



Determination of amino acids misincorporation in recombinant protein by mass spectrometry

MZ Alam^{1*}, L Regioneiri², MAS Santos²

¹Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh; ² RNA Biology Laboratory, Department of Biology and CESAM, University of Aveiro, Aveiro 3810-193, Portugal

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 homeostasis. 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, *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 soluble protein fractions by mass-spectrometry to determine how much misincorporation occur. 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.

Key words: Amino acids misincorporation, mass spectrometry, recombinant protein, β -galactosidase

Bangladesh Animal Husbandry Association. All rights reserved.

Bang. J. Anim. Sci. 2013. 42 (1): 11-19

Introduction

Protein synthesis errors can be defined as any disruption in the conversion of a coding sequence into a functional protein (Drummond et al. 2009). They arise at all steps of gene expression, from transcription to protein folding, and have many widespread phenotypic consequences. Normally, translation is a remarkably accurate process and it is believed the fidelity of translation is more accurate in mammalian cells (Santos et al. 1993). Cells have their own mechanism to control over translation. However, a variety of translational errors have been observed during expression of proteins in eukaryotes, including frame shifts, premature truncation, read-through, tRNA misacylation and amino acid misincorporation (Drummond et al. 2009). Among these errors, amino acid misincorporations during translation are estimated to occur once in every 1,000-10,000 codons translated (Parker 1989) and at a much higher frequency in recombinant proteins that are overexpressed (Bogosian et al. 1990). 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 (Powers et al. 2008). For example, overexpression of bovine somatotropin in *E. coli* caused up to 14% of all the methionine residues to be substituted by norleucine (Bogosian et al. 1989).

One of the source of amino acid misincorporation is 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 (Dix et al. 1989). 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 (Sharp et al. 1986; Kane 1995).

In this paper, we reported the misincorporation of amino acids in 1254 amino acid long GST- β galactosidase protein when *E. coli LacZ* gene with GST tag was expressed in yeast *Saccharomyces cerevisiae*. We have chosen yeast as a homologous model of lower eukaryotic system for the heterologous protein expression and many pharmaceutically important proteins have been produced in yeast since long before. Moreover, we exposed yeast cells with different types of stressors in order to increase protein aggregation because polypeptide errors due to amino acid misincorporation can induce protein misfolding and aggregation. Here three stressors, namely ethanol, chromium and geneticin (antibiotic) were chosen to examine the effect of stress on the solubility and misincorporation of amino acids of β -galactosidase protein.

*Corresponding Author: mzalamgen@yahoo.com

Current proteomic technologies are being used to identify and quantify amino acids misincorporated into proteins by analyzing theoretical and experimental mass spectrometry data. Mass spectrometry (MS) is a high-throughput technology for protein identification and quantification with high-accuracy (Walther et al. 2010). 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 (Dingyi et al. 2009). In the present study we used a MS technique known as Matrix Assisted Laser Desorption Ionization (MALDI) –Time of Flight (TOF) which has become a major ionization method for analysis of large biological samples and synthetic polymers (Karin and peter 2007). Therefore, the main objective of this study was to quantify amino acid misincorporation in a recombinant protein GST- β gal under environmental stress condition.

Materials and methods

The experiment was conducted at the RNA biology laboratory under the Department of Biology, University of Aveiro, Portugal from October 2011 to July 2012. Mass spectrometry analysis was done at the Aveiro University Mass Spectrometry centre under the Department of Chemistry.

Yeast strain and plasmid

Diploid *Saccharomyces cerevisiae* CEN.PK2 (MATa/MATa, ura3-52/ura3-52, trp1-289/trp1-289, leu2-3_112/leu2-3_112, his3 Δ 1/his3 Δ 1, MAL2-8C/MAL2-8C, SUC2/SUC2) was used in this study. The multicopy plasmid vector pGL-C1 carrying *E. coli* lacZ gene, producing β -galactosidase protein, was transformed in CEN.PK2 strain under the control of the GPD promoter as a fusion with glutathione S-transferase (GST- β gal). Plasmid DNA preparation and transformation was performed according to standard protocol (Sambrook et al. 2001 and Gietz et al. 2002). The strain was preserved at -80° C in minimal medium (MM) with 40% glycerol.

Growth medium and conditions

S. Cerevisiae strain CEN.PK2 was grown at 30° C with agitation (180 rpm) in minimal medium

(0.67% bacto yeast nitrogen without amino acids, 100 μ g/ml of each of the required amino acids, 2% glucose and 2% bacto agar for solid medium) lacking tryptophan. The yeast strain was grown until late exponential phase (between 0.6 and 0.8 OD₆₀₀). 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 OD₆₀₀ of 0.02. Since the stressors affect the yeast growth rate, OD₆₀₀ was measured during three days (around 96 hours) until the cultures reached stationary phase (OD₆₀₀ 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. After several preliminary screens (data not shown) the sub lethal dose of each stressor was determined. The final concentrations of the three different stressors were ethanol 5% (v/v), CrO₃ 0.1 mM, and Geneticin® (G418) 0.2 mg/L.

β –galactosidase thermal stability assay

β -galactosidase was used as reporter protein to quantify mistranslation induced by amino acid misincorporation in this protein. The assay was adapted from Rose et al. 1983 with some modifications. β -galactosidase thermal inactivation was also monitored in *S. cerevisiae* cell lysates when the cultures reached an OD₆₀₀ of ~ 1.5 as described by Sambrook et al. 2008 with small modifications describe elsewhere (unpublished data).

SDS-PAGE Gel electrophoresis and Western blotting

Sodium Dodecyl Sulphate (SDS) - Polyacrylamide Gel Electrophoresis (PAGE) was performed according to standard protocol (Sambrook et al. 2001). Since the GST- β galactosidase protein has a MW of 142 kDa as determined by using the BioEdit (version 7.0.9) sequence alignment software and the β galactosidase alone has a MW of 116 kDa, a 4% stacking and 10% resolving gels were used. The protein samples for SDS-PAGE were prepared by adding 6X loading buffer in a ratio of 1:6, followed by denaturing the samples for seven minute at 95° C. Around 15 μ L of protein preparation were loaded and the gels were run with 1X SDS running buffer at 80 V for the first 30 minutes. Then, the voltage was increased to 130 V and the gels were run for 2 hours. The gel was stained with Coomassie Brilliant Blue dye for 1-2 hours and destained overnight with destain solution. The presence of β

Amino acids misincorporation in recombinant protein

–galactosidase protein was confirmed by western blotting (Harry et al. 1979) using the primary antibody (anti β -galactosidase IgG fraction A-11132 raised in rabbit, Molecular probes) and a mouse anti-rabbit antibody reactive against the primary antibody and coupled to a fluorochrome that allows subsequent visualization.

Extraction, purification and GST- β galactosidase preparation for mass spectrometry analysis

For the extraction of protein (soluble and insoluble) from yeast, frozen cells were defrosted on ice and 5 mL of 1X PBS was added and the pellets were re-suspended by vortexing. Then 1mM PMSF and a Roche Protease Inhibitor cocktail were added. Around $\sim 1/3$ 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 μ m filter into a new 15 ml falcon tube ("total soluble" protein) and a small fraction (100 μ l) 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 μ m filter to a new 50 ml falcon tube ("total insoluble" protein) and a small fraction (100 μ l) was saved for SDS-PAGE and western blot analysis. Then the GST- β galactosidase protein was purified from the whole protein yeast lysates using Glutathione Sepharose 4B (GE Healthcare, Sweden) medium.

The Glutathione Sepharose medium was prepared according to manufacturer protocol and was incubated overnight. 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. The eluted protein fraction was collected in

a new falcon tube. The protein was then ready for SDS-PAGE and western blotting analysis. After purification the β -galactosidase protein was concentrated. The eluted GST- β galactosidase were concentrated using an Amicon Ultra-15 ml filters (Millipore) of 3,000 molecular weight cut-off (MWCO) ultra filtration membranes and the concentration of purified protein was measured using NanoDrop[™] spectrophotometer. The purified protein was then analyzed with MALDI TOF mass spectrometer in the form of in-solution. For in-solution digestion the protein samples were mixed with an aqueous buffer (NH_4HCO_3) containing trypsin (5%) and sent to mass spectrometry centre for further peptide mapping.

Results and Discussion

Transformations of the lacZ gene into *Saccharomyces cerevisiae*

The *E. coli* lacZ gene in the pGL-C1 plasmid was confirmed by amplification of the gene by polymerase chain reaction from the yeast strains CEN.PK2. The primers successfully amplified the GST-lacZ gene (1254 aa). In this sequence AA position from 5-218 (214 AA) matched with the GST sequence (UniProt identifier accession no. P08515 and entry 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 that was the product of lacZ gene.

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

The effect of different stressors on the growth of *S. cerevisiae* transformed with plasmid pGL-C1 was monitored by measuring the optical density at different time points at wavelength 600 nm. 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 1B). The Geneticin had the strongest effect on the growth rate (OD_{600} 0.018 h^{-1}) compared to other treatments and 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.

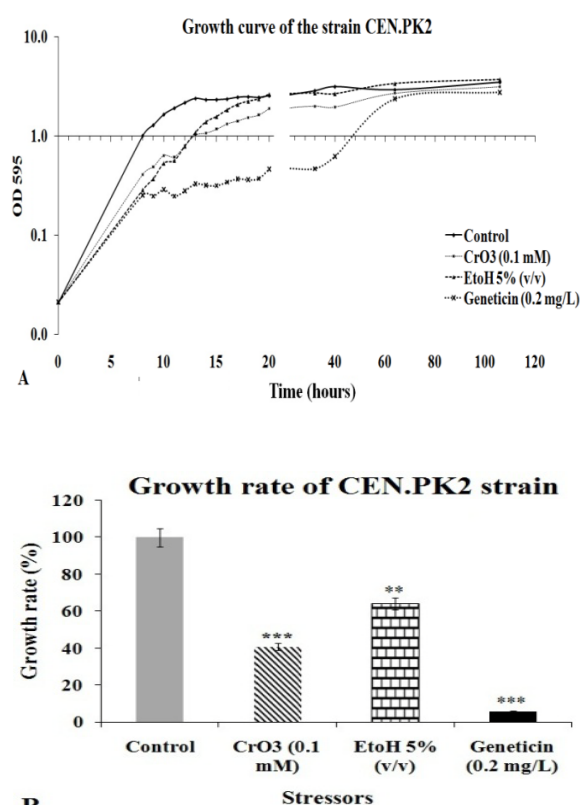


Figure 1. Growth curves (A) and rate (B) of *Saccharomyces cerevisiae* strain CEN.PK2 carrying pGL-C1 plasmid. The difference in growth rate of yeast highly significant at both 5% and 1% level of significance (student t-test: $t_{cal} > t_{tab}$)

Previous studies reported 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 (Stanley et al. 2010). At the same time yeast down regulate expression of many genes associated with energy demanding processes, such as growth (Stanley et al. 2010). In the present study, 5% ethanol decreases the growth rate of the yeast strains CEN.PK2 by 40%. The yeast showed slight increase in OD, which could be explained by ethanol being metabolized after 64 hours. 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). While in case of 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. 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 growth of 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) and amino acids misincorporation into the chromium treated cells increased (7 vs. 9 aa) relative to control cells.

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. We observed a decrease in activity of the β -galactosidase over heat inactivation time, while the strongest decrease of enzymatic activity was observed after 6 minutes of heat inactivation. However, relative differences in thermal stability in the β -gal 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 2).

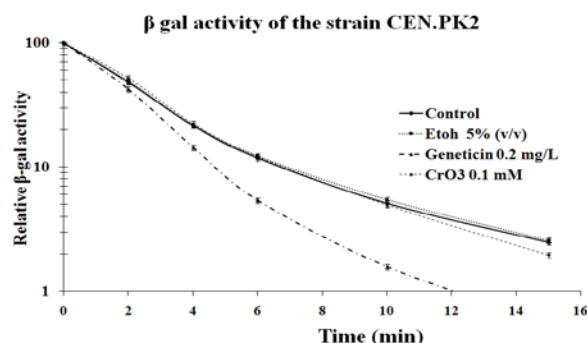


Figure 2. β -galactosidase activity of yeast strain CEN.PK2 carrying pGL-C1 plasmid. β -galactosidase thermal inactivation profiles were determined by measuring β -galactosidase activity directly in *S. cerevisiae* transformed with pGL-C1 plasmid

The enzyme 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 (Branscomb et al. 1975). However, in this study misincorporation of amino acids in GST- β gal purified from control, ethanol and chromium treated cells did not show visible effects on the heat stability of expressed enzyme. Only geneticin treated yeast showed a relative decrease in β -gal activity (%) during first twelve minutes of thermal inactivation. Therefore,

Amino acids misincorporation in recombinant protein

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 Arg₆₆₂ (to His₆₆₂) and His₇₇₁ (to Met₇₇₁). 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.

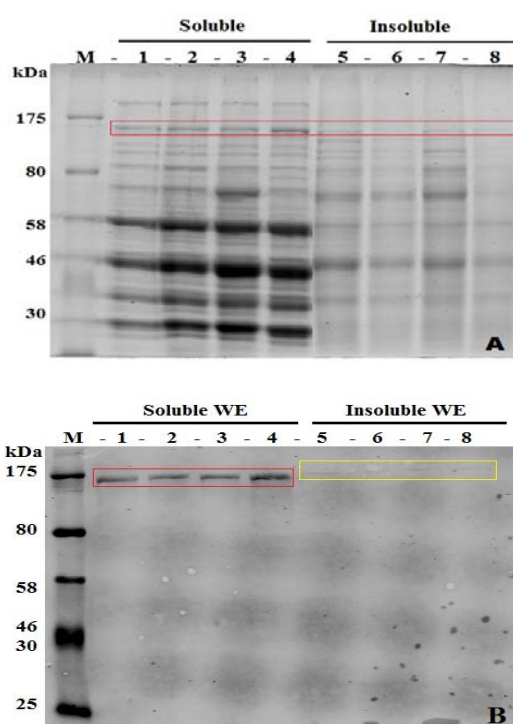


Figure 3. SDS-PAGE [A] and Western blotting [B] analysis showing the presence of GST- β galactosidase in the whole protein extract from yeast strain CEN.PK2 carrying multicopy plasmid pGL-C1: lane M-marker, lane 1, 2, 3 and 4 indicates soluble protein extract from untreated control, Ethanol, CrO₃ and Geneticin respectively. Lane 5, 6, 7, and 8 indicates insoluble protein extract from control, Ethanol, CrO₃ and Geneticin, respectively. The rectangle in Figure [A] indicates the position of GST- β galactosidase protein confirmed by western blotting. The rectangle in the figure [B] shows the GST- β -galactosidase bands in soluble and insoluble extract.

Purification of β -galactosidase protein from *Saccharomyces cerevisiae*

The presence of GST- β galactosidase protein was confirmed by SDS-PAGE and western blotting (Figure 3) and intense bands in soluble protein were observed in three different treatments as well as in untreated cells. The protein was detected both in soluble and insoluble fractions (Figure 3A), but the intensity of bands in insoluble fraction was very weak. After western blotting the bands were only observed in soluble fraction and were not detected in the insoluble whole extract. The predicted MW of mistranslated GST- β galactosidase protein was also calculated using the BioEdit (version 7.0.9) sequence alignment software, which showed 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 3B).

The β -galactosidase protein produced from lacZ gene has a molecular weight of ~116 kDa and the fusion with GST can increase the molecular weight up to ~142 kDa. Most of the GST- β galactosidase protein was soluble (Figure 3B), 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 concentration of purified GST- β galactosidase protein was recorded 0.57, 0.38, 0.30 and 0.21 mg/ml in control, ethanol, chromium and geneticin treated yeast respectively and confirmed by SDS-PAGE (data not shown).

Mass spectrometry analysis

Mass spectrometry data confirmed the amino acids misincorporations in GST- β galactosidase protein from yeast strain CEN.PK2 shown in Table 1. Only 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, this misincorporation of amino acid is not random, but depends on position and amino acids specificity.

Table 1. Amino acids misincorporation (bold and underlined) in GST – β galactosidase analyzed with MALDI-TOF mass spectrometry.

Treatments	Peptide fragments	AA → Misincorporated AA	After misincorporation of AA	Position in GST- β gal sequence
Control (Without Stressor)	NK K FELGLEFPNLPYYI	Lys → Arg	NK R FELGLEFPNLPYYIDGDVK	47
	DGDVK	Trp → Ala	D A ENPGVTQLNR	247
	D W ENPGVTQLNR	Arg → His	NSEEA H TDRPSQQLR	274
	NSEEA R TDRPSQQLR	Ser → Ala	W A DGSYLEDDQMWR	423
	WSDGSYLEDDQMWR	Cys → Pro	LIL P EYAHAMGNSLGGFAK	767
	LIL C EYAHAMGNSLGG FAK	Ser → Asp	H D DNELLHWMVAL	878
	H S DNELLHWMVAL	Arg → His	CGT H ELNYGPHQWR	1173
	CGT R ELNYGPHQWR			
	Ethanol 5% (v/v)	AK L PILGYWK	Leu → Ser	AKSPILGYWK
NK K FELGLEFPNLPYYI		Lys → Arg	NK R FELGLEFPNLPYYIDGDVK	47
DGDVK		Gln → Pro	GSGHHLEWLVPDPV L PR	243
GSGHHLEWLVPDPVV		Trp → Pro	D P ENPGVTQLNR	247
L Q R		Arg → His	NSEEA H TDRPSQQLR	274
D W ENPGVTQLNR		Trp → Ala;	A ADGSYLEDDQMWR	422, 423
NSEEA R TDRPSQQLR		Ser → Ala	LIL D EYAHAMGNSLGGFAK	767
W SDGSYLEDDQMWR		Cys → Asp	H P DNELLHWMVAL	878
LIL C EYAHAMGNSLGG FAK		Ser → Pro	CGT H ELNYGPHQWR	1173
H S DNELLHWMVAL		Arg → His		
CGT R ELNYGPHQWR				
Chromium 0.1 mM	AK L PILGYWK	Leu → Ser	AKSPILGYWK	4
	GSGHHLEWLVPDPVV	Leu → Ser	GSGHHLEWLVPDPVV S QR	242
	L Q R	Trp → Val	LAHPPFAS V R	267
	LAHPPFAS W R	Arg → His	NSEEA H TDRPSQQLR	274
	NSEEA R TDRPSQQLR	Trp → Ala;	A ADGSYLEDDQMWR	422; 423
	W SDGSYLEDDQMWR	Ser → Ala	CGT H ELNYGPHQWR	1173
	CGT R ELNYGPHQWR	Arg → His	A AQQQLMETSHR	1193; 1194
	Y SQQQLMETSHR	Tyr → Ala; Ser → Ala		
Geneticin 0.2 mg/L	AK L PILGYWK	Leu → Ser	AKSPILGYWK	4
	D W ENPGVTQLNR	Trp → Ala	D A ENPGVTQLNR	247
	NSEEA R TDRPSQQLR	Arg → His	NSEEA H TDRPSQQLR	274
	W SDGSYLEDDQMWR	Trp → Ala;	A ADGSYLEDDQMWR	422; 423
	LTDDP R WLPAMSER	Ser → Ala	LTDDP H WLPAMSER	662
	LILCEYA H AMGNSLGG FAK	Arg → His	LILCEYA M AMGNSLGGFAK	771
	H S DNELLHWMVAL	His → Met	CGT H ELNYGPHQWR	1173
	CGT R ELNYGPHQWR	Arg → His	A AQQQLMETSHR	1193; 1194
	Y SQQQLMETSHR	Tyr → Ala; Ser → Ala		

There were two substitutions of amino acids that were position specific, namely Ser → Ala in position 423 (Figure 4B) and Arg → His in position 274 (Figure 4A) and 1173 (Figure 4C) in β -galactosidase from the yeast treated with all stressors. The residues positions are mentioned here according to GST- β galactosidase protein (1-1254 aa) sequenced in this study (give accession no). Alanine may misincorporate at highest level (Figure 5B]. 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 gal protein.

Amino acids misincorporation in recombinant protein

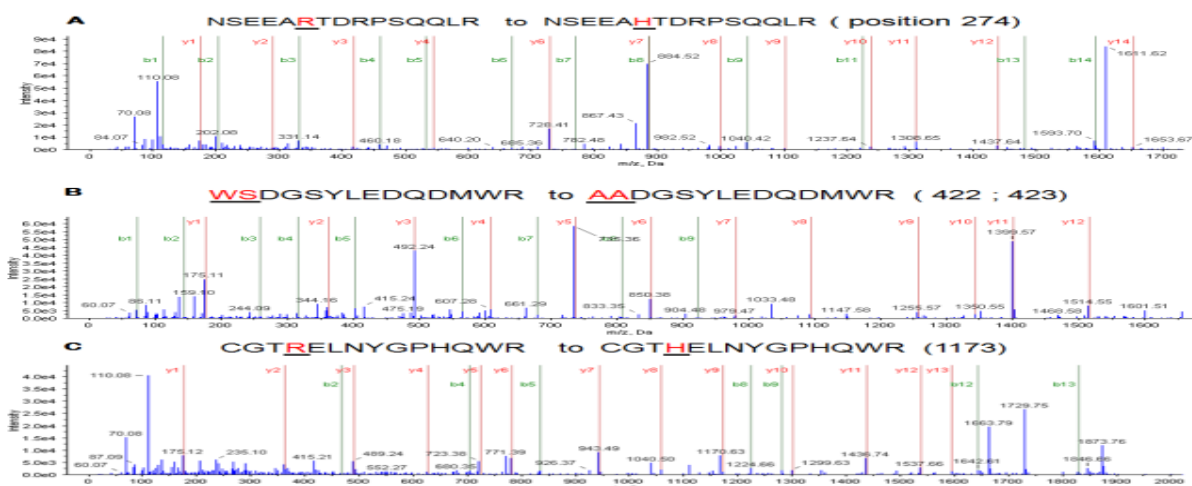
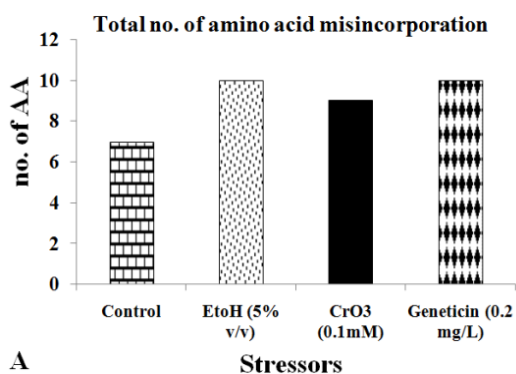
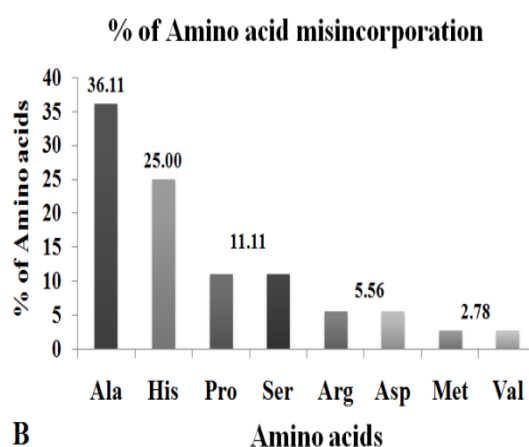


Figure 4. Misincorporation of amino acids in GST- β galactosidase protein (1254 aa) A) Histidine at arginine position, B) Alanine at tryptophan and serine position and C) Histidine at arginine position

The *E. coli* lacZ gene aligns with positions 240 to 1254 of the GST- β gal produced in the yeast 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 galactosidase protein. The rest of the misincorporations in the position 4 and 47 denote GST protein.

From the mass spectrometry data, we observed that 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} (Parker 1989), 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 (Parker 1989). The histidine was misincorporated only at position [CGC₆₆₂] 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.

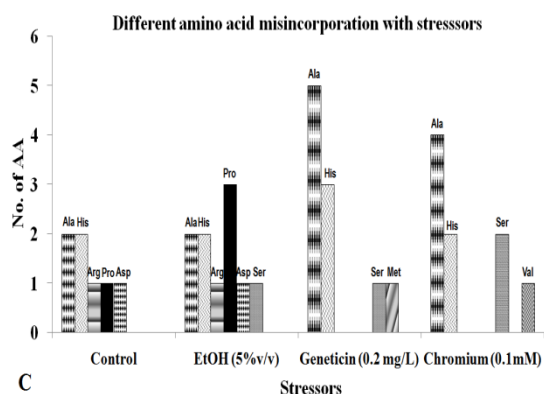


Figure 5. Bar charts showing A) Total number of amino acids misincorporation in GST- β galactosidase protein produced in presence of different stressors, B) Percentage of amino acids misincorporation in GST- β galactosidase protein and C) Number of different amino acids in GST- β galactosidase protein produced in presence of the stressors indicated

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 (Dingyi et al. 2009). Although misreading at the second codon position is very rare (Parker 1989, Freeland et al. 1998), misincorporations at this position have been reported in several studies. For example, G•U mismatches that include serine to asparagine mistranslation (AGC to AAC) (Freeland et al. 1998) and arginine to asparagines (CGG to CAG) mistranslations (McNulty et al. 2003) have been demonstrated. Some of the misincorporation such as proline for glutamine codon CAA₂₄₃ (ethanol), serine for leucine codon UUA₂₄₂ (chromium), valine for tryptophan codon UGG₂₆₇ (chromium) and methionine for histidine codon CAC₇₇₁ (geneticin) occurred once in GST- β -galactosidase produced in presence of different stressors. Further studies with other type of protein are needed to be consistent of these results.

Conclusion

Fidelity in protein synthesis is essential whether protein is produced in native or heterologous hosts. Faithful translation from messenger RNA (mRNA) to protein depends on many factors and gene translation into corresponding amino acid sequences is not always accurate. With the aim of quantifying misincorporation of amino acids in recombinant proteins, GST- β galactosidase was expressed in yeast and mass spectrometry was used for detection of errors with high precisions.

Our data show that environmental stressors increase the misincorporation at least 60-80 folds relative to that observed in control cells. The unexpectedly high error rate has implications for the therapeutic use of the heterologous host derived recombinant proteins. 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. The yeast can be a model organism to study amino acids misincorporation when pharmaceutically important animal protein expressed in yeast.

Acknowledgements

This work was supported by the RNA biology laboratory, Department of Biology, University of Aveiro, Portugal. M Z Alam was involved in all aspects of the experimental design, works, data collection, analysis and interpretation.

References

- Bogosian G, Violand BN, Dorward-King EJ, Workman WE, Jung PE, Kane JF (1989). Biosynthesis and incorporation into protein of norleucine by *Escherichia coli*. *Journal of Biological Chemistry*, 264: 531-539.
- Bogosian G, Violand BN, Jung PE, and Kane JF (1990). Effect of protein overexpression on mistranslation in *Escherichia coli*, RR 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, P. 546.
- Branscomb EW and Galas DJ (1975). Progressive decrease in protein synthesis accuracy induced by streptomycin in *Escherichia coli*. *Nature*, 254: 161-163.
- De Godoy LMF, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, Walther TC, Mann M. (2008). Comprehensive massspectrometry- based proteome quantification of haploid versus diploid yeast. *Nature*, 455: 1251-1254.
- Dingyi W, Malgorzata MV, Sheng G, Lihe S, Jana D, Yao-Ming H, Susan FF, Ellen G, Nels P, and Werner M (2009). Discovery and Investigation of Misincorporation of Serine at Asparagine Positions in Recombinant Proteins Expressed in Chinese Hamster Ovary Cells. *Journal of Biological Chemistry* 284: 32686-32694.
- Dix DB and Thompson RC (1989). Codon choice and gene expression: synonymous codons

Amino acids misincorporation in recombinant protein

- differ in translational accuracy. Proceeding of National Academy of Science, USA, 86: 6888.
- Drummond DA and Wilke CO (2009). The evolutionary consequences of erroneous protein synthesis. *Nature Review Genetics* 10: 715-724.
- Freeland SJ, Hurst LD (1998). The genetic code is one in a million. *Journal of Molecular Evolution*, 47: 238-248.
- Gietz RD and RA Woods. (2002). Transformation of yeast by the Liac/SS carrier DNA/PEG method. *Methods in Enzymology*, 350: 87-96.
- Harry Towbin, Theophil S and Julian G (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Biochemistry*, 76: 4350-4354.
- Kane JF (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology*, 6: 494-500.
- Karin H and Peter H (2007). Peptide Mass Mapping by MALDI-TOF MS. Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense.
- McNulty DE, Claffee BA, Huddleston MJ, Porter ML, Cavnar KM (2003). Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Expression and Purification*, 27: 365-374.
- Parker J (1989). Errors and alternatives in reading the universal genetic code. *Microbiological Reviews*, 53: 273-298.
- Powers ET and Balch WE (2008). Costly mistakes: translation infidelity and protein homeostasis. *Cell*, 134: 204-206.
- 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.
- Sambrook J, Fritsch EF and Maniatis T (2008). *Molecular cloning: a laboratory manual*. (2nd ed.) Cold Spring Harbor Laboratory Press.
- Sambrook J, Russell DW (2001). *Molecular cloning, A Laboratory Manual*. Third edition. CSHL Press, Cold Spring Harbour, New York, USA.
- Santos MA, Tuite MF (1993). New insights into messenger-RNA decoding - implications for heterologous protein-synthesis. *Trends in Biotechnology*. 11, 500-505.
- Sharp PM, Tuohy TM and Mosurski KR (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Research*, 14: 5125-5143.
- Stanley D, Bandara A, Fraser S, Chambers PJ, Stanley GA (2010). The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*. *Journal of Applied Microbiology*, 109: 13-24.
- Wada K, Wada Y, Ishibashi F, Gojobori T, Ikemura T (1992). Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Research*, 20: 2111-2118.
- Walther TC, Mann M (2010) Mass spectrometry based proteomics in cell biology. *Journal of Cell Biology*, 190: 491.