The color of the mixture will turn to yellow.

3. Place a QIAquick spin column in a provided 2 ml collection tube.

4. To bind DNA, apply the sample to the QIAquick column and centrifuge for 30 – 60 s.

5. Discard flow-through. Place the QIAquick column back into the same tube. Collection tubes are re-used to reduce plastic waste.

6. To wash, add 0.75 ml Buffer PE to the QIAquick column and centrifuge for 30 – 60 s.

7. Discard flow-through and place the QIAquick column back in the same tube. Centrifuge the column for an additional 1 min.

IMPORTANT: Residual ethanol from Buffer PE will not be completely removed unless the flow-through is discarded before this additional centrifugation.

8. Place QIAquick column in a clean 1.5 ml microcentrifuge tube.

9. To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water (pH 7.0–8.5) to the center of the QIAquick membrane and centrifuge the column for 1 min. Alternatively, for increased DNA concentration, add 30 µl elution buffer to the center of the QIAquick membrane, let the column stand for 1 min, and then centrifuge.

IMPORTANT: Ensure that the elution buffer is dispensed directly onto the QIAquick membrane for complete elution of bound DNA. The average eluate volume is 48 µl from 50 µl elution buffer volumes, and 28 µl from 30 µl elution buffer. Elution efficiency is dependent on pH. The maximum elution efficiency is achieved between pH 7.0 and 8.5. When using water, make sure that the pH value is within this range, and store DNA at –20°C as DNA may degrade in the absence of a buffering agent. The purified DNA can also be eluted in TE buffer (10 mM Tris·Cl, 1 mM EDTA, pH 8.0), but the EDTA may inhibit subsequent enzymatic reactions.

10. If the purified DNA is to be analyzed on a gel, add 1 volume of Loading Dye to 5 volumes of purified DNA. Mix the solution by pipetting up and down before loading the gel. Loading dye contains 3 marker dyes (bromophenol blue, xylene cyanol, and orange G) that facilitate estimation of DNA migration distance and optimization of agarose gel run time.

- **Preparation of Glutathione Sepharose media for use in batch purification:**

1. Gently shake the bottle to re-suspend the slurry.
2. Collect 200 µL for each sample of bed volume of Glutathione Sepharose and transfer in a 15 mL falcon tube
3. Sediment the chromatography medium by centrifugation at 500 g for 5 min.
4. Carefully decant the supernatant (using a pipet!).
5. Wash the Glutathione Sepharose 4B by adding 2 ml of 1X PBS for every 200 µL.
6. Sediment the chromatography medium by centrifugation at 500 g for 5 min.
7. Carefully decant the supernatant.
8. Repeat steps 5 and 6 once for a total of two washes (to remove ethanol!).
9. Re-suspend the washed slurry into 200 µL of 1X PBS (final slurry concentration of ~ 50% - Ready to use!)

❖ Immunoprecipitation Protocol Using Dynabeads® Protein A or Dynabeads® Protein G

Binding of Antibody (Ab)

1. Completely resuspend Dynabeads by pipetting or rotating on a roller (5 min).

2. Transfer 50 µl Dynabeads to a tube, place on magnet and remove supernatant.

3. Remove tube from magnet and resuspend the Dynabeads in 200 µl Ab Binding & Washing Buffer containing Ab of choice. *(Typically 1 - 10 µg Ab, the optimal amount needed will depend on the individual Ab used). Here, 1µl of* anti–b-galactosidase rabbit IgG fraction was used.

4. Incubate 10 minutes with rotation at room temperature.

5. Place tube on magnet and remove supernatant.

6. Remove tube from magnet and wash the Dynabeads-Ab complex by resuspending in 200 µl Ab Binding & Washing Buffer.

Immunoprecipitation of Antigen (Ag)

7. Place tube on magnet and remove supernatant

8. Add Ag-containing sample 100 µl (typical range 100 -1000 µl) to the Dynabeads-Ab complex and gently resuspend by pipetting.

9. Incubate 10 minutes at room temperature with rotation.

10. Place tube on magnet, transfer supernatant to a clean tube.

11. Wash the Dynabeads-Ab-Ag complex 3 times, using 200 µl Washing Buffer for each wash. Mix gently by pipetting.

12. Resuspend the Dynabeads-Ab-Ag complex in 100 µl Washing Buffer and transfer the suspension to a clean tube. Place tube on magnet and remove supernatant.

Elution of Ab/Ag complex *(alternatives A: denaturing or B: non-denaturing)*

A. Gently resuspend the Dynabeads-Ab-Ag complex in 20 µl Elution Buffer. Add 10 µl NuPAGE® LDS Sample Buffer / NuPAGE® Reducing Agent mix and incubate 10 minutes at 70°C. Place tube on magnet and load supernatant/sample onto a gel. *(Alternatively, the Dynabeads-Ab-Ag complex can be resuspended in the SDS sample buffer of your choice and heated as per your standard protocol prior to gel loading.)*

B. Gently resuspend the Dynabeads-Ab-Ag complex in 20 µl Elution Buffer. Incubate 2 minutes at room-temperature. Place tube on magnet and transfer supernatant/sample

APPENDIX B

PBS 1x (1000 mL)

8.2 g NaCl

0. 2 g KCl

1.41 g Na2HPO4;

0.24 g KH2PO4;

Volume adjusted to 1000 mL of distilled water and pH was adjusted to 7.4 with HCl and sterilized by autoclaving.

TGM 1x

25mM Tris-Base (3.03 g/L)

193 mM Glycine (14.4g/L)

20 % Methanol (200 mL) and Volume adjusted to 1000 mL of distilled water

TBS 1X (1000ml)

6.05 g Tris;

8.76 g NaCl;

800 ml of distilled water. The pH was adjusted to 7.5 with 1 M HCl and volume adjusted to 1 L with distilled water.

TBS-T (1000 ml)

1 ml of Tween-20 in 1L of 1X TBS buffer.

Coomassie Brilliant Blue (R-250)

40% Methanol

10% Glacial Acetic Acid

0.05% Coomassie Brilliant Blue R-250

Elution Buffer (100 mL)

0.61 g Tris-Base (50 mM)

0.31 g Reduced Glutathione (10 mM), adjust volume and P^H at 8 with HCl and then sterilized by filtration (0.20 µm)

Destain solution (1L)

100 mL Absolute ethanol (10%)

75 mL Glacial acetic acid (7.5%) and the volume adjusted to 1 L with distilled water.

SDS-PAGE gel recipe (for 2 gels):

Components	Resolving gel (10%)	Stacking gel (4%)
Distilled Water (Milli-Q)	3.6 ml	3.465 ml
SDS (10% w/v)	100 µl	50 µl
Tris-HCl	3.75 ml (1.0M pH 8)	1.0 ml (0.625M pH6.8)
Acrilamide/ Bis acrilamide (40% stock)	2.5 ml	0.5 ml
APS (10mg/100 µl)	50 µl	25 µl
TEMED	20 µl	20 µl

Z Buffer (1 liter):

16.1g Na_2HPO_4-$7H_2O$

5.5g NaH_2PO_4-H_2O

0.75g KCl

0.246g $MgSO_4$-$7H_2O$

H_2O to 1 liter (autoclave and store at 4°C)

2.7ml β-mercaptoethanol (add 270ul/100ml but only before use)

ONPG (stock solution)

4 mg/ml in Z buffer, also known as 2-nitrophenyl β-D-galactopyranoside

6x loading dye (10 ml)

6 ml Glyercol

1.2 ml EDTA (0.5M)

2.8 ml ddH2O (sterile)

pinch of bromophenol blue

pinch of xylene cyanol

- Pour glycerol into a graduated 10-15ml tube (do not pipette). Add other ingredients and mix by vortex. Do not use too much dye and store at room temperature.

APPENDIX C

Primer sequences: 5’→ 3’

Forward 1: TAATACGACTCACTATAGGG [Universal T7 promoter]

Reverse 1: CGGATACTGACGAAACGCCT

Forward 2: GATGAAGATCAGCCGTTTCC

Reverse 2: CCTATTGCTATAACCGCACT

Forward 3: TGGCAATTTAACCGCCAGTC

GST-β Gal Sequence (GST-LacZ)

MAKLPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIR
YIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHV
THPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVP
RGSGHHLEWLVPDPVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWRFAWFPAPEAV
PESWLECDLPEADTVVVPSNWQMHGYDAPIYTNVTYPITVNPPFVPTENPTGCYSLTFNVDESWLQEGQTRIIFD
GVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLRAGENRLAVMVLRWSDGSYLEDQDMWRMSGIFRDVSLLHKPT
TQISDFHVATRFNDDFSRAVLEAEVQMCGELRDYLRVTVSLWQGETQVASGTAPFGGEIIDERGGYADRVTLRLN
VENPKLWSAEIPNLYRAVVELHTADGTLIEAEACDVGFREVRIENGLLLLNGKPLLIRGVNRHEHHPLHGQVMDE
QTMVQDILLMKQNNFNAVRCSHYPNHPLWYTLCDRYGLYVVDEANIETHGMVPMNRLTDDPRWLPAMSERVTRMV
QRDRNHPSVIIWSLGNESGHGANHDALYRWIKSVDPSRPVQYEGGGADTTATDIICPMYARVDEDQPFPAVPKWS
IKKWLSLPGETRPLILCEYAHAMGNSLGGFAKYWQAFRQYPRLQGGFVWDWVDQSLIKYDENGNPWSAYGGDFGD
TPNDRQFCMNGLVFADRTPHPALTEAKHQQQFFQFRLSGQTIEVTSEYLFRHSDNELLHWMVALDGKPLASGEVP
LDVAPQGKQLIELPELPQPESAGQLWLTVRVVQPNATAWSEAGHISAWQQWRLAENLSVTLPAASHAIPHLTTSE
MDFCIELGNKRWQFNRQSGFLSQMWIGDKKQLLTPLRDQFTRAPLDNDIGVSEATRIDPNAWVERWKAAGHYQAE
AALLQCTADTLADAVLITTAHAWQHQGKTLFISRKTYRIDGSGQMAITVDVEVASDTPHPARIGLNCQLAQVAER
VNWLGLGPQENYPDRLTAACFDRWDLPLSDMYTPYVFPSENGLRCGTRELNYGPHQWRGDFQFNISRYSQQQLME
TSHRHLLHAEEGTWLNIDGFHMGIGGDDSWSPSVSAEFQLSAGRYHYQLVWCQK* = 1254 Amino acid

Note: **GST** - (214 aa) matches from position 5 to 218 aa (UniProt identifier accession no. P08515 and entry name GST26_SCHJA) and **LacZ** - (1015 aa) matches from aa position 240 to 1254 (UniProt identifier accession no. Q8VNN2 and entry name BGAL_ECOLX).

APPENDIX D

Difference of molecular weight with native GST-β gal protein: Control 119 Daltons (Da), Ethanol 265 Da, Geneticin 443 Da and chromium 475 Da.

Protein: GST-β gal protein

Length = 1254 amino acids
MW = 143310.47 Daltons

Amino Acid		Number	Mol%
Ala	A	87	6.93
Cys	C	20	1.59
Asp	D	82	6.53
Glu	E	79	6.29
Phe	F	47	3.75
Gly	G	87	6.93
His	H	42	3.35
Ile	I	51	4.06
Lys	K	42	3.35
Leu	L	126	10.04
Met	M	31	2.47
Asn	N	51	4.06
Pro	P	78	6.22
Gln	Q	63	5.02
Arg	R	76	6.06
Ser	S	68	5.42
Thr	T	60	4.78
Val	V	75	5.98
Trp	W	44	3.51
Tyr	Y	45	3.59

Protein: GST-β gal from Control sample
Length = 1254 amino acids
MW = 143191.25 Daltons

Amino Acid		Number	Mol%
Ala	A	89	7.09
Cys	C	19	1.51
Asp	D	83	6.61
Glu	E	79	6.29
Phe	F	47	3.75
Gly	G	87	6.93
His	H	44	3.51
Ile	I	51	4.06
Lys	K	41	3.27
Leu	L	126	10.04
Met	M	31	2.47
Asn	N	51	4.06
Pro	P	79	6.29
Gln	Q	63	5.02
Arg	R	75	5.98
Ser	S	66	5.26
Thr	T	60	4.78
Val	V	75	5.98
Trp	W	43	3.43
Tyr	Y	45	3.59

Protein: GST-β gal from Etoh treated sample
Length = 1254 amino acids
MW = 143045.06 Daltons

Amino Acid		Number	Mol%
Ala	A	89	7.09
Cys	C	19	1.51
Asp	D	83	6.61
Glu	E	79	6.29
Phe	F	47	3.75
Gly	G	87	6.93
His	H	44	3.51
Ile	I	51	4.06
Lys	K	41	3.27
Leu	L	125	9.96
Met	M	31	2.47
Asn	N	51	4.06
Pro	P	81	6.45
Gln	Q	62	4.94
Arg	R	75	5.98
Ser	S	67	5.34
Thr	T	60	4.78
Val	V	75	5.98
Trp	W	42	3.35
Tyr	Y	45	3.59

Protein: GST-β gal from GENETICIN treated sample
Length = 1254 amino acids
MW = 142866.96 Daltons

Amino Acid		Number	Mol%
Ala	A	92	7.33
Cys	C	20	1.59
Asp	D	82	6.53
Glu	E	79	6.29
Phe	F	47	3.75
Gly	G	87	6.93
His	H	44	3.51
Ile	I	51	4.06
Lys	K	42	3.35
Leu	L	125	9.96
Met	M	32	2.55
Asn	N	51	4.06
Pro	P	78	6.22
Gln	Q	63	5.02
Arg	R	73	5.82
Ser	S	67	5.34
Thr	T	60	4.78
Val	V	75	5.98
Trp	W	42	3.35
Tyr	Y	44	3.51

Protein: GST-β gal from CHROMIUM Treated sample
Length = 1254 amino acids
MW = 142834.86 Daltons

Amino Acid		Number	Mol%
Ala	A	91	7.25
Cys	C	20	1.59
Asp	D	82	6.53
Glu	E	79	6.29
Phe	F	47	3.75
Gly	G	87	6.93
His	H	44	3.51
Ile	I	51	4.06
Lys	K	42	3.35
Leu	L	124	9.88
Met	M	31	2.47
Asn	N	51	4.06
Pro	P	79	6.29
Gln	Q	63	5.02
Arg	R	73	5.82
Ser	S	68	5.42
Thr	T	60	4.78
Val	V	76	6.06
Trp	W	42	3.35
Tyr	Y	44	3.51