



**Universidade de  
Aveiro  
2005**

Departamento de Biologia

**Raquel Monteiro  
Marques da Silva**

**Reconstrução molecular de uma alteração ao código  
genético**

**Molecular reconstruction of a genetic code alteration**



**Universidade de  
Aveiro  
2005**

Departamento de Biologia

**Raquel Monteiro  
Marques da Silva**

**Reconstrução molecular de uma alteração ao código genético**

dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Biologia, realizada sob a orientação científica do Doutor Manuel Santos, Professor Auxiliar do Departamento de Biologia da Universidade de Aveiro

Apoio financeiro do POCTI no âmbito do III Quadro Comunitário de Apoio.

Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio.

Dedico este trabalho aos meus pais, à minha irmã e ao meu marido, pelo seu apoio incondicional e por, juntamente comigo, acreditarem e sonharem.

## **o júri**

presidente

**Doutor João Lemos Pinto**

Professor Catedrático do Departamento de Física da Universidade de Aveiro

**Doutor Cláudio Enrique Sunkel Cariola**

Professor Catedrático do Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto

**Doutor Amadeu Mortágua Velho da Maia Soares**

Professor Catedrático do Departamento de Biologia da Universidade de Aveiro

**Doutor António Carlos Matias Correia**

Professor Associado do Departamento de Biologia da Universidade de Aveiro

**Doutor Manuel António da Silva Santos**

Professor Auxiliar do Departamento de Biologia da Universidade de Aveiro (Orientador)

**Doutor Hélian Boucherie**

Directeur de Recherche CNRS à l'IBGC, Bordeaux

**Doutor Mick Tuite**

Director of Research in the Department of Biosciences at the University of Kent

## **Agradecimentos**

## **Acknowledgements**

I would like to thank my supervisor, Doutor Manuel Santos, for guidance and advice, and to all members of the lab, past and present, for support and fruitful discussions.

I am very grateful to Doutor Hélian Boucherie and all members of his group, for teaching me all about 2D-PAGE and for their support during my stay in Bordeaux. In particular my thanks go to Christelle Monribo-Espagne for technical assistance and Doutor Michel Perrot for his help with data analysis and for his good mood. A big thank you to Carine de Marcos, for making me feel at home!

I am also grateful to Doutor Frank Holstege and all members of his group, for the help and guidance on microarrays and for support during my stay in Utrecht, in special Marian G. Koerkamp.

I acknowledge all teachers and students from the GABBA program, namely those from the 1999-2000 edition, for sharing with me the enthusiasm of the first steps into science.

I am indebted to FCT, for financial support to this work.

A very special thanks to my parents and sister, for their encouragement and care. I also want to thank all the friends and family that have always supported me. In particular, my gratitude goes to D. Noémia Vital, for teaching me to look further and push my limits. Finally, I thank my husband João, for being by my side at all moments.

## palavras-chave

Código genético, ambiguidade, tRNA, erros de tradução, transcriptoma, proteoma.

## resumo

Várias espécies de *Candida* traduzem o codão CUG de leucina como serina. Esta alteração ao código genético é mediada por um novo tRNA (ser-tRNA<sub>CAG</sub>), que pode ser carregado com serina (97 %) e leucina (3 %) *in vivo*. Por esta razão o codão CUG é ambíguo, pois pode ser descodificado como serina ou leucina. Para elucidar o impacto da ambiguidade do código genético na expressão génica e na fisiologia da célula, o ser-tRNA<sub>CAG</sub> de *C. albicans* foi expresso em *Saccharomyces cerevisiae*. Isto induz a descodificação ambígua do codão CUG, devido à competição entre o tRNA endógeno que traduz o codão CUG como leucina e o *C. albicans* ser-tRNA<sub>CAG</sub>, que o traduz maioritariamente como serina.

A caracterização do transcriptoma e do proteoma das linhas celulares manipuladas de *S. cerevisiae* mostra que a ambiguidade do código genético induz alterações globais na expressão de genes e proteínas, com alterações na resposta ao stress, metabolismo dos hidratos de carbono e dos aminoácidos, estrutura e função da parede celular, síntese e degradação de proteínas. Adicionalmente, os resultados indicam que a tradução errada do codão CUG regula a expressão génica ao nível da tradução. A ambiguidade do codão CUG gera instabilidade do proteoma e genoma, contudo, estas células não perdem viabilidade. Pelo contrário, os dados sugerem que a resposta ao stress despoletada pela ambiguidade do codão CUG aumenta o potencial de adaptação, como é demonstrado pela tolerância que as células ambíguas têm a várias condições de stress.

Por estas razões, a reconstrução da alteração na descodificação do codão CUG providenciou dados importantes sobre o impacto que alterações ao código genético têm na adaptação e evolução das células. Este estudo também trouxe novas ideias acerca dos mecanismos que permitem a tolerância das células eucarióticas a elevados níveis de erro na tradução do mRNA.

**keywords**

Genetic code, ambiguity, tRNA, mistranslation, transcriptome, proteome.

**abstract**

Several *Candida* species translate the standard leucine CUG codon as serine. This genetic code alteration is mediated by a novel tRNA (ser-tRNA<sub>CAG</sub>), which can be charged both with serine (97 %) and leucine (3%) *in vivo*. Therefore, the CUG codon is ambiguous, since it can be decoded either as serine or leucine. To elucidate the impact of genetic code ambiguity on gene expression and cell physiology, the *C. albicans* ser-tRNA<sub>CAG</sub> was expressed in *Saccharomyces cerevisiae*. This induces ambiguous decoding of the CUG codon, due to competition between the endogenous tRNA that decodes the CUG codon as leucine and the *C. albicans* ser-tRNA<sub>CAG</sub>, which decodes it mainly as serine.

Transcriptome and proteome characterization of the engineered *S. cerevisiae* cell lines show that genetic code ambiguity induces global gene and protein expression changes, with alterations in the stress response, carbohydrate and amino acid metabolism, cell wall structure and function, protein synthesis and protein degradation. Additionally, the results indicate that CUG mistranslation regulates gene expression at the translational level. CUG ambiguity generates proteome and genome instability, however, these cells do not lose viability. Instead, the data suggests that the stress response triggered by CUG ambiguity increases adaptation potential, as shown by the tolerance of ambiguous cells to several stress conditions.

Therefore, the reconstruction of the CUG reassignment pathway provided important insight on the impact that genetic code alterations have on cell adaptation and evolution. This study also sheds new light on the mechanisms that allow eukaryotic cells to tolerate high levels of mRNA mistranslation.



## Index

Index.....	1
List of abbreviations.....	6

### **I. Introduction**

<b>1. The standard genetic code.....</b>	<b>11</b>
1.1. Transfer RNAs.....	14
1.2. Aminoacyl-tRNA synthetases.....	21
1.3. The genetic code and the origin of life.....	25
1.4. Mistranslation.....	27
<b>2. Evolution of the genetic code.....</b>	<b>31</b>
2.1. Special decoding events.....	34
2.1.1. Asparagine, glutamine and initiator methionine.....	34
2.1.2. The 21 <sup>st</sup> and 22 <sup>nd</sup> amino acids: Selenocysteine and Pyrrolysine.....	39
2.1.3. Natural genetic code alterations.....	43
2.2. Theories for the evolution of the genetic code.....	48
2.3. Artificial genetic codes.....	52
<b>3. The <i>Candida</i> spp. genetic code.....</b>	<b>56</b>
3.1. The ser-tRNA <sub>CAG</sub> .....	57
3.2. LeuRS and SerRS.....	61
3.3. CUG identity redefinition through decoding ambiguity.....	63
3.4. <i>Candida</i> spp. as a model system for studying the evolution of genetic code alterations.....	65
3.5. Objectives of this thesis.....	70

## **II. Materials and methods**

<b>1. Strains and growth conditions.....</b>	<b>73</b>
<b>2. Transcriptome analysis.....</b>	<b>75</b>
2.1. Total RNA extraction.....	75
2.2. mRNA enrichment.....	76
2.3. cDNA synthesis.....	78
2.3.1. RT reaction.....	78
2.3.2. Cleanup with Microcon-30 concentrators.....	78
2.4. cDNA labelling.....	79
2.4.1. Coupling monofunctional NHS-ester Cy-dyes.....	79
2.4.2. Cleanup with Chromaspin-30 columns.....	79
2.5. RNA quantification.....	81
2.6. Microarray hybridization.....	82
2.6.1. Enhanced prehybridization.....	82
2.6.2. Hybridization.....	82
2.6.3. Washing hybridized arrays.....	83
2.7. LifterSlip cleaning.....	83
2.8. Microarray production.....	84
2.9. Data analysis.....	84
<b>3. Proteome analysis.....</b>	<b>86</b>
3.1. Kinetics of <sup>35</sup> S-Methionine incorporation.....	86
3.2. Radioactive labelling.....	88
3.3. 2D-PAGE.....	88
3.3.1. Protein extraction.....	88
3.3.2. Radioactivity counting.....	89
3.3.3. First dimension.....	89
3.3.4. Second dimension.....	90
3.3.5. Washing of the material.....	91
3.3.6. Data analysis.....	92
3.4. Proteasome purification.....	93

3.4.1. Protein extraction.....	93
3.4.2. Complex purification.....	94
3.4.3. Column regeneration.....	95
<b>4. Physiological characterization.....</b>	<b>96</b>
4.1. Cell viability assays.....	96
4.2. Proteasome activity assay.....	96
4.3. Trehalose and glycogen quantification.....	97
4.4. SDS-PAGE.....	98
4.4.1. SDS-PAGE.....	98
4.4.2. Coomassie staining.....	99
4.4.3. Silver staining.....	99
4.5. Western Blot analysis.....	99
4.6. Protein carbonylation assays.....	101
4.7. Ubiquitylation assays.....	102
4.8. HSF-1 analysis.....	102
4.9. Phospho-Serine analysis.....	102
4.10. Karyotype analysis.....	103

### **III. Results**

<b>1. Transcriptome analysis of <i>S. cerevisiae</i> cells expressing CUG ambiguity.....</b>	<b>107</b>
1.1. Introduction.....	107
1.2. Transcriptome of <i>S. cerevisiae</i> cells expressing CUG ambiguity.....	109
1.2.1. CUG ambiguity induces the expression of stress genes.....	112
1.2.2. CUG ambiguity alters the expression of cell wall genes.....	114
1.2.3. CUG ambiguity induces the expression of phosphate metabolism genes.....	115
1.2.4. CUG ambiguity induces the expression of carbohydrate metabolism genes.....	117

1.2.5. CUG ambiguity represses the expression of protein synthesis genes.....	118
1.2.6. CUG ambiguity alters the expression of amino acid metabolism genes.....	119
1.2.7. Effect of CUG ambiguity on additional cellular functions.....	120
1.2.8. The response to CUG ambiguity is different from other stress responses.....	122
1.3. Transcriptome of <i>S. cerevisiae</i> expressing the ser-tRNA <sub>CAG-T<sub>33</sub></sub> .....	125
1.4. Transcriptome of <i>S. cerevisiae</i> expressing the ser-tRNA <sub>CAG-G<sub>33</sub></sub> .....	132
1.5. Discussion.....	138
<b>2. Proteome analysis of <i>S. cerevisiae</i> cells expressing CUG ambiguity...</b>	<b>142</b>
2.1. Introduction.....	142
2.2. Proteome of <i>S. cerevisiae</i> cells expressing CUG ambiguity.....	146
2.2.1. CUG ambiguity induces the expression of stress proteins.....	148
2.2.2. CUG ambiguity alters the expression of carbohydrate metabolism enzymes.....	152
2.2.3. CUG ambiguity induces the expression of proteasome subunits.....	154
2.2.4. CUG ambiguity represses the expression of amino acid metabolism enzymes.....	156
2.2.5. CUG ambiguity represses protein synthesis.....	158
2.2.6. CUG ambiguity regulates gene expression at the translational level.....	159
2.3. Proteome alterations under additional stress conditions.....	164
2.3.1. Proteome alterations under growth at 37°C.....	167
2.3.2. Proteome alterations under Heat-Shock.....	172
2.3.2.1. Characterization of the response of control cells to heat shock...	172
2.3.2.2. Characterization of the response of CUG ambiguous cells to heat shock.....	173
2.4. Proteasome subunit identification on 2D-maps.....	182
2.5. Discussion.....	188

<b>3. Physiological characterization of <i>S. cerevisiae</i> cells expressing CUG ambiguity</b> .....	<b>191</b>
3.1. Introduction.....	191
3.2. Effect of CUG ambiguity on cell viability.....	192
3.3. Effect of CUG ambiguity on proteasome activity.....	196
3.4. Effect of CUG ambiguity on trehalose and glycogen accumulation.....	200
3.5. Effect of CUG ambiguity on protein oxidation and ubiquitylation.....	207
3.6. Effect of CUG ambiguity on phosphoserine proteins and HSF-1.....	212
3.7. Karyotype of <i>S. cerevisiae</i> cells expressing CUG ambiguity.....	217

#### **IV. General Discussion**

<b>1. The cellular response to CUG ambiguity</b> .....	<b>223</b>
1.1. CUG ambiguity alters gene expression.....	225
1.2. The stress response triggered by CUG ambiguity.....	228
1.3. The role of molecular chaperones and the proteasome.....	229
1.4. Is the <i>C. albicans</i> ser-tRNA <sub>CAG</sub> a generator of phenotypic diversity?.....	231
1.5. Is there a role for CUG ambiguity during <i>C. albicans</i> infection?.....	233
<b>2. Conclusions</b> .....	<b>234</b>
<b>3. Future work</b> .....	<b>238</b>

<b><u>V. References</u></b> .....	<b>239</b>
-----------------------------------	------------

<b><u>VI. Annexes</u></b> .....	<b>275</b>
---------------------------------	------------



## List of abbreviations

A	Ampère
APS	ammonium persulphate
ATP	adenosine 5'-triphosphate
AZC	azetidine-2-carboxylic acid
cAMP	cyclic adenosine 5'-monohosphate
cDNA	complementary deoxyribonucleic acid
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
μCi	microcurie
cm	centimeter
Cy3	cyanine 3
Cy5	cyanine 5
DEPC	diethyl pyrocarbonate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNP	2,4-dinitrophenylhydrazine
2D-PAGE	two dimensional polyacrylamide gel electrophoresis
DTT	dithiotreitol
ECL	enhanced chemiluminescence
EDTA	ethylenediamine tetracetic acid
EGTA	ethyleneglycol-bis(B-aminoethylether)N,N,N',N'-tetracetic acid
g (mg, μg, ng)	gram (miligram, microgram, nanogram)
GTP	guanosine 5'-triphosphate
H <sub>2</sub> O <sub>2</sub>	hydrogen peroxide
HEPES	N-[-2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
IEF	isoelectric focusing
IPG	immobilized pH gradient
kDa	kilodalton
l (ml, μl)	liter (mililiter, microliter)
MCA	4-methylcoumaryl-7-amide

mQ	milliQ
mRNA	messenger ribonucleic acid
nm	nanometers
NP-40	nonylphenyl-polyethylene glycol
OD <sub>600</sub>	optical density at 600 nm
PMSF	phenylmethylsulfonyl fluoride
RNA	ribonucleic acid
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
Suc-LLVY-MCA	succinyl-leucyl-leucyl-valyl-tyrosine 4-methylcoumaryl-7-amide
TCA	trichloroacetic acid
TEMED	N,N,N,'N,'-tetramethylethylenediamine
TEV	tobacco etch virus
TFA	trifluoroacetic acid
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
tRNA	transfer ribonucleic acid
U	units
V	Volt
W	Watt

Other abbreviations used are explained where appropriate in text.

# **I. Introduction**

1. The standard genetic code
2. Evolution of the genetic code
3. The *Candida* spp. genetic code



## 1. The standard genetic code

Genetic information is stored in DNA as a sequence of letters that contain the guidelines required to produce all proteins of a cell. This code forms a genetic language whose alphabet contains only four letters, the nucleotides adenosine (A), thymidine (T), guanosine (G) and cytidine (C) that are arranged in three-letter words designated as codons. In order to decode the information contained on the DNA sequence it has to be copied to the messenger RNA (mRNA), whose language is similar to the DNA but uses uridine (U) instead of T. The combination of the 4 ribonucleotides (A, U, G, C) into triplets results in the 64 different codons that make the genetic code (Table 1). From those 64 codons, 61 triplets specify the 20 amino acids and the additional 3 specify termination codons. Some amino acids are specified by only one codon, namely Methionine (Met) and Tryptophan (Trp) that are decoded by the AUG and UGG codons, respectively, but the remaining amino acids have more than one codon, and thus the genetic code is degenerate. The codons that are assigned to the same amino acid are termed “synonymous” and in most cases share the first two nucleotides. Therefore, a group of four synonymous codons is a “family box” or a “four-codon box”, and a group of two synonymous codons is a “two-codon set” (Table 1). Amino acids that are encoded by family boxes are Alanine (Ala), Glycine (Gly), Proline (Pro), Threonine (Thr) and Valine (Val), while Asparagine (Asn), Aspartate (Asp), Cysteine (Cys), Glutamine (Gln), Glutamate (Glu), Histidine (His), Lysine (Lys), Phenylalanine (Phe) and Tyrosine (Tyr) are encoded by two-codon sets. Exceptions are found for Arginine (Arg), Leucine (Leu) and Serine (Ser), which have six synonymous codons consisting of a family box and a two-codon set. Isoleucine (Ile) has three synonymous codons, and the remaining three codons (UAG, UAA and UGA) are the signal for termination of protein synthesis (Table 1).

**Table 1 - The standard genetic code.** The 64 codons are shown, divided into codon boxes that are defined by the first two nucleotides. There are 61 codons corresponding to the 20 amino acids and 3 codons for translation termination (adapted from Agris, 2004).

<b>UUU</b>	<b>Phe</b>	<b>UCU</b>	<b>Ser</b>	<b>UAU</b>	<b>Tyr</b>	<b>UGU</b>	<b>Cys</b>
<b>UUC</b>	<b>Phe</b>	<b>UCC</b>	<b>Ser</b>	<b>UAC</b>	<b>Tyr</b>	<b>UGC</b>	<b>Cys</b>
<b>UUA</b>	<b>Leu</b>	<b>UCA</b>	<b>Ser</b>	<b>UAA</b>	<b>STOP</b>	<b>UGA</b>	<b>STOP</b>
<b>UUG</b>	<b>Leu</b>	<b>UCG</b>	<b>Ser</b>	<b>UAG</b>	<b>STOP</b>	<b>UGG</b>	<b>Trp</b>
<b>CUU</b>	<b>Leu</b>	<b>CCU</b>	<b>Pro</b>	<b>CAU</b>	<b>His</b>	<b>CGU</b>	<b>Arg</b>
<b>CUC</b>	<b>Leu</b>	<b>CCC</b>	<b>Pro</b>	<b>CAC</b>	<b>His</b>	<b>CGC</b>	<b>Arg</b>
<b>CUA</b>	<b>Leu</b>	<b>CCA</b>	<b>Pro</b>	<b>CAA</b>	<b>Gln</b>	<b>CGA</b>	<b>Arg</b>
<b>CUG</b>	<b>Leu</b>	<b>CCG</b>	<b>Pro</b>	<b>CAG</b>	<b>Gln</b>	<b>CGG</b>	<b>Arg</b>
<b>AUU</b>	<b>Ile</b>	<b>ACU</b>	<b>Thr</b>	<b>AAU</b>	<b>Asn</b>	<b>AGU</b>	<b>Ser</b>
<b>AUC</b>	<b>Ile</b>	<b>ACC</b>	<b>Thr</b>	<b>AAC</b>	<b>Asn</b>	<b>AGC</b>	<b>Ser</b>
<b>AUA</b>	<b>Ile</b>	<b>ACA</b>	<b>Thr</b>	<b>AAA</b>	<b>Lys</b>	<b>AGA</b>	<b>Arg</b>
<b>AUG</b>	<b>Met</b>	<b>ACG</b>	<b>Thr</b>	<b>AAG</b>	<b>Lys</b>	<b>AGG</b>	<b>Arg</b>
<b>GUU</b>	<b>Val</b>	<b>GCU</b>	<b>Ala</b>	<b>GAU</b>	<b>Asp</b>	<b>GGU</b>	<b>Gly</b>
<b>GUC</b>	<b>Val</b>	<b>GCC</b>	<b>Ala</b>	<b>GAC</b>	<b>Asp</b>	<b>GGC</b>	<b>Gly</b>
<b>GUA</b>	<b>Val</b>	<b>GCA</b>	<b>Ala</b>	<b>GAA</b>	<b>Glu</b>	<b>GGA</b>	<b>Gly</b>
<b>GUG</b>	<b>Val</b>	<b>GCG</b>	<b>Ala</b>	<b>GAG</b>	<b>Glu</b>	<b>GGG</b>	<b>Gly</b>

The genetic code is also non-universal, since although most organisms use a standard genetic code, some organisms and eukaryotic organelles use genetic codes that differ slightly from the norm.

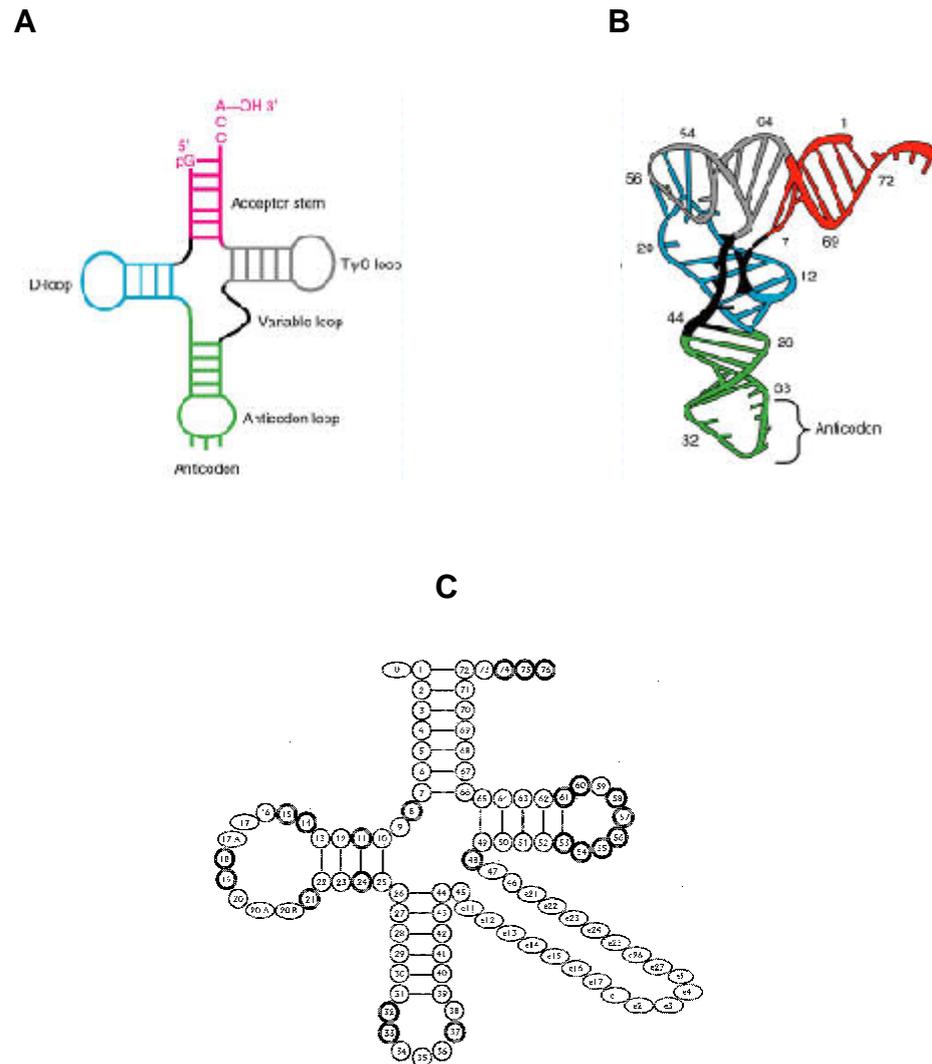
Decoding the genetic information contained in the mRNA into protein occurs on the ribosome, during the process of translation. This is achieved by pairing each amino acid-specific codon from the mRNA with a complementary sequence, the anticodon triplet of the transfer RNA (tRNA). The tRNA thus connects the worlds of nucleic acids and proteins, as it binds a specific amino acid corresponding to the anticodon. The enzymes that charge the tRNAs with the respective amino acids are the aminoacyl-tRNA synthetases (aaRS). Fidelity of translation relies on the accuracy of codon-anticodon recognition and tRNA aminoacylation and, therefore, maintenance of the genetic code is assured by tRNAs, aaRSs, translation factors and the ribosome, which will be discussed in the next chapters.

## **1.1. Transfer RNAs**

In 1958 Crick's "adaptor hypothesis" predicted the existence of a factor linking codons and amino acids, and RNA could be such an adaptor as it is possible to envisage a base-pairing mechanism to read the codons while carrying the correct amino acids (Crick, 1958). Transfer RNA (tRNA) is indeed the molecule that performs these tasks in the cell, having the dual role of base-pairing with the codon of the mRNA complementary to its anticodon sequence and being aminoacylated with the correct amino acid.

The tRNAs can be subdivided into families of isoacceptors, defined by the corresponding or cognate amino acid they specify, and each family is normally recognized by a single aminoacyl-tRNA synthetase (aaRS) that charges the individual isoacceptor tRNAs with the correct amino acid. The function of the tRNAs is ultimately to deliver the correct amino acid for incorporation into the nascent peptide chain in response to the cognate codon of the mRNA. Aminoacylated tRNA forms a complex with the elongation factor 1A (EF-Tu in prokaryotes and eEF1A in eukaryotes) and GTP. The function of this ternary complex is to carry the aminoacyl-tRNA to the A site of the ribosome, where GTP is hydrolyzed and a peptide bond is formed.

The typical tRNA cloverleaf secondary structure consists of three arms composed of a stem and a loop, a variable region and the acceptor stem with the 3' single-stranded CCA end (Figure 1A). The amino acid is attached to the ribose of the A residue on the 3' end. The tRNA arms are the D arm, named after the presence of dihydrouridine residues, the anticodon arm, where the triplet complementary to the codon is located, and the T $\psi$ C arm, whose name derives from the modified nucleosides ribothymidine and pseudouridine. The number of residues in the stem and loop regions is conserved and can therefore be referenced by a standard number (Figure 1C).

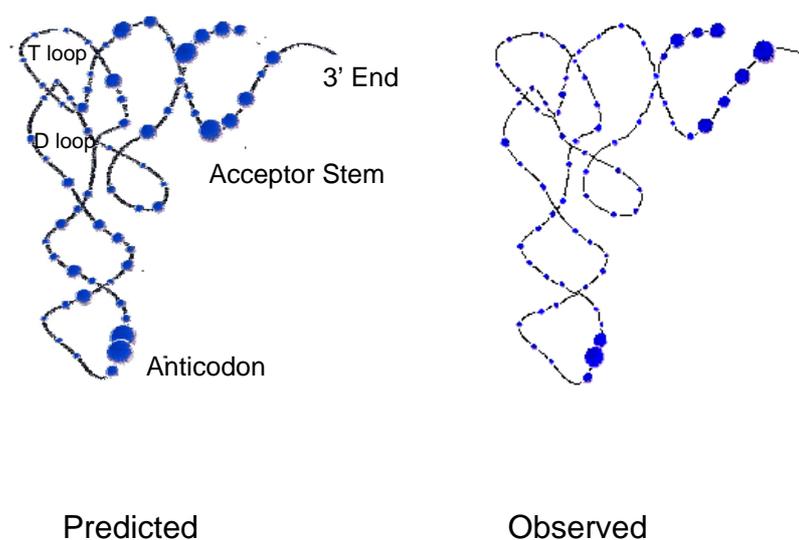


**Figure 1 - tRNA secondary and tertiary structures.** A – The cloverleaf structure of a tRNA, indicating in colours the different arms. B – The three-dimensional folding of a tRNA. The colours correspond to the domains indicated in A (adapted from Rich and Kim, 1978). C – Numbering of nucleotides in tRNAs. Circles correspond to nucleotides present in all tRNAs, ovals represent nucleotides that are not always present, namely the nucleotides in the variable loop (Sprinzl *et al.*, 1998).

There are two families of tRNAs according to the length of the variable region. Class I tRNAs have short variable loops of four or five nucleosides, while those of class II have long variable arms of 10 to 24 bases. Class II is formed by leucine and serine tRNAs in eukaryotes, and in eubacteria by leucine, serine and tyrosine tRNAs (Dirheimer *et al.*, 1995).

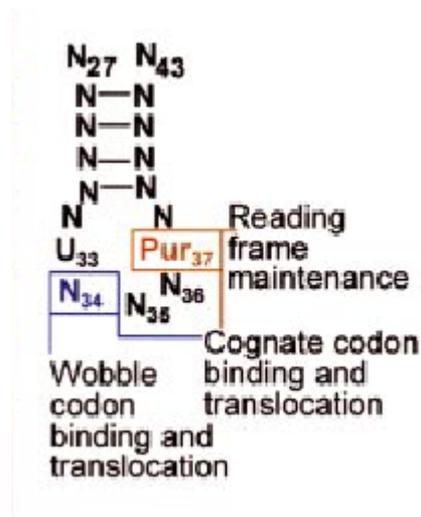
The tertiary three-dimensional structure of a tRNA is an L-shaped molecule, with the D-arm stacked onto the anticodon-arm and the T $\psi$ C-arm stacked onto the acceptor-stem, defining two functional domains (Figure 1B). The site that interacts with the mRNA template and the amino acid attachment site are at opposite ends of the tRNA. These distinct domains also bind different parts of the aaRSs (chapter I.1.2.), and there is evidence that the T $\psi$ C-acceptor minihelix can function as an independent unit since, when charged, this structure behaves like a charged tRNA in its ability to interact with the elongation factor (Schimmel and Ribas de Pouplana, 1995). It is, therefore, possible that the two domains had independent origins, with the T $\psi$ C-acceptor minihelix evolving before the emergence of the anticodon domain. Thus, in its origin, the genetic code might have been based on aminoacylation of T $\psi$ C-acceptor minihelix-like molecules by ribozymes. The posterior evolution of the anticodon domain originated tRNA-like molecules, and the respective anticodon-binding domain was acquired by synthetases (Schimmel and Ribas de Pouplana, 1995; Ribas de Pouplana and Schimmel, 2001b) (chapter I.1.2.).

The features of a tRNA that determine which synthetase will recognize it are designated as identity determinants (Figure 2). These elements located in the acceptor region consist of the three first base pairs (1:72, 2:71 and 3:70) and the discriminator base, the unpaired position 73. tRNAs for chemically similar amino acids are likely to have the same nucleoside at position 73, and, thus, this position potentially identifies present and even past tRNA forms specific for related amino acids (Crothers *et al.*, 1972; McClain, 1993). The anticodon is also important for the recognition of most tRNA families. Besides the positive signals for tRNA and cognate synthetase efficient interaction, there are also negative determinants that block recognition by non-cognate aaRSs (Pallanck *et al.*, 1995).



**Figure 2 – Comparison between computer-predicted and experimentally observed identity determinants of *Escherichia coli* tRNAs.** Each blue circle corresponds to a nucleoside position, and the diameter is proportional to the fraction of the 20 tRNA acceptors for which the position was a predicted or observed determinant (adapted from Meinnel *et al.*, 1995).

tRNAs contain modified nucleosides that influence the structure of the molecule and the efficiency of translation, due to their direct and indirect involvement in codon recognition and action as determinants of cognate acylation (Yokoyama and Nishimura, 1995; Björk, 1995b; Agris, 2004). These nucleosides derive from the normal adenosine (A), guanosine (G), cytidine (C) and uridine (U) that are modified after the synthesis of the tRNA, with the exception of queuosine (Q). The biggest diversity of modifications is found at positions 34 and 37 in the anticodon-loop of the tRNA (Figure 3) (reviewed by Agris, 2004).



**Figure 3 – Nucleoside modifications in the anticodon-arm of a tRNA.** Several modifications at the wobble position 34 and purine 37 of the anticodon-loop of the tRNA are important for decoding and reading frame maintenance (adapted from Agris, 2004).

Modified bases at the wobble position (34) alter the decoding properties of the tRNA. Unmodified uridine is able to base pair with all four nucleosides, although C is recognized with less efficiency. Therefore, the presence of modified U derivatives at the wobble position functions either to extend or restrict the decoding properties of the tRNAs. Inosine (I) is a purine formed by the

deamination of A and is common in the anticodon loop of eukaryotic tRNAs, at nucleoside 34, and can base pair with A, U and C. Queuosine (Q) is a hypermodified nucleoside inserted in substitution of G after its excision from the ribophosphate backbone and, with the exception of yeast, Q is found at position 34 of tRNAs that read NAU or NAC codons, where N is any of the four bases (Yokoyama and Nishimura, 1995).

The often extensively modified nucleosides at position 37 are thought to have evolved to strengthen the base pairing between the last base of the anticodon (position 36) and the first base of the codon (Björk, 1995a). Position 37 has frequently hydrophobic nucleosides such as  $i^6A$  or its derivatives, namely in tRNAs that read codons starting with U, both to improve A36-U interaction and to prevent base pairing of A36 with other bases. The isopentenyl group ( $i^6$ ) is produced from mevalonic acid, a precursor to the synthesis of several products such as haem A, ubiquinone or cholesterol, underlying the connection between tRNA modification and metabolism (Björk, 1995a).

The most conserved modifications found in position 37 are the  $m^1G$ , which is important to maintain the translational reading frame, and the  $t^6A$  modification, which promotes cognate codon binding (Agris, 2004). These observations prompt the hypothesis that both modifications were present early in evolution and even in the tRNA of the last common ancestor (Björk, 1995a). Several positions outside the anticodon arm have simpler modifications, like methylated or thiolated derivatives that might induce structural alterations and hence also affect decoding indirectly.

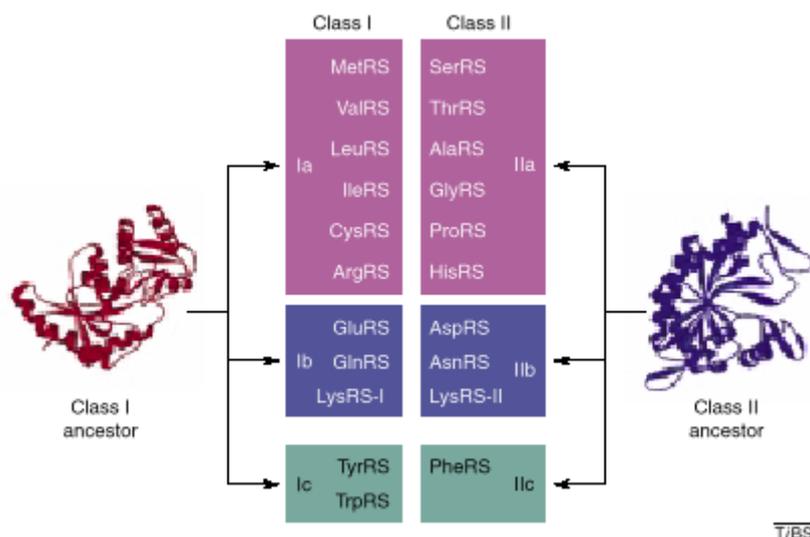
Modified nucleosides on the anticodon can modulate the recognition of the tRNA by the cognate aminoacyl-tRNA synthetase. Modifications at other positions may also act as identity determinants or antideterminants, and conformational changes on the tRNA induced by modified bases can play an indirect role in the aminoacylation process. For example, yeast has two isoleucine isoacceptors with the modified nucleosides I and  $\Psi$  at the wobble position (nucleotide 34), which

might be important for IleRS recognition. In *E. coli* tRNA<sup>Ile</sup>, the presence of lysidine (k<sup>2</sup>C) instead of C34 alters both codon reading, as it base pairs with A rather than G, and prevents misacylation with methionine. This suggests that, during evolution, a post-transcriptional modification on a Met-tRNA decoding the AUG codon might have converted it into an Ile-tRNA that reads the AUA codon instead (Muramatsu *et al.*, 1988).

Since modified nucleosides derive from compounds involved in intermediary metabolism, tRNA modification is sensitive to metabolic stress conditions. For instance, lack of methionine or cysteine results in deficient methylation or thiolation, and oxygen or iron limitation also alter tRNA modification (Björk, 1995a). Taking into account the role of modified nucleosides in the decoding properties of the tRNAs, mutations and stress conditions that might originate undermodified tRNAs will affect the accuracy of translation. For example, the lack of a modified uridine, due to a mutation on the mitochondrial leu-tRNAs<sub>UUR</sub>, resulted in mistranslation of leucine codons with consequent mitochondrial dysfunction (Yasukawa *et al.*, 2000) (chapter I.1.4.). Mutation of the mitochondrial leu-tRNAs<sub>UUR</sub> was also associated with diabetes (Suzuki *et al.*, 2005).

## 1.2. Aminoacyl-tRNA synthetases

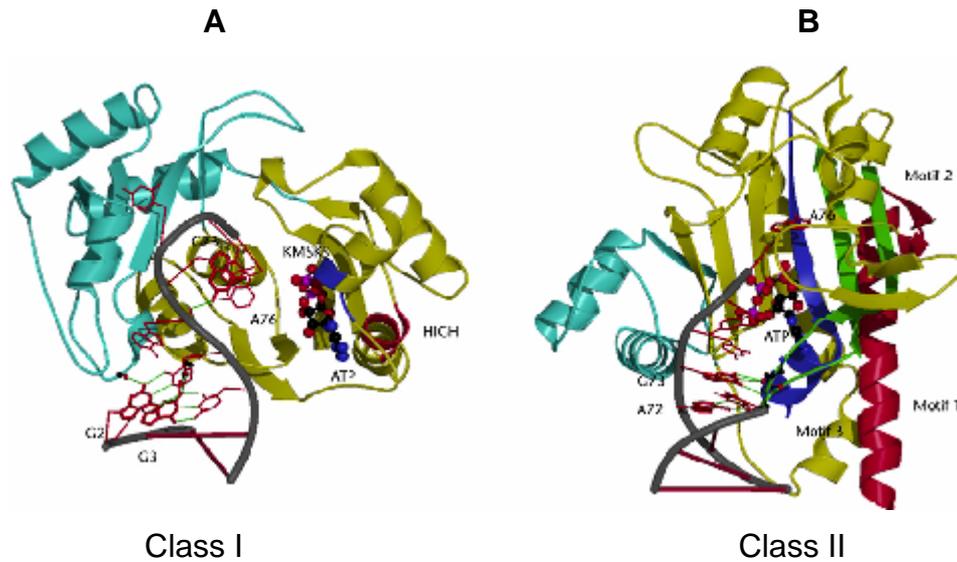
The aminoacyl-tRNA synthetases (aaRSs) establish the genetic code by charging tRNAs with the cognate amino acids. There are two classes of aaRSs that were originated from two distinct single-domain proteins (Schimmel *et al.*, 1993), and members of each class share common sequence motifs that distinguish the active site structures (Figure 4). Class I enzymes have a nucleotide binding domain, the Rossmann fold, characterized by the HIGH and KMSKS motifs and includes ArgRS, CysRS, GluRS, GlnRS, IleRS, LeuRS, MetRS, TrpRS, TyrRS and ValRS. Class II synthetases contain an antiparallel  $\beta$ -sheet domain with three degenerate sequence motifs, a N-terminal motif 1, a central motif 2 and a C-terminal motif 3, and AlaRS, AsnRS, AspRS, GlyRS, HisRS, LysRS, PheRS, ProRS, SerRS and ThrRS belong to this group (Meinzel *et al.*, 1995; Woese *et al.*, 2000).



**Figure 4 – Subclasses of aminoacyl-tRNA synthetases.** The two main subclasses of aaRSs have evolved from two independent ancestors, as represented in this diagram (Ribas de Pouplana and Schimmel, 2001a).

Both classes of synthetases can be further divided into three subgroups and the enzymes in each subclass generally recognize chemically or sterically similar amino acids (Figure 4). For example, synthetases from class Ic and IIc recognize the aromatic amino acids, whereas class Ib and IIb enzymes recognize charged amino acids and their derivatives (Ribas de Pouplana and Schimmel, 2001a). The fact that class I enzymes systematically aminoacylate the 2'-OH group of the ribose from the terminal adenosine of the tRNA while class II aminoacylate the 3'-OH group supports the hypothesis of a common ancestor for the synthetases of the same class (Meinzel *et al.*, 1995; Woese *et al.*, 2000). Furthermore, the two synthetase classes also differ in the binding of the tRNA, as class I enzymes approach the minor groove of the acceptor arm and bind the tRNA on the side opposite to the variable loop, whereas class II aaRSs interact with the major groove side of the acceptor arm of the tRNA and bind on the variable loop side (Ruff *et al.*, 1991) (Figure 5).

The existence of two distinct classes of synthetases favours the hypothesis of evolution of the tRNA aminoacylation mechanism at least twice. There is evidence to propose that aaRS families evolved independently and were subject to horizontal gene transfer, mainly due to their universal function (reviewed in Woese *et al.*, 2000). But, what are the implications of the existence of two different classes of aaRSs to the evolution and establishment of the modern genetic code? The symmetry between the two classes has been interpreted as the result of interaction of synthetase pairs with tRNA, during evolution (Ribas de Pouplana and Schimmel, 2001a). According to this model, the primitive genetic code would have encoded a reduced number of amino acids, and tRNAs were recognized by synthetase pairs. During expansion of the genetic code, the incorporation of new amino acids was achieved by duplication of tRNAs and synthetases. To discriminate between amino acids, aaRS pairs were split and each one have evolved to recognize distinct, but related, amino acids and tRNAs specific for a subset of codons from the original set. This theory receives further support from the observation of patterns correlating codons to synthetase subclasses (Ribas de Pouplana and Schimmel, 2001a).



**Figure 5 – AaRSs complexed with the acceptor arms of their cognate tRNAs.** A – Class I GluRS, showing in yellow the Rossman fold with the characteristic motifs HIGH and KMSKS highlighted in red and dark blue, respectively. B – Class II AspRS, showing the characteristic motifs 1, 2 and 3 highlighted in red, green and dark blue, respectively (adapted from Arnez and Moras, 1997).

To accomplish their two main functions, activation of the amino acid and recognition of the tRNA, aaRSs are organized into distinct specialized domains, which may also reflect their gradual evolution. Two distinct ancestors, corresponding to different aminoacylation mechanisms, could have acquired additional domains capable of binding tRNA-like molecules. This increased complexity allowed the development of specificity and accuracy to achieve better regulation of the translational process (Delarue and Moras, 1993). The ancestry of aaRSs also supports that evolution of aaRSs was a critical step on the evolution of the modern genetic code (Szathmáry, 1999; Woese *et al.*, 2000; Lee *et al.*, 2000).

AaRSs are highly selective for their amino acid and tRNA substrates, and in most cases directly acylate the tRNAs in an ATP-dependent two-step reaction. First ATP and the amino acid bind to the active site of the aaRS to form an aminoacyl-

adenylate, and then the amino acid is attached to the tRNA by 3'-esterification (reviewed by Ibba and Soll, 2000).

Specificity of aaRSs implicates that these enzymes must be able to discriminate between competing substrates, i.e., similar amino acids. The editing activity of some synthetases prevents the incorporation of noncognate amino acids by a mechanism of proofreading that proceeds either by hydrolysis of the misactivated amino acid or by deacylation of the mischarged tRNA. For example, ValRS needs to distinguish between valine and threonine and IleRS has to recognize isoleucine and discriminate against the similar amino acid valine (Meinzel *et al.*, 1995).

The presence of identity determinants and antideterminants is important for cognate tRNA selection, as all tRNAs are structurally similar. Recognition of the correct tRNA from the cellular pool of tRNAs by a particular synthetase depends on the functional interaction of the aaRS with specific sites of the tRNA, that for most aaRSs consist of the acceptor stem that includes the discriminator base (N73), the three bases of the anticodon and the D loop. Interactions with other regions of the tRNA, such as the extra arm, and particular structural elements unique to specific tRNAs may also have a role in cognate aminoacylation (Figure 2 on previous chapter).

AaRSs contribute to other cellular functions in addition to protein synthesis, namely quality control of translation, maturation of tRNA precursors and porphyrin biosynthesis. Besides regulating the expression of their own genes, some aaRS have also been implicated in translational regulation, amino acid metabolism and intron splicing (Meinzel *et al.*, 1995; Ibba and Soll, 2000). Some of these functions are not dependent on aminoacylation activity, but instead rely on the ability of aaRSs to bind RNA molecules. Therefore, the roles of aaRSs beyond maintaining the fidelity of mRNA translation place them as important regulators of cellular function by coupling translation and metabolism.

### 1.3. The genetic code and the origin of life

The origin of the genetic code and the translational process meant a major breakthrough in evolution, as they allowed task division between proteins and nucleic acids on a world previously dominated by RNA (Szathmary, 1999).

On the RNA world, metabolic reactions would have been performed by RNA enzymes or ribozymes. There is evidence that some ribozymes have aminoacyl transferase activity and are able to catalyze the formation of amide bonds (Wiegand *et al.*, 1997). Furthermore, extant rRNA participates in peptidyl transfer reactions during translation (Nitta *et al.*, 1998) establishing the ribosome as a ribozyme (Nissen *et al.*, 2000), and some coenzymes have nucleotide-like moieties, possibly remains from the RNA world. The fact that several enzymes use the same coenzyme points to an evolutionary conservation of coenzyme chemical structure. Additionally, since RNA can selectively bind coenzymes and ribozymes can use cofactors in their activity, it is possible that amino acids were used as cofactors in the RNA world (Szathmary, 1999). The attachment of amino acids might have been accomplished through their linkage to oligonucleotides that could reversibly basepair with the ribozymes, thus assigning amino acids to specific sequences that evolved into tRNAs. The ribozymes that catalyzed these reactions evolved into aminoacyl-tRNA synthetases in a coding mechanism that preceded translation, while other ribozymes might have lost their enzymatic activity turning into mRNA molecules. Aminoacyl-tRNA synthetase ribozymes were, thus, fundamental to the evolution of the genetic code and the process of translation (Lee *et al.*, 2000), and the increased versatility of the protein world, with enhanced properties and novel chemistries of amino acids over nucleic acids, contributed to their selection.

The genetic code might have arisen with a limited number of amino acids, and the others were added later. The addition of new amino acids, with novel and distinct chemical properties, introduced a significant selective advantage. Thus, this would have occurred in a progenitor lineage that displaced all other existing codes

(Weiner and Maizels, 1987). In alternative to this view of a common genetic code in the ancestor, the modern genetic code might be a reductive instead of an expanded code. According to this hypothesis, there was a gradual replacement of rare amino acids with more abundant ones, based on the observation that extant proteinaceous amino acids have more than one codon. The genetic code would have evolved towards keeping functional or structural amino acids, such as selenocysteine and pyrrolysine (chapter 1.2.1.2.) that are chemically active residues (Fenske *et al.*, 2003).

In the primitive genetic code, it is likely that a single codon specified more than one amino acid with similar properties. Early proteins would, therefore, consist of a heterogeneous mixture of closely related peptide sequences, with some species more active than others that were favoured by selection. Cells with ambiguous codes producing statistical proteomes could have been important during the early evolution of organisms, namely under conditions of amino acid limitation due to the possibility of using available alternatives (Pezo *et al.*, 2004; Hendrickson *et al.*, 2004). The existence of ambiguous codes in extant organisms is considered by some authors as a reminiscence of these ancestral codes, which have evolved to eliminate ambiguity (Nangle *et al.*, 2002). Alternatively, genetic code alterations could have evolved from the standard code and not from primitive lineages that existed in the RNA world (Knight *et al.*, 2001). Thus, these ambiguous codes might be functional and subjected to positive selection (Santos *et al.*, 1996; Santos *et al.*, 1997; Santos *et al.*, 1999; this work).

#### 1.4. Mistranslation

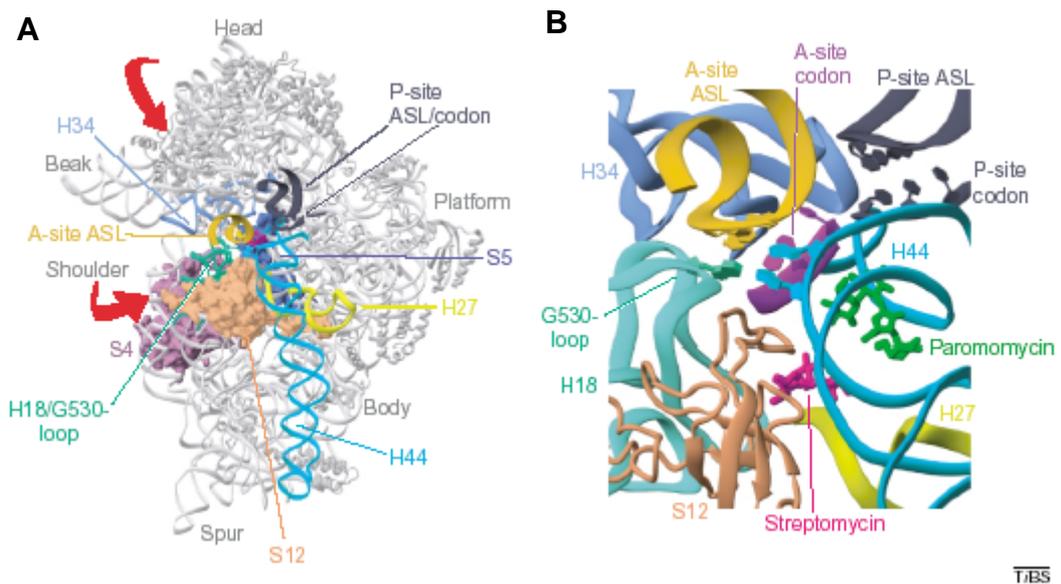
Fidelity of the translational process assures the production of a stable and functional proteome, hence the translational machinery has to monitor every step of protein synthesis (reviewed by Valente and Kinzy, 2003). In normal conditions, mistranslation errors, which result from the insertion of the wrong amino acid at a given codon, are in the order of 1 in  $10^{-4}$  (Edelmann and Gallant, 1977). However, this error frequency increases under stress conditions, namely during amino acid starvation (Parker and Precup, 1986). Therefore, mistranslation events that result in the formation of aberrant proteins are often linked to disease. For example, mutations in mitochondrial tRNAs are correlated with severe neuromuscular and neurodegenerative pathologies, namely MELAS and MERRF, which are mitochondrial encephalomyopathies (Yasukawa *et al.*, 2000; Rabilloud *et al.*, 2002). Mitochondrial dysfunction can be originated by mutation in the mitochondrial leu-tRNAs<sub>UUR</sub>, resulting in the absence of uridine modification at the first position of the anticodon with consequent mistranslation of leucine codons (Yasukawa *et al.*, 2000). More recently, a mitochondrial tRNA mutation has been associated to hypertension and dyslipidemia, which are risk factors for cardiovascular diseases such as infarction or stroke (Wilson *et al.*, 2004). In this study, the mutation was found at position 33 of the tRNA, where the replacement of the conserved uridine by cytidine impairs ribosome binding and causes lipid and glycogen accumulation, among other metabolic disturbances (Wilson *et al.*, 2004). Misreading has also been implicated in Alzheimer's disease and other age-related disorders, namely due to the production of an out-of-frame ubiquitin, which is not able to tag proteins for degradation, accumulates and blocks the proteasome (reviewed by van Leeuwen *et al.*, 2002).

The errors that might occur during translation include frameshifting, due to reading frame alteration, nonsense errors originated by premature termination or failure in stop codon recognition, and missense errors resulting from non-cognate amino acid insertion, either by misacylation of the tRNA by the aaRS or incorrect interaction between codon-anticodon on the ribosome (Stansfield *et al.*, 1998).

Indeed, the ribosome plays an important role in maintaining the fidelity of translation, by selecting the cognate aminoacyl-tRNA (aa-tRNA), in the form of a ternary complex with the elongation factor 1A and GTP, from the pool of ternary complexes available. Besides the ribosome, elongation factors are also central molecules in the decoding process. Elongation factor 1 brings the aa-tRNA complexed with GTP to the ribosomal A site, where it signals correct codon-anticodon interaction through GTP hydrolysis, and elongation factor 2 contributes to the maintenance of the proper reading frame by controlling translocation of the peptidyl-tRNA to the ribosomal P-site (reviewed in Kapp and Lorsch, 2004).

The simultaneous speed (4 to 8 amino acids per second) and accuracy ( $10^{-4}$  errors per codon decoded) of translation indicates that the ribosome has a proofreading system for tRNA discrimination (Thompson and Stone, 1977; Ruusala *et al.*, 1982). In fact, there are two steps in aa-tRNA selection by the ribosome that rely on the stability of codon-anticodon interaction. In the initial selection, an A-site incoming aa-tRNA is tested for codon-anticodon pairing, and ternary complexes with very dissimilar anticodons dissociate rapidly without GTP hydrolysis (Pape *et al.*, 1999). However, this step is insufficient to reject near-cognate aa-tRNAs that match the codon, and an additional proofreading step is required. During proofreading, correct codon-anticodon interaction stabilizes the ternary complex, slowing down the dissociation rate and activating GTP hydrolysis, in an induced-fit mechanism (Rodnina and Wintermeyer, 2001a; Rodnina and Wintermeyer, 2001b). Accordingly, binding of the cognate aa-tRNA induces movement of the ribosomal 30S subunit domains and conformational changes of the conserved essential bases A1492, A1493 and G530 of the 16S ribosomal RNA (rRNA), which interact with the first two codon-anticodon base pairs (Figure 6) (Ogle *et al.*, 2001). Moreover, cognate aa-tRNA binding triggers a transition from an open to a closed form in the ribosomal 30S subunit, required for the subsequent steps of selection (Ogle *et al.*, 2002). The induced GTP hydrolysis results in a conformational change of the elongation factor bound to GDP, which dissociates from the ribosome and allows the cognate aa-tRNA to be accommodated on the ribosomal 50S subunit, and the peptidyl transferase reaction takes place (reviewed by Ogle

*et al.*, 2003). Peptide bond formation is catalysed by the 23S rRNA, on the peptidyl transferase centre of the large ribosome subunit, making the ribosome a ribozyme (Nissen *et al.*, 2000). The fact that the accuracy of translation is dependent on RNA-RNA interactions is also consistent with the development of translation in an RNA world (Yarus and Smith, 1995) (chapter I.1.3.).



**Figure 6 – Induced fit of the ribosomal 30S subunit around the decoding centre.** A – Overview of the 30S subunit structure. Red arrows indicate the domain movement during the transition to the closed conformation. B – Close-up of selected 30S elements around the decoding centre, showing the A-site codon (in purple) interacting with the nucleotides G530 (in turquoise), A1492 and A1493 (in cyan) of the 16S RNA (from Ogle *et al.*, 2003).

The direct involvement of the ribosome in the recognition of the cognate aa-tRNA hints on why antibiotics, such as streptomycin or paromomycin, increase translation error rate. For example, paromomycin induces the ribosome to switch to the closed form even when a near-cognate aa-tRNA is present, promoting mistranslation (Ogle *et al.*, 2003).

However, as stressed above, errors in translation may also arise during termination. Again, ribosomal RNA and translation factors assure the efficiency of protein synthesis termination. Translation termination occurs when a stop codon is present in the ribosomal A site, resulting in the release of the synthesized polypeptide. This reaction is catalyzed by the ribosome in response to the class I release factor (eRF1) that mimics tRNA structure to perform the task of nonsense codon recognition (reviewed by Poole and Tate, 2000). Several mutations of the rRNA have been identified, both in the 16S and 23S rRNAs from *E. coli* and in yeast 18S rRNA that either enhance stop codon read-through, or counteract suppressor mutations in the release factors. Thus, such mutations can impair or restore translation termination (Kapp and Lorsch, 2004). Furthermore, mutations of the residues 32 and 123 of eRF1 result in loss of stop codon recognition, suggesting that these positions are crucial for termination (Lehman, 2001). Specific mutations at these positions determine which stop codons are recognized by eRF1, as shown in organisms that have genetic code alterations involving stop codons (Lozupone *et al.*, 2001) (next chapter). Therefore, eRF1 mutations have implications not only for translation termination but also for the evolution of the genetic code.

## 2. Evolution of the genetic code

The genetic code was once thought to be universal and immutable. According to the frozen accident theory, any alteration to the genetic code would be detrimental or lethal to the organism (Crick, 1968). For this reason, the discovery of alternative genetic codes was interpreted as aberrations of nature, without evolutionary consequences (Osawa and Jukes, 1989). However, recent studies indicate that some genetic code alterations are not neutral but rather represent novel mechanisms for the fast evolution of new phenotypes (Santos *et al.*, 1999; this work). These studies underline the physiological and evolutionary consequences for the organisms that redefine codon identity, supporting the hypothesis that the genetic code has intrinsic flexibility and is still evolving (Yokobori *et al.*, 2001; Knight *et al.*, 2001).

Genetic code alterations involve mainly stop codons (chapter I.2.1.3.), suggesting that release factors and in particular eRF1 had a major role in the evolution of the genetic code (Figure 7). The residues 32 and 123 of eRF1 are determinant for carrying out termination functions, as mutations in these positions promote loss of stop codon recognition (Lehman, 2001). Moreover, the presence of specific residues at these positions is associated with the acquired new meaning of the reassigned termination codon. Indeed, the replacement of the canonical leucine at position 123 of eRF1 by a phenylalanine permits UAR (R = purine) decoding as glutamine (Lozupone *et al.*, 2001).

**Figure 7 – The genetic code 2005.** The standard genetic code is shown in black and in colour the genetic code alterations. Green, alterations found in mitochondria; blue, alterations in prokaryotic cytoplasmic and eukaryotic nuclear systems. The recently added amino acids selenocysteine and pyrrolysine are also shown. Red highlights the reassignment of the leucine-CUG codon to serine in *Candida* species, which will be described in detail in chapter I.3.



Some codon sets are more prone to change than others (chapter 1.2.1.3.), for instance codons starting with U or A often reassign their identity. Conversely, alterations in codons with G in the first position are not known, suggesting that the strength of codon-anticodon interaction is important for the evolution of non-standard genetic codes. Accordingly, codons starting with C are rarely changed, with the only known examples being the reassignment of the leucine CUN codon family to threonine in yeast mitochondria (Pape *et al.*, 1985) and the CUG decoding as serine in *Candida* species (Santos and Tuite, 1995).

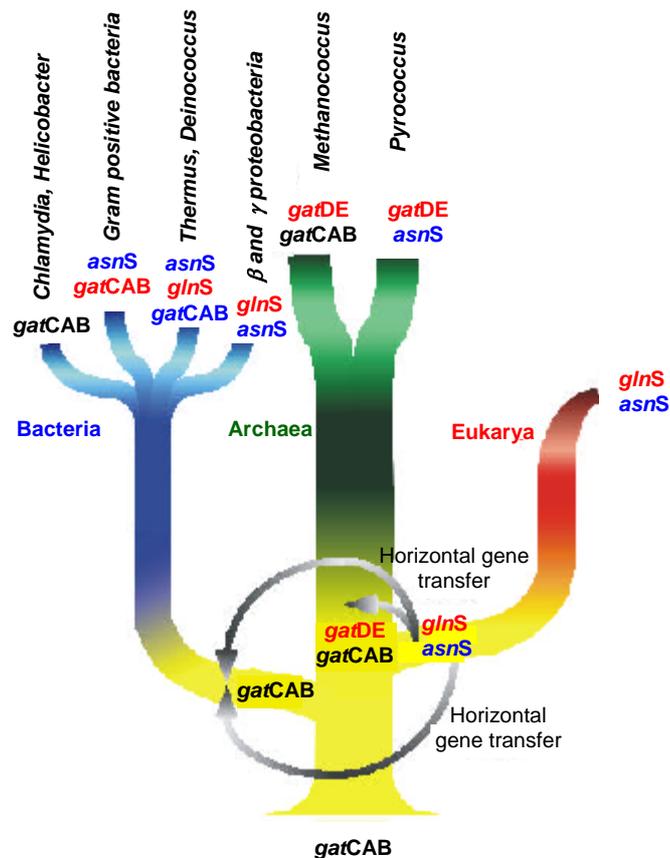
## **2.1. Special decoding events**

### **2.1.1. Asparagine, glutamine, and initiator methionine**

The discovery of an alternative route for aminoacyl-tRNA synthesis, such as the indirect synthesis of asparaginyl-, glutaminyl-, and formylmethionyl-tRNAs from aminoacyl-tRNA precursors (reviewed in Ibba and Soll, 2000; Praetorius-Ibba and Ibba, 2003), has provided important insight on the evolution of aaRSs and the modern genetic code (Ruan *et al.*, 2001).

The asparaginyl-tRNA synthetase (AsnRS) and glutaminyl-tRNA synthetase (GlnRS) directly synthesize Asn-tRNA and Gln-tRNA in the cytoplasm of eukaryotes and some eubacteria. However, gram-positive bacteria, eukaryotic organelles and archaea lack the GlnRS, and the AsnRS is also absent in some archaea (Praetorius-Ibba and Ibba, 2003). In these cases asparagine and glutamine are synthesized in a tRNA-dependent transamidation pathway, which is the most common divergence from the canonical aminoacyl-tRNA synthesis (Ruan *et al.*, 2001). Asn-tRNA<sup>Asn</sup> and Gln-tRNA<sup>Gln</sup> are formed from tRNAs misacylated with aspartic (Asp-tRNA<sup>Asn</sup>) and glutamic acid (Glu-tRNA<sup>Gln</sup>) that are produced by nondiscriminating AspRS and GluRS. These mischarged tRNAs are not recognized by translation elongation factors and hence do not participate in protein synthesis (Praetorius-Ibba and Ibba, 2003), rather are recognized by the tRNA-dependent amidotransferases, AspAdT (Asp-tRNA<sup>Asn</sup> amidotransferase) and GluAdT (Glu-tRNA<sup>Gln</sup> amidotransferase) (Curnow *et al.*, 1996; Curnow *et al.*, 1997). These enzymes produce the correct Asn-tRNA<sup>Asn</sup> and Gln-tRNA<sup>Gln</sup> from Asp-tRNA<sup>Asn</sup> and Glu-tRNA<sup>Gln</sup>, respectively.

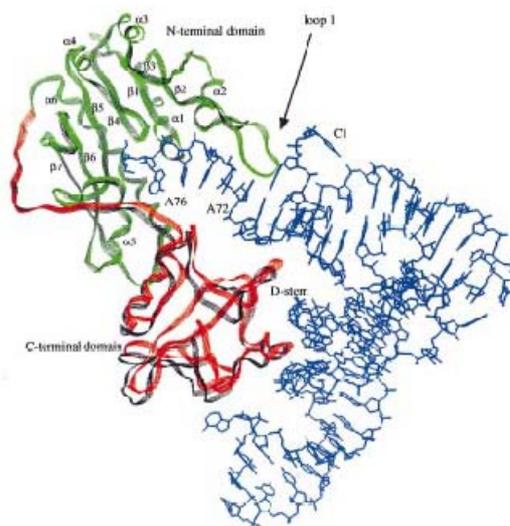
Interestingly, the amide aminoacyl-tRNA synthesis proceeds through distinct pathways in Bacteria, Archaea and Eukarya (Figure 8).



**Figure 8 – Phylogenetic distribution of the indirect and direct pathways of amide aminoacyl-tRNA synthesis.** In blue, Asn-tRNA synthesis; in red, Gln-tRNA synthesis; and in green, both Asn-tRNA and Gln-tRNA synthesis. The direct route is indicated by the presence of AsnRS and GlnRS, whereas the indirect route is indicated by the tRNA-dependent amidotransferases *gatCAB*, encoding a heterotrimeric enzyme with both aspartyl and glutamyl amidotransferase activity, and *gatDE*, encoding a heterodimeric enzyme with glutamyl amidotransferase activity only (Ruan *et al.*, 2001).

In the eukaryotic cytoplasm, a direct mechanism is used for amide aminoacyl-tRNA synthesis, via AsnRS and GlnRS. In most bacteria and eukaryotic organelles, the asparaginyl-tRNAs are also directly synthesized, and glutaminyl-tRNAs are produced indirectly using the tRNA-dependent transamidation pathway, whereas in archaea, the indirect pathway is the most used (Ruan *et al.*, 2001). The exceptions found are the result of horizontal gene transfer, as shown by the phylogenetic analysis of aaRSs (Woese *et al.*, 2000). These analyses also showed that both AsnRS and GlnRS have arisen later and evolved from the ancestral AspRS and GluRS (Woese *et al.*, 2000), suggesting that asparagine and glutamine might have been absent from the primitive genetic code and were added posteriorly, using an indirect pathway for aminoacyl-tRNAs synthesis. It is likely that the common ancestor for GluRS and GlnRS evolved into distinct synthetases after the divergence of gram-positive and gram-negative eubacteria (Sherman *et al.*, 1995). Indeed, the modern GlnRS may have arisen from the eukaryotic GluRS lineage (Woese *et al.*, 2000), as it is absent from archaea (Ruan *et al.*, 2001), therefore, after the establishment of the three organismal domains (Bacteria, Archaea and Eukarya), whereas AsnRS appeared prior to this event (Woese *et al.*, 2000).

Protein synthesis normally starts with methionine or its derivative formyl-methionine, so translation initiation requires a special methionine initiator tRNA different from the elongator, used to insert methionine at internal sites of the peptide chain (reviewed by RajBhandary and Ming Chow, 1995). In eubacteria and eukaryotic organelles, chloroplasts and mitochondria, this initiator is formylmethionyl-tRNA<sup>fMet</sup>. The initiator tRNA is charged with methionine by MetRS that recognizes mainly the anticodon sequence (Schulman and Pelka, 1988; Schulman, 1991). Methionine is then formylated by methionyl-tRNA formyltransferase (MTF), which involves a highly specific recognition process as MTF only formylates the initiator Met-tRNA<sup>fMet</sup> and not any elongator Met-tRNA (Figure 9) (Schmitt *et al.*, 1998).



**Figure 9 – The formylase-fMet- tRNA<sup>fMet</sup> complex.** In blue, the tRNA<sup>fMet</sup>; in green, the N-terminal domain of the formylase and in red, the C-terminal of the enzyme. The interaction of the formylase with the acceptor stem of the tRNA<sup>fMet</sup> is highlighted (Schmitt *et al.*, 1998).

Special structural features of the acceptor stem of this tRNA are crucial for this process, since the base pairs 2:71 and 3:70 are formylation determinants, and the base pair at the top of the acceptor stem is never a Watson-Crick base pair. This mismatch also acts as an antideterminant for interaction with the elongation factor thus excluding initiator tRNA from elongation (Schulman and Her, 1973; Seong and RajBhandary, 1987). The crystal structure of the *E. coli* initiator tRNA showed the disruption of the C1:A72 base pair, and an unusual U33 turn pointing into the loop rather than away from it (Dirheimer *et al.*, 1995).

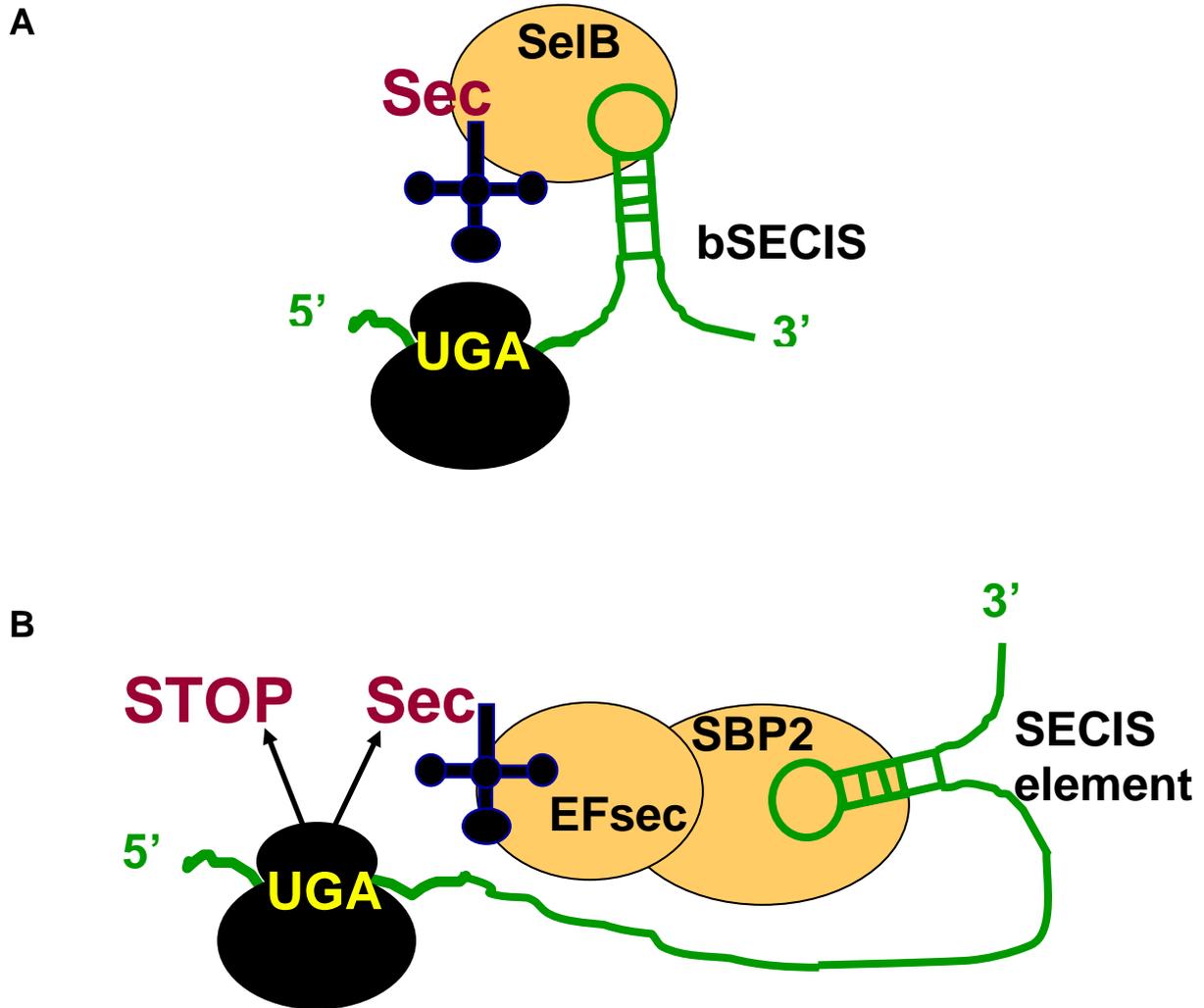
Initiation of protein biosynthesis in the eukaryotic cytoplasm and archaeobacteria does not involve formylation, relying instead on  $\text{tRNA}_i^{\text{Met}}$  whose structure differs from that of the  $\text{tRNA}_e^{\text{Met}}$ . Unlike bacteria, eukaryotic initiator tRNAs have a Watson-Crick base pair at the end of the acceptor stem, which is almost always A1:U72 that is not found in eukaryotic elongator tRNAs (RajBhandary and Ming Chow, 1995). In yeast, the crystal structure of the initiator tRNA showed the typical features of other, although not all, eukaryotic initiators, namely a short D-loop with A20 instead of D20, an unusual T $\psi$ C loop with A54 replacing T54, and a modified ribose at position A64. This modification is an antideterminant that prevents formation of the ternary complex (Kiesewetter *et al.*, 1990). Although not related, it is noteworthy that both prokaryotic and eukaryotic initiator tRNAs have evolved a structural peculiarity with an antideterminant role against interaction with the elongation factor (Dirheimer *et al.*, 1995). Additionally, initiator tRNAs bind directly to the ribosomal P-site, in contrast to elongator tRNAs that bind to the A-site and are then translocated to the P-site. All initiator tRNAs have commonly three consecutive G:C base pairs in the anticodon-stem, which are important for targeting them to the ribosomal P-site (RajBhandary and Ming Chow, 1995).

### 2.1.2. The 21<sup>st</sup> and 22<sup>nd</sup> amino acids: Selenocysteine and Pyrrolysine

Adding to the canonical 20 amino acids, in recent years two new naturally occurring amino acids, namely selenocysteine and pyrrolysine, designated respectively as the 21<sup>st</sup> and 22<sup>nd</sup> amino acids, were discovered (Chambers *et al.*, 1986; Zinoni *et al.*, 1986; Srinivasan *et al.*, 2002; Hao *et al.*, 2002).

Selenium is an essential micronutrient with many potential health benefits and is present in selenium-containing proteins, as a dissociable cofactor or as the amino acid selenocysteine (Sec) located at the enzymes active centre. Sec is cotranslationally inserted into selenoproteins in response to UGA (opal) stop codons. Unlike the 20 canonical amino acids, it is biosynthesized on its tRNA and requires additional factors in order to be incorporated into protein during mRNA translation (reviewed by Hatfield and Gladyshev, 2002; Driscoll and Copeland, 2003). Although the biosynthesis of Sec from serine is well established in prokaryotes, the detailed steps of the pathway in eukaryotes and archaea are still uncharacterized. In *E. coli*, Sec-tRNA<sup>[Ser]Sec</sup> is aminoacylated with serine by the seryl-tRNA synthetase (Leinfelder *et al.*, 1988). A Sec synthase removes the hydroxyl group from serine giving rise to the intermediate that serves as acceptor for activated selenium, leading to the formation of selenocysteyl-tRNA<sup>[Ser]Sec</sup> (Forchhammer and Bock, 1991). The Sec-tRNA can then be used in protein synthesis, in order to incorporate Sec at Sec-specific UGA codons.

But how does Sec get inserted only at some UGA codons? Decoding of the UGA codon as Sec is dependent on the presence of a structural element on the mRNA, designated the Sec insertion sequence (SECIS). SECIS elements consist of stem-loop structures that are present in the 3' untranslated regions of eukaryotic and archaeal selenoprotein genes, whereas in bacteria they are present in the coding regions immediately downstream of the Sec-encoding UGA codons (Hatfield and Gladyshev, 2002).



**Figure 10 – Mechanisms of Sec insertion.** A – In prokaryotes the Sec-specific elongation factor SelB connects the tRNA<sup>Sec</sup> and the SECIS (Sec insertion sequence) element, allowing the introduction of Sec at UGA codons (adapted from Driscoll and Copeland, 2003). B – In eukaryotes, the Sec-tRNA is recruited by the Sec-specific elongation factor EFsec, which binds the protein SBP2. SBP2 is a SECIS-binding protein that interacts with the SECIS element to incorporate Sec at the UGA codon (adapted from Hatfield and Gladyshev, 2002).

In prokaryotes, the elongation factor SelB is specific for Sec incorporation, since it binds the tRNA<sup>Sec</sup> and the SECIS element (Figure 10A) (Driscoll and Copeland, 2003). In eukaryotes, the SECIS elements bind the SECIS-binding protein SBP2 in a tight complex, which in turn recruits the Sec-specific elongation factor EFsec. EFsec is bound to the Sec tRNA<sup>[Ser]Sec</sup> (Figure 10B), thus incorporating Sec in the nascent polypeptide in response to an UGA codon (Tujebajeva *et al.*, 2000).

A key player of the mechanism of Sec insertion into proteins is the tRNA<sup>Sec</sup>, which has an unusual structure and length (reviewed by Commans and Bock, 1999) essential for SerRS, Sec synthase and EFsec recognition. A longer acceptor stem with one or two additional base pairs, in bacteria and eukaryal tRNA<sup>Sec</sup> respectively, is important for interaction with the translation factor EFsec and also for conversion of serine into Sec by Sec synthase (Baron and Bock, 1995; Ibba and Soll, 2000).

Some authors favour the hypothesis that the Sec system is very old, a relic from the early stages of evolution of the translational apparatus, based on the fact that Sec is encoded by UGA. This key feature is conserved in all three lines of descent, therefore considered as part of the universal genetic code and not a deviation (Baron and Bock, 1995). Given the distinct properties of Sec biosynthesis and insertion into proteins, it is also possible that Sec was added at a later stage in evolution to a pre-existent genetic code in order to profit the chemistry of selenium in antioxidant defence and evolve new functions (Gladyshev and Kryukov, 2001).

More recently, the amino acid pyrrolysine (Pyl) was found in archaeal proteins, encoded by the UAG (amber) termination codon (Srinivasan *et al.*, 2002; Hao *et al.*, 2002). A new tRNA<sub>CUA</sub> and aminoacyl-tRNA synthetase, which could charge the tRNA<sub>CUA</sub> with lysine *in vitro*, were recently discovered. The mechanisms of Pyl synthesis and insertion into proteins were, therefore, thought to follow a similar pathway to selenocysteine (Ibba and Soll, 2002), however recent *in vivo* studies suggest that Pyl is charged on the tRNA<sub>CUA</sub> and incorporated into proteins by a

mechanism similar to the 20 canonical amino acids (Blight *et al.*, 2004). The finding that the Pyl-tRNA synthetase is able to activate the amino acid pyrrolysine and charge it on the corresponding tRNA showed that a new naturally occurring synthetase-tRNA pair is responsible for the introduction of an additional amino acid into the genetic code, and opens the possibility that other such pairs exist in nature. Moreover, expression of the archae *pylT* (tRNA<sub>CUA</sub>) and *pylS* (Pyl-tRNA synthetase) in *E. coli* results in the decoding of UAG as Pyl (Blight *et al.*, 2004). How the specific Pyl-encoding UAG codons are found and distinguished from the UAG stop codons is a question that remains unsolved, but mRNA structures and additional factors could be involved in Pyl-specific UAG recognition like for selenocysteine (Schimmel and Beebe, 2004). Although the detailed mechanisms of UAG translation as Pyl are not yet known, the demonstration that additional amino acids occur naturally supports the notion that the genetic code is more flexible than once thought and can still be a box of surprises.

### 2.1.3. Natural genetic code alterations

Variations of the standard genetic code were first found in human mitochondria (Barrell *et al.*, 1979) and since then many more examples of alternative genetic codes were discovered in prokaryotic and eukaryotic nuclear and mitochondrial genomes (Osawa *et al.*, 1992). In fact, mitochondrial genetic codes are particularly prone to change with the exception of plants. The phylogenetic distribution of genetic code changes shows the deviations to the standard genetic code that occur in mitochondria (Figure 11).

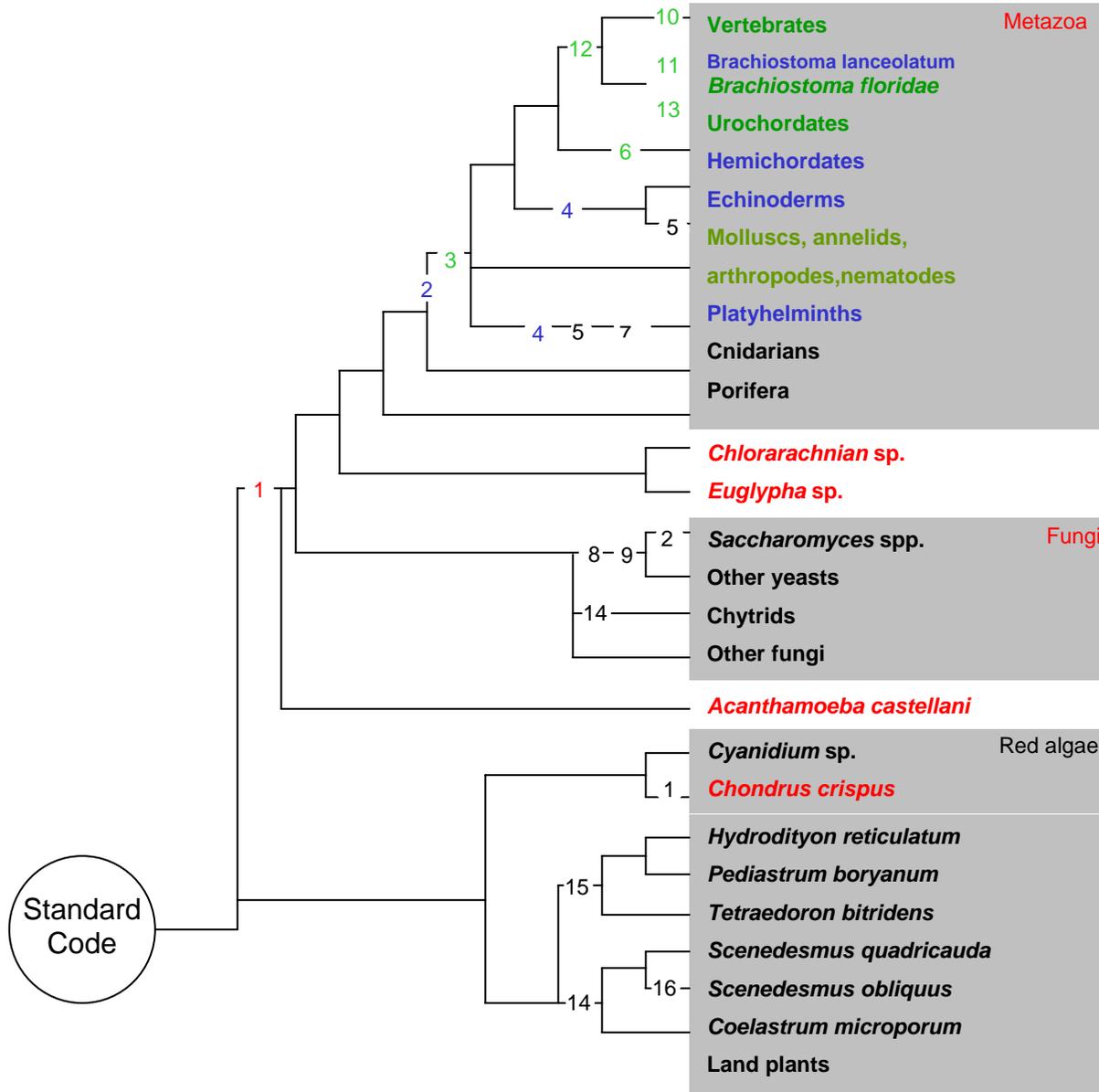
In metazoan mitochondria reassignments involve the codons UGA, AGN (N is A, U, G or C), AUN and AAN (Yokobori *et al.*, 2001). In non-metazoan mitochondria most changes are related to the assignment of the termination codons UAG and UGA, but also the reassignment of sense codons to nonsense (termination), and the unassignment of sense codons. Some of the changes evolved early, for example the standard UGA stop codon codes for tryptophan in the mitochondria of most organisms (Yokobori *et al.*, 2001).

Other codon reassignments, like the arginine AGR codons code for different amino acids in different organisms. AGA and AGG are decoded as serine in platyhelminths, nematodes, annelids, arthropods, molluscs, echinoderms and hemichordates, but in urochordates the AGR codons have been reassigned to glycine. In other metazoa, these codons have been unassigned and further altered for termination in vertebrates, and the AGA codon changed back to glycine or serine in different species of *Brachiostoma* (reviewed by Knight *et al.*, 2001).

Also curious is the case of the isoleucine AUA codon that codes for methionine in most metazoan mitochondria, but in platyhelminths, echinoderms and hemichordates this codon changed again to the standard isoleucine. Also in the mitochondria of platyhelminths and echinoderms the standard AAA-lysine codon has been reassigned to asparagine (Castresana *et al.*, 1998).

In the mitochondria of the yeast *Saccharomyces cerevisiae*, the canonical stop codon UGA is decoded as tryptophan, the standard AUA-isoleucine codes for methionine and the entire leucine-CUN family has been reassigned to threonine (Pape *et al.*, 1985).

**Figure 11 – Phylogenetic distribution of genetic code changes that occur in mitochondria** (adapted from Knight *et al.*, 2001).



①	UGA	Stop	→	Trp
②	AUA	Ile	→	Met
③	AGR	Arg	→	Ser
④	AUA	Met	→	Ile
⑤	AAA	Lys	→	Asn
⑥	AGR	Ser	→	Gly
⑦	UAA	Stop	→	Tyr
⑧	CUN	Leu	→	Thr
⑨	CGN	Arg	→	UN
⑩	AGR	UN	→	Stop
⑪	AGA	UN	→	Gly
⑫	AGR	Ser	→	UN
⑬	AGA	UN	→	Ser
⑭	UAG	Stop	→	Leu
⑮	UAG	Stop	→	Ala
⑯	UCA	Ser	→	Stop

The nuclear-cytoplasmic genetic code changes involve mostly the reassignment of stop codons to sense codons and the unassignment of sense codons, and can be considered a subset of the mitochondrial alterations (Figure 12).

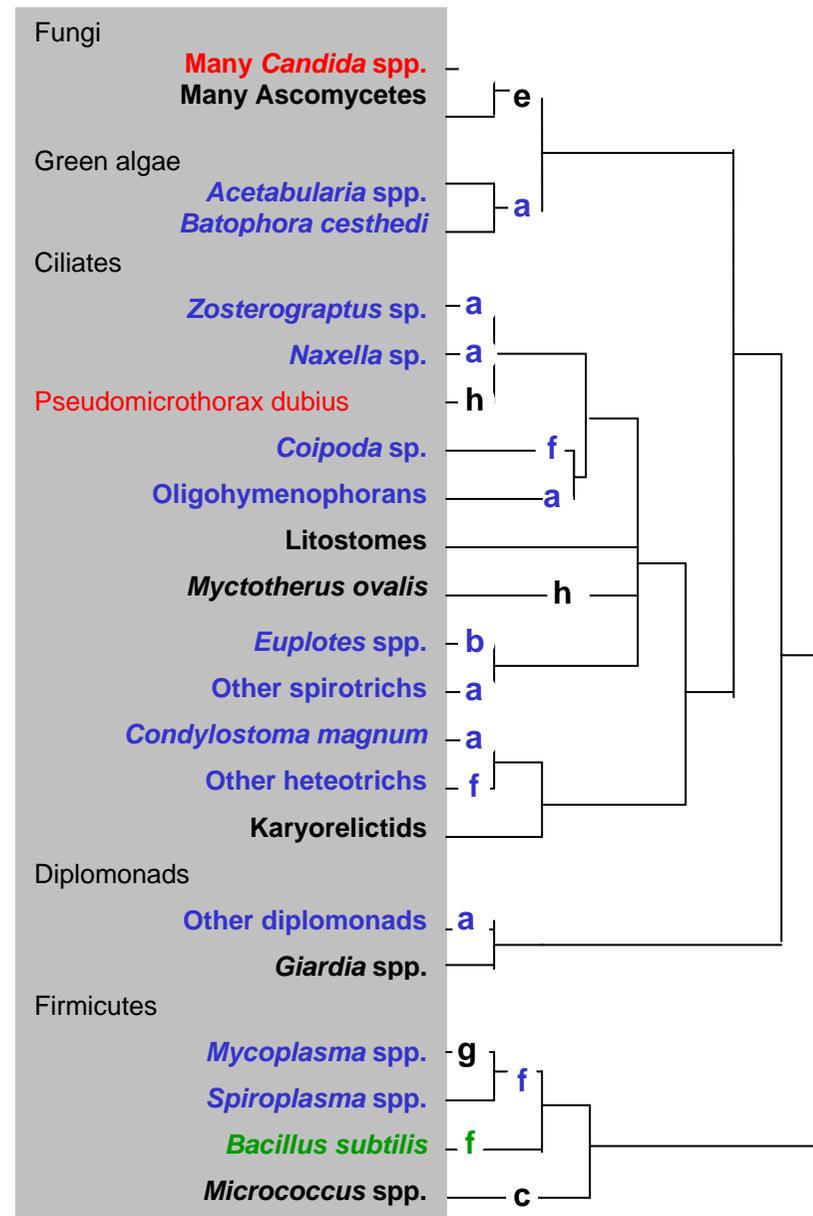
In ciliates, one (UGA) or two (UAR) of the three termination codons are frequently translated into different amino acids. The genera *Oxytricha*, *Paramecium* and *Tetrahymena* decode the UAR codons (UAA and UAG) as glutamine, UGA means cysteine in *Euplotes* and recent work demonstrates that three peritrich species translate UAA into glutamate (Sanchez-Silva *et al.*, 2003). Some bacteria such as *Mycoplasma* and *Spiroplasma* species also reassigned the UGA stop codon to tryptophan. Alterations to the standard genetic code in *Mycoplasma* species further include the unassignment of the canonical arginine-CGG codon (Oba *et al.*, 1991). In *Micrococcus*, the AGA-arginine and AUA-isoleucine codons are also unassigned, since they do not code for any amino acid (Ohama *et al.*, 1990; Kano *et al.*, 1991).

Interestingly, in *Bacillus subtilis* the UGA stop codon has been reassigned to tryptophan, however this codon remained ambiguous. Therefore, it can be both decoded as tryptophan or can be used for translation termination (Lovett *et al.*, 1991).

The only known sense-to-sense reassignment in the eukaryotic nuclear genetic code is the decoding of the standard leucine CUG codon as serine, which occurs in several *Candida* species (Santos and Tuite, 1995). This is a very peculiar alteration of the genetic code, as the CUG codon has not been fully reassigned and remains ambiguous in some species, coding for two amino acids simultaneously (discussed in chapter I.3.).

**Figure 12 – Phylogenetic distribution of genetic code changes that occur in bacteria and eukaryotes** (adapted from Knight *et al.*, 2001).

(a)	UAR	Stop	→	Gln
(b)	UGA	Stop	→	Cys
(c)	CUG	Leu	→	Ser
(d)	AGA	Arg	→	UN
(e)	AUA	Ile	→	UN
(f)	UGA	Stop	→ ↔	Trp*
(g)	CGG	Arg	→	UN
(h)	UGA	Stop	→	UN



## **2.2. Theories for the evolution of the genetic code**

Two theories have been put forward to explain the evolution of alternative genetic codes, the “Codon Capture” theory (Osawa and Jukes, 1989) and the “Ambiguous Intermediate” theory (Schultz and Yarus, 1994), reviewed by Massey *et al.* (2003).

The “Codon Capture” theory proposes a neutral mechanism that excludes any function for genetic code deviations. According to this hypothesis, the reassigned codon firstly disappears from the genome, and then its cognate tRNA also disappears. Next, a natural or mutant near-cognate tRNA able to decode the lost codon via misreading captures the missing codon reintroduced by mutation (Osawa and Jukes, 1989). Complete codon loss is a necessary step for the reassignment process, and the driving force of codon loss and gain is GC pressure. This model fits very well for some genetic code alterations, such as the unassignment of the arginine CGG codon in *Mycoplasma capricolum*. In this AT-rich (75%) genome the G-ending CGG codon and its cognate tRNA<sup>Arg</sup> have been lost (Oba *et al.*, 1991). Similarly, in the GC-rich (74%) genome of *Micrococcus luteus* the A-ending arginine AGA and the isoleucine AUA codons have disappeared (Ohama *et al.*, 1990; Kano *et al.*, 1991).

Although widely accepted, the “Codon Capture” theory establishes that codon reassignment could only be possible with rarely used codons, as a codon has to disappear from the entire ORFeome (Osawa *et al.*, 1992). In mitochondria this could be possible, due to the small size of the genome, however, in prokaryotes and eukaryotes the higher number of protein coding genes turns this pathway hard to follow. Indeed, in an extension of the “Codon Capture” theory, Watanabe and co-workers suggested that complete loss of a codon might not be required for codon reassignment (Yokobori *et al.*, 2001). Additionally, if the aaRS recognizes the anticodon loop region of the tRNA, the near-cognate tRNA that captures the lost codon cannot be altered through mutation in the anticodon sequence, unless a new aaRS specific for the mutant tRNA also appears.

The “Ambiguous Intermediate” theory suggests a model that establishes genetic code ambiguity as the critical step for codon reassignment (Schultz and Yarus, 1994). In contrast to the “Codon Capture”, the “Ambiguous Intermediate” theory assumes that a transitional codon can be decoded simultaneously by two tRNAs, one cognate and another near-cognate with expanded decoding properties (Schultz and Yarus, 1994). Codon reassignment is predicted to occur by mutations that alter tRNA abundance or decoding and aminoacylation fidelity, and render a codon ambiguous by competition of the cognate and a near-cognate tRNAs for this codon. Selection might favour further mutations that improve decoding by the near-cognate tRNA, thus changing the meaning of the codon. This theory has been recently reformulated, based on new data that support the hypothesis that codon ambiguity can drive codons to extinction (Massey *et al.*, 2003). Indeed, the use of comparative and functional genomics has highlighted the impact that genetic code ambiguity had on the usage of the CUG codon in *C. albicans*. The CUG reassignment from leucine to serine in *Candida* species is mediated by a novel tRNA (chapter 1.3.) and is consistent with the “Ambiguous Intermediate” theory. *Candida* spp. and all other eukaryotes have an intra- and inter-chromosomal mosaic distribution pattern of GC content, making it difficult to eliminate specific codons from the genome by GC pressure alone (Santos *et al.*, 1997). Indeed, aligning orthologous genes from *C. albicans*, *S. cerevisiae* and *S. pombe* showed that CUG decoding as both leucine and serine forced the CUG codons to disappear from the genome of the ancestor, since a decreased usage of the CUG codon would minimize the negative impact of ambiguous CUG decoding (Massey *et al.*, 2003). This finding suggests that ambiguity could be able to drive codon reassignment similarly to GC pressure, and that either biased GC content or codon ambiguity or both can be used to explain some observed codon unassignments. Interestingly, the *C. albicans* genome has approximately 17000 CUG codons. How can this paradox be explained? Sequence alignments showed that the CUG codons present in extant *C. albicans* genes are represented by serine codons on *S. cerevisiae* and *S. pombe* homologous genes, therefore representing codons that evolved recently from serine and not leucine codons

(Massey *et al.*, 2003). This replacement of old leucine with new serine CUG codons was shaped by the genetic code ambiguity introduced by the ser-tRNA<sub>CAG</sub>.

The ambiguous transitional state of a codon to be reassigned might not endanger cell survival as much as once thought, as readthrough of stop codons is known to be tolerated in *Escherichia coli*, yeast or mammalian cells (Feng *et al.*, 1990; Beier and Grimm, 2001). Moreover, half of the total amino acid substitutions yield active proteins (Zabin *et al.*, 1991; Huang *et al.*, 1992). This was shown in bacteriophages and in *E. coli*, where substitutions that changed size, hydrophobicity, and charge of the amino acid, did not decrease the activities of the altered proteins (Hellinga *et al.*, 1992). Indeed, ambiguity of the genetic code is not incompatible with cell survival in bacteriophages (Bacher *et al.*, 2003), *E. coli* (Pezo *et al.*, 2004) and yeast (Santos *et al.*, 1996; Santos *et al.*, 1999; this work), supporting the “Ambiguous Intermediate” theory.

The reassignment of termination codons is common and they may be explained by the “Codon Capture” theory, as stop codons are relatively infrequent and their substitution produces extension of the protein instead of amino acid replacement. According to the Osawa-Jukes model, following the reintroduction of a lost codon in the ORFeome, any tRNA could capture the unassigned codon. However, stop codon reassignments show a pattern of repetition where the same codons are systematically reassigned to the same amino acids (chapter 1.2.1.3). The “Ambiguous Intermediate” theory is able to justify these non-random changes based on tRNA-structure mediated codon reassignments. Competition between the release factor complex and a nonsense suppressor tRNA creates an intermediary stage where the stop codon is ambiguous, meaning both translation termination and sense decoding. Further mutation of the release factor prevents its recognition of the stop codon (Lozupone *et al.*, 2001), while efficient decoding is assured by tRNA mutation. Therefore, both the “Codon Capture” and the “Ambiguous Intermediate” theories are valid to explain genetic code changes and are not mutually exclusive. In fact, for some examples even a combination of the two can be at work (Knight *et al.*, 2001).

These theories are subjacent to two different approaches to engineer alternative genetic codes, as will be discussed in the next chapter. “Top-down” approaches, in which the proteome is globally affected, correspond basically to the “Ambiguous Intermediate” theory and have been performed with *B. subtilis*, *E. coli* and bacteriophages grown in the presence of amino acid analogues. Survival of these evolved organisms with ambiguous proteomes was possible, although their fitness has decreased, suggesting that genetic code alterations through codon decoding ambiguity might be limited to similar amino acids or rare codons (reviewed by Bacher *et al.*, 2004). “Bottom-up” approaches are in agreement with the “Codon Capture” theory, as genetic code alterations have been mainly obtained by engineering specific aaRSs and tRNAs to incorporate non-natural amino acids into a determined stop codon. The number of codons and proteins altered is smaller with these methods but the chemical diversity of the substitutions is larger, so a combination of both approaches that could be reflected on the organism as a whole will be perfect to study genetic code evolution (Bacher *et al.*, 2004).

### **2.3. Artificial genetic codes**

Additional chemical groups generated by post-translational modification, such as phosphorylation and glycosylation, or cofactors are required for many proteins to carry out their normal functions, suggesting that life with a 20 amino acid code might be sufficient but not optimal (Cropp and Schultz, 2004). Alteration or expansion of the genetic code, either by insertion of nonnatural amino acids or by shuffling the natural occurring building blocks of proteins, could improve enzyme stability, activity and even introduce new properties. Besides the potential applications, genetic code manipulation provides important insight on the evolution of life and the genetic code itself (Hendrickson *et al.*, 2004). Peptides can be modified to create novel pharmaceuticals, such as protease inhibitors used against HIV (Kiso *et al.*, 1999; Mak *et al.*, 2003) and *C. albicans* infections (Bein *et al.*, 2002; Hruby, 2002). A handful of different compounds useful in research as biophysical probes has also been produced by incorporation of nonnatural amino acids (reviewed by Hendrickson *et al.*, 2004).

Recent efforts towards altering or expanding the genetic code in several organisms have been successful. Incorporation of nonnatural amino acids was achieved in bacteriophages (Bacher *et al.*, 2003), *E. coli* (Wang *et al.*, 2001; Döring *et al.*, 2001; Mehl *et al.*, 2003) and *S. cerevisiae* (Chin *et al.*, 2003), and also the naturally archae-occurring 22<sup>nd</sup> amino acid, pyrrolysine, was introduced in *E. coli* (Blight *et al.*, 2004).

An evolved bacteriophage that can incorporate a tryptophan analogue was obtained with only few mutations. Continued growth in the presence of 6-fluorotryptophan improved the initial poor fitness of the phage as the proteome gradually adapted to ambiguity, suggesting that evolution of organisms with alternative genetic codes is possible with ambiguous intermediates (Bacher *et al.*, 2003).

To expand the genetic code of *E. coli*, an orthogonal tRNA-aaRS pair was obtained based on the *Methanococcus jannaschii* tyrosyl-tRNA synthetase (TyrRS) and mutant tyrosine amber suppressor tRNA (Wang *et al.*, 2001). Orthogonality ensures that the new synthetase does not aminoacylate any other cellular tRNA and that the new tRNA is not recognized by the existing synthetases. After selection, a mutant TyrRS that could charge the synthetic amino acid O-methyl-L-tyrosine was obtained and used to incorporate this amino acid in response to the amber nonsense codon UAG (Wang *et al.*, 2001).

In these studies, the unnatural amino acid has to be added to the growth medium for uptake by the cell. However, this exogenous limitation of amino acids was recently overcome by engineering a bacterium, which could synthesize and incorporate the secondary metabolite *p*-aminophenylalanine (*p*AF) at the amber stop codon (Mehl *et al.*, 2003). The orthogonal tRNA-aaRS pair used was the same as before (Wang *et al.*, 2001). Mutant TyrRS that recognize *p*AF were selected and expressed in *E. coli* as well as the *S. venezuelae* genes for *p*AF biosynthesis, allowing the introduction of this compound into proteins with high fidelity (Mehl *et al.*, 2003).

In a different approach, the genetic code of *E. coli* was expanded by incorporation of unnatural amino acids at codons that normally specify natural amino acids. The method developed consisted on disrupting the editing domain of valyl-tRNA synthetase (ValRS) so that it can mischarge the cognate tRNA<sup>Val</sup> with cysteine or the similar noncanonical aminobutyrate, introducing these amino acids at valine codons (Döring *et al.*, 2001). This changes the amino acid composition of many cellular proteins, and can confer selective advantage to the engineered cells (Döring *et al.*, 2001).

Schultz and co-workers were also able to expand an eukaryotic genetic code, by addition of an orthogonal pair consisting of the amber suppressor tRNA<sub>CUA</sub> and tyrosyl-tRNA synthetase from *E. coli* into *S. cerevisiae* (Chin *et al.*, 2003). Using forced selection methodologies, mutant TyrRS that charge the tRNA<sub>CUA</sub> with the

chosen compound were obtained, thus constituting a route for the insertion of additional amino acids at the amber codon. 5 distinct unnatural amino acids with novel properties were added by this method to the genetic code of yeast cells (Chin *et al.*, 2003).

Other authors managed to incorporate pyrrolysine into an *E. coli* strain by expressing *pylT* (tRNA<sub>CUA</sub>) and *pylS* (pyrrolysil-tRNA synthetase), and envisage the possibility of adding this natural amino acid to proteins of other species (Blight *et al.*, 2004).

Most strategies for the introduction of non-natural amino acids into the proteins of living cells minimize ambiguity, by limiting the target codons and the amino acid substitutions. This was the approach followed in some of the studies described above (Wang *et al.*, 2001; Mehl *et al.*, 2003; Chin *et al.*, 2003). The use of an orthogonal tRNA-aaRS pair derived from a foreign species will result in the incorporation of the synthetic amino acid at specific codons, however the new amino acid will be limited to a subset of proteins and not all proteome will be altered. But genetic code alterations with ambiguity of tRNA aminoacylation, based for instance on the disruption of the editing activity of aaRSs (Döring *et al.*, 2001), will have an impact on all the proteome and can provide important data to understand the evolution of the genetic code (Bacher *et al.*, 2003; Pezo *et al.*, 2004).

Ambiguity of the genetic code is toxic and inhibits growth proportionally to the degree of misacylation (Nangle *et al.*, 2002), however it can confer growth advantages in certain environments, regardless of its artificial (Pezo *et al.*, 2004) or natural origin (Santos *et al.*, 1996; Santos *et al.*, 1999; this work). The production of statistical proteins, groups of related heterogeneous polypeptides in which several amino acids can be inserted at a single position, might enhance adaptation capacity in the presence of selective pressure. Therefore, ambiguity should have had an important role in the early evolution of life, when organisms needed to adapt to rapidly changing environments. The production of statistical

proteins by an ambiguous genetic code allowed cell survival, as distinct subsets of the same proteins would be more appropriate to grow under different stress conditions, improving fitness. In particular, during nutrient limitation the use of available alternative amino acids for the production of essential proteins could provide growth advantage (Pezo *et al.*, 2004).

The work described in this thesis characterizes eukaryotic cells with an altered genetic code, in which the CUG codon is ambiguous due to the expression of a mutant tRNA that allows incorporation of serine instead of the standard leucine (see chapter 1.3.). Although there is a natural amino acid replacing the cognate one, and despite the generation of a statistical proteome, these cells are viable and display stress-resistance phenotypes, strengthening the hypothesis that it would be possible to create a new organism in which the genetic code has been changed through ambiguous intermediates.

### 3. The *Candida* spp. genetic code

A unique genetic code change is the decoding of the leucine-CUG codon as serine in *C. albicans* and several other *Candida* species. This genetic code change is mediated through ambiguous codon decoding by a novel transfer RNA, the Ser-tRNA<sub>CAG</sub> (Santos, 1992; Santos *et al.*, 1993; Santos *et al.*, 1996; Suzuki *et al.*, 1997), which translates the CUG codon both as leucine and serine. This is consistent with the “Ambiguous Intermediate” theory rather than the “Codon Capture” theory (chapter 1.2.2.). Nevertheless, lowering the usage frequency of the CUG codon, favouring other leucine codons instead, might have been important in CUG reassignment by minimizing the deleterious consequences of ambiguous CUG decoding. Such disappearance or reduction of codon usage to a tolerable minimum was driven by codon misreading and not by GC pressure alone (Massey *et al.*, 2003).

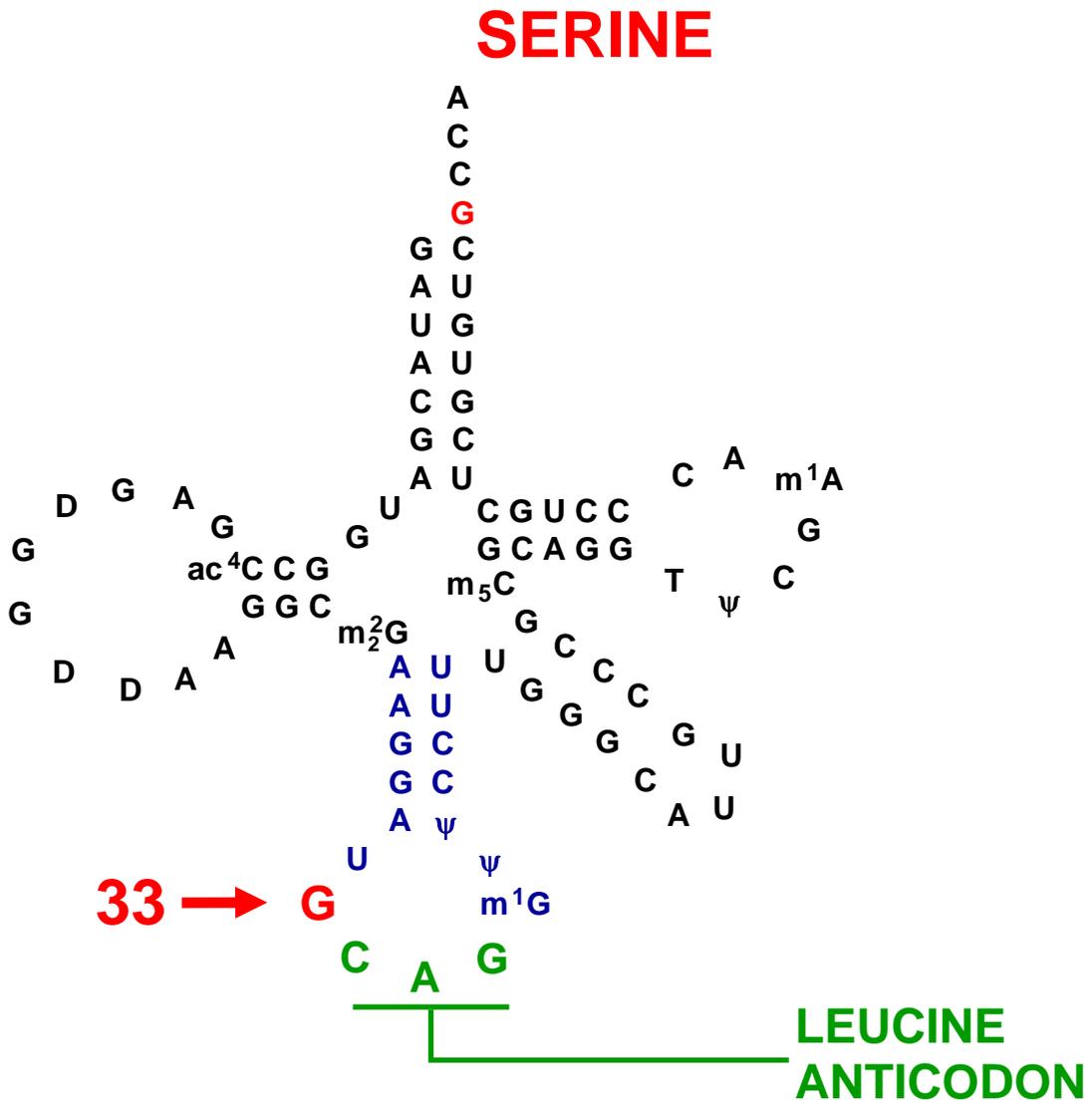
Some *Candida* species translate the CUG codon exclusively as leucine, such as *C. glabrata* and *C. krusei*, while others like *C. cylindracea* decode it as serine only. However, in many species the CUG codon is ambiguous, meaning simultaneously leucine and serine, as the ser-tRNA<sub>CAG</sub> can be charged with 3 % leucine in *C. zeylanoides* (Suzuki *et al.*, 1997) and *C. albicans* (Catarina Gomes, unpublished results). These facts raise interesting questions about the *Candida* genetic code. For example, is CUG reassignment still evolving in different *Candida* species? Is serine-CUG decoding the final step of the evolutionary pathway or is CUG ambiguity advantageous and selected for? What is the impact of CUG reassignment to the cell? How can organisms cope with the negative effects of CUG mistranslation?

### 3.1. The Serine-tRNA<sub>CAG</sub>

The reassignment of the CUG codon from leucine to serine in several *Candida* species occurs due to the unusual decoding and aminoacylation properties of the Ser-tRNA<sub>CAG</sub> (Santos, 1992; Santos *et al.*, 1993; Santos *et al.*, 1996; Suzuki *et al.*, 1997). The structure of the *C. albicans* ser-tRNA<sub>CAG</sub> is shown on Figure 13.

This tRNA was shown to be a chimera of a leucine-tRNA and a serine-tRNA (Suzuki *et al.*, 1997), as the anticodon-arm is similar to leucine-tRNAs and has a leucine anticodon, whereas the remaining of the molecule is similar to serine-tRNAs, with a long variable arm and the discriminator base for serine. Remarkably, ambiguity of the CUG codon results from the existence of identity determinants for serine and leucine on the same tRNA, allowing for its recognition by both synthetases, LeuRS and SerRS. A<sub>73</sub> is the discriminator base for LeuRS, while G<sub>73</sub> is the discriminator base for SerRS (Soma *et al.*, 1996), and therefore, the presence of G<sub>73</sub> indicates that the ser-tRNA<sub>CAG</sub> is charged with serine. Additionally, A<sub>35</sub> and the m<sup>1</sup>G at position 37 are leucylation determinants (Suzuki *et al.*, 1997), suggesting that the ser-tRNA<sub>CAG</sub> can also be charged with leucine. Indeed, the *C. albicans* ser-tRNA<sub>CAG</sub> can be both serylated and leucylated. In contrast, the *C. cylindracea* ser-tRNA<sub>CAG</sub> has an A<sub>37</sub> and, thus, cannot be recognized by LeuRS nor charged with leucine, which explains why the CUG codon is decoded exclusively as serine in this species (Suzuki *et al.*, 1997).

One of the most striking structural features of this ser-tRNA<sub>CAG</sub> is the presence of guanosine at position 33 where all other tRNAs have a pyrimidine, generally a uridine. U<sub>33</sub> is conserved, as it is required for the correct turn of the phosphate backbone and stacking of the anticodon bases (Ladner *et al.*, 1975; Woo *et al.*, 1980) and, therefore, the replacement of U<sub>33</sub> by G<sub>33</sub> may have had an important role on CUG reassignment in *Candida* species (Santos *et al.*, 1996; Santos *et al.*, 1997; Suzuki *et al.*, 1997).



**Figure 13 – The *Candida albicans* Ser-tRNA<sub>CAG</sub>.** This tRNA has an anticodon arm similar to leu-tRNAs (in blue) and the leucine anticodon CAG (in green), but the remaining part of the molecule is similar to ser-tRNAs (in black). G<sub>33</sub> (in red) played a critical role during CUG reassignment from leucine to serine (Santos, 1992).

A tRNA with amino acid acceptor identity different from its cognate codon-anticodon interaction, inserting the wrong amino acid in response to a given codon, is expected to be quite toxic to the cell, since the replacement of one amino acid with another throughout cellular proteins would lead to loss of function of the proteins (Pallanck *et al.*, 1995). Therefore, known missense suppressors are generally inefficient (Yarus and Smith, 1995), and in the case of the ser-tRNA<sub>CAG</sub>, G<sub>33</sub> lowered its decoding efficiency (Santos *et al.*, 1996). Thus, G<sub>33</sub> allowed the cell to tolerate the deleterious effects of CUG mistranslation, by reducing the number of proteins altered by the incorporation of serine instead of leucine. It is conceivable that this structural change lowers both leucylation and the decoding efficiency of the ser-tRNA<sub>CAG</sub>, as observed for the selenocysteine-tRNA whose unusual structure lowers its aminoacylation efficiency (Ibba and Soll, 2000). Moreover, modified nucleosides in the anticodon loop play major roles in codon reassignment, strengthening the idea that tRNA structural alteration is crucial for the evolution of alternative genetic codes (Yokobori *et al.*, 2001).

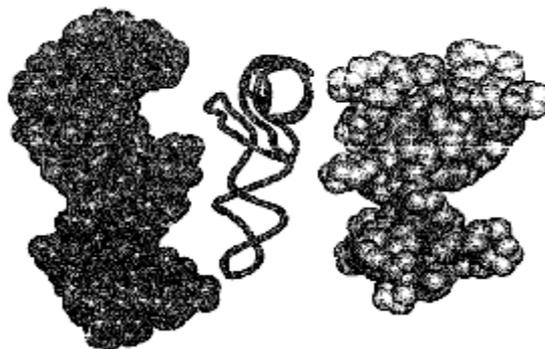
Since the unusual structure of the ser-tRNA<sub>CAG</sub> was essential for the evolution of CUG reassignment from leucine to serine, it poses the important question of “Did the ser-tRNA<sub>CAG</sub> originated from a ser-tRNA or a leu-tRNA”? If the ser-tRNA<sub>CAG</sub> evolved from a leu-tRNA, the mutation of position 33 from U to G, giving rise to a leucine anti-determinant, decreased efficient recognition and charging of the tRNA by the LeuRS. This allows the tRNA to be captured by SerRS and then additional mutations improved serylation efficiency. The discriminator position 73 is obviously an important target, since A<sub>73</sub> is required for interaction with LeuRS (Soma *et al.*, 1996), however the *C. albicans* ser-tRNA<sub>CAG</sub> has a G<sub>73</sub>. Indeed, a single mutation from A<sub>73</sub> to G<sub>73</sub> changes the identity of a human leu-tRNA to a ser-tRNA *in vitro* (Breitschopf and Gross, 1994). Additionally, one nucleotide insertion on the variable arm of a *S. cerevisiae* leu-tRNA conferred serine acceptor activity *in vitro* (Himeno *et al.*, 1997). Therefore, few mutations are required to change a leu-tRNA into a ser-tRNA, supporting the hypothesis that the ser-tRNA<sub>CAG</sub> originated from a leu-tRNA. The simultaneous presence of m<sup>1</sup>G at position 37, which is crucial for leucylation (Suzuki *et al.*, 1997), and G at position 73, which is a determinant for

SerRS recognition, allowed the novel ser-tRNA<sub>CAG</sub> to be charged with the two distinct amino acids, leucine and serine, conferring ambiguity to the CUG codon at an early evolutionary stage of the CUG reassignment pathway.

In contrast, the ser-tRNA<sub>CAG</sub> might have evolved from a ser-tRNA, as it has higher nucleotide identity to ser-tRNAs than leu-tRNAs (Santos *et al.*, 1997). If so, its CUG-anticodon sequence, corresponding to leucine, should have been originated from altered splicing of an intron present in the ser-tRNA<sub>IGA</sub> (Yokogawa *et al.*, 1992; Ueda *et al.*, 1994). The novel ser-tRNA<sub>CAG</sub>, created in this altered splicing event, would be charged exclusively with serine, considering that all the identity determinants for interaction with SerRS would be present, although the anticodon sequence corresponded to leucine. Mischarging of the ser-tRNA<sub>CAG</sub> with leucine would require additional mutation of A<sub>37</sub> to m<sup>1</sup>G<sub>37</sub>, since ser-tRNAs usually have A<sub>37</sub> and the ser-tRNA<sub>CAG</sub> has m<sup>1</sup>G at position 37, which acts as a leucylation determinant (Suzuki *et al.*, 1997). Such mutation would have occurred later during evolution, suggesting that ambiguity could be the final step of the CUG reassignment pathway by introducing some kind of evolutionary advantage (Santos *et al.*, 1997). m<sup>1</sup>G<sub>37</sub> is also important for reading frame maintenance of leu-tRNAs (Agris, 2004), therefore, it may have appeared due to the need to prevent frameshifting by the ser-tRNA<sub>CAG</sub>. If so, is the ambiguity a consequence of the need to maintain the reading frame rather than an evolutionary imperative of the codon reassignment?

### 3.2. LeuRS and SerRS

LeuRS and SerRS are class I and II aaRSs, respectively. As described earlier (chapter I.1.3.), synthetases from distinct classes bind the tRNA from opposite sites (Figure 14), and such symmetry between classes could reflect a primitive stage during the evolution of the genetic code, when synthetase pairs interacted with the same tRNA (Ribas de Pouplana and Schimmel, 2001a). Interestingly, the correlation between the codons used to specify leucine and serine suggests that LeuRS and SerRS might have been one of such pairs (Ribas de Pouplana and Schimmel, 2001). If so, it is possible that the LeuRS may mischarge ser-tRNAs at low levels, or that SerRS can residually mischarge leu-tRNAs.



**Figure 14 – Interaction of the two distinct classes of aaRSs with a tRNA.** A class I synthetase is represented on the left and on the right a class II synthetase is shown, highlighting the mirror-symmetrical interaction with the tRNA (on the centre) (adapted from Arnez and Moras, 1997).

Accordingly, LeuRS and SerRS distinguish different regions in their cognate tRNAs. In *S. cerevisiae*, the LeuRS recognizes the anticodon stem and the D-loop (Dietrich *et al.*, 1990; Soma *et al.*, 1996), while SerRS recognizes the long extra-arm instead (Dock-Bregeon *et al.*, 1989). As discussed in the previous chapter, both synthetases can aminoacylate the ser-tRNA<sub>CAG</sub>, because both ser and leu identity elements are present simultaneously on the ser-tRNA<sub>CAG</sub>. The leucine anticodon, namely the base A<sub>35</sub>, and m<sup>1</sup>G<sub>37</sub> contribute to LeuRS recognition (Soma *et al.*, 1996), whereas the discriminator base G<sub>73</sub> and the extra-arm are determinant for SerRS recognition. The G at position 33 decreases affinity for the LeuRS, therefore the ser-tRNA<sub>CAG</sub> is mainly charged with serine (Suzuki *et al.*, 1997).

The fact that the ser-tRNA<sub>CAG</sub> can be charged with both amino acids raises the possibility that a structural change in one or both synthetases might have had a role in CUG reassignment from leucine to serine. Study of the *C. albicans* leucyl-tRNA synthetase (CaLeuRS) and seryl-tRNA synthetase (CaSerRS) showed that these synthetases have significant amino acid identity with homologous LeuRSs and SerRSs from other organisms and fully complement *S. cerevisiae* LeuRS and SerRS null strains, respectively (O'Sullivan *et al.*, 2001a; O'Sullivan *et al.*, 2001b). These observations suggest that CaLeuRS and CaSerRS recognize and charge *S. cerevisiae* leu-tRNAs and ser-tRNAs similarly to the *S. cerevisiae* LeuRS and SerRS, supporting the hypothesis that the key element in CUG reassignment is the ser-tRNA<sub>CAG</sub> rather than the LeuRS or the SerRS.

### 3.3. CUG identity redefinition through decoding ambiguity

According to the “Codon Capture” theory (chapter 1.2.2.) genetic code alterations are neutral and so have no evolutionary implications, since there is no production of novel proteins with different functions (Osawa and Jukes, 1989). In the opposite scenario, codon reassignments through genetic code ambiguity potentially create diversity due to the mutations introduced, which might be advantageous to the organism (Santos *et al.*, 1996; Santos *et al.*, 1999). One of these genetic code changes that evolved via ambiguous intermediates is the reassignment of the CUG codon from leucine to serine.

The fact that different *Candida* species translate the CUG codon with distinct meanings raises the possibility that the pathway of CUG reassignment is still evolving. *Candida* species that translate the CUG codon as serine can be divided into two groups, one with a ser-tRNA<sub>CAG</sub> that can be leucylated, as is the case of *C. albicans* and *C. zeylanoides*, and the other where the ser-tRNA<sub>CAG</sub> is serylated only, namely *C. cylindracea* (Suzuki *et al.*, 1997). On the first group, the CUG codon is ambiguous, however in *C. cylindracea* the reassignment seems to be complete as the CUG codon is translated exclusively as serine. This suggests that the ambiguous group would have arisen first during evolution, and the decoding of the CUG codon as serine would be the final stage of the reassignment pathway (Suzuki *et al.*, 1997). This hypothesis is based on the observations that *C. cylindracea* has high copy numbers of the ser-tRNA<sub>CAG</sub> and uses CUG as a main serine codon, while ambiguous *Candida* species have low copy numbers of the ser-tRNA<sub>CAG</sub> and use the CUG codon infrequently. Therefore, taking into account that low ser-CUG usage is important to minimize the negative impact of ambiguity, it is likely that high CUG usage may have appeared late in the evolution of the genus *Candida*, when ambiguity no longer restricts CUG usage (Suzuki *et al.*, 1997).

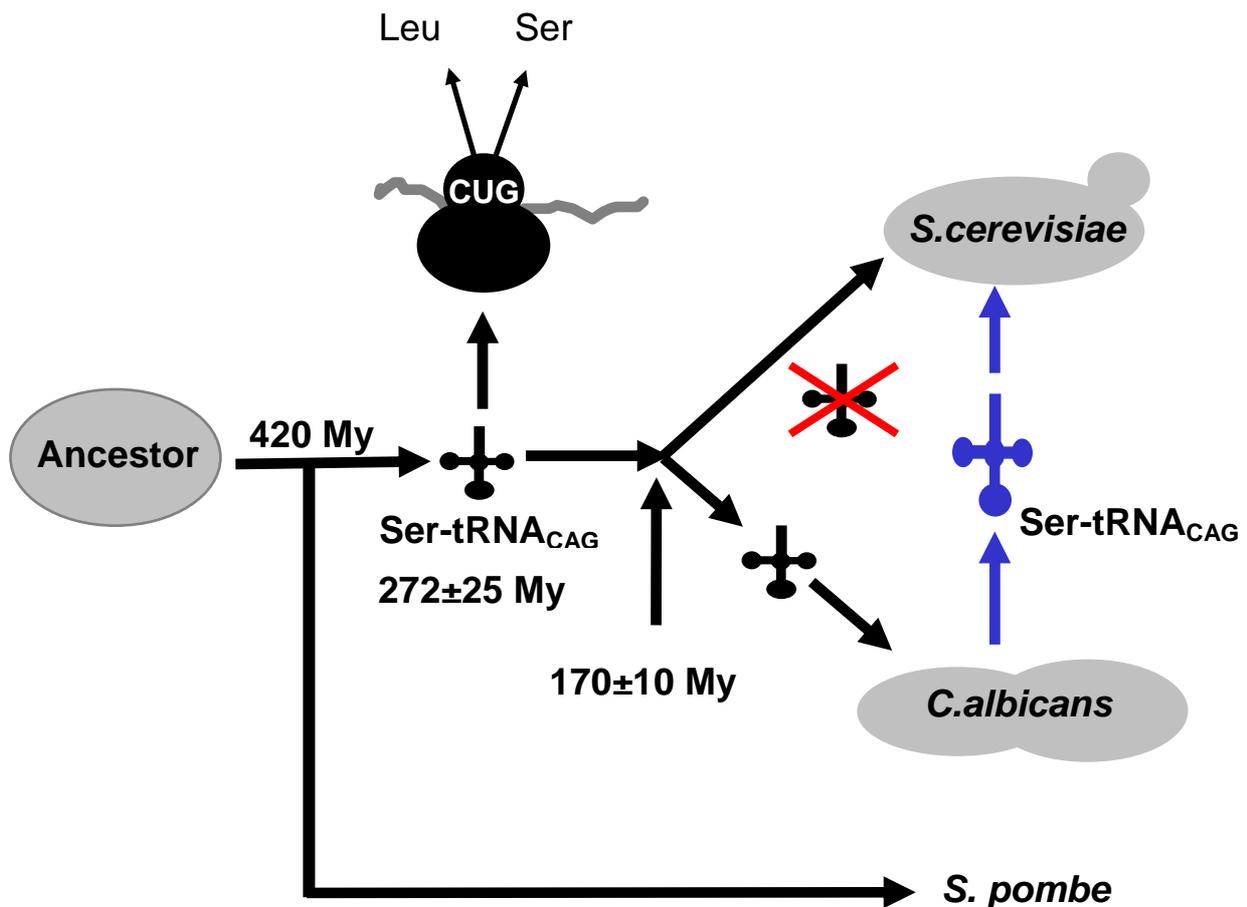
As mentioned above, a partial reassignment is observed in *Candida* species where the CUG codon is ambiguous, namely *C. zeylanoides* and *C. albicans* that

decode the CUG codon both as leucine and serine. In this case, CUG ambiguity should introduce some selective advantages to the cell in order to counteract the negative impact on fitness, originated by translational misreading. Acquisition of novel phenotypes could be such an advantage, as *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> show increased stress resistance (Santos *et al.*, 1996; Santos *et al.*, 1999). Although these cells have decreased growth rates, they can still compete and grow under certain unfavourable conditions where the wild-type cells cannot (Santos *et al.*, 1996; Santos *et al.*, 1999). Therefore, the balance between the deleterious effects of ambiguity and the potential to create functional diversity, allowing the organism to colonize novel ecological niches, might be the driving force on the evolution of genetic code changes through decoding ambiguity.

### 3.4. *Candida* spp. as a model system for studying the evolution of genetic code alterations

The alterations to the standard genetic code found in several organisms led to the formulation of important new questions, namely “what is the mechanism for the evolution of alternative codes?”, “how can microorganisms survive a genetic code change?”, and “what advantages do they create to allow for their selection?”. The study of the *Candida* genetic code provides a good model system to understand the evolution of genetic codes through codon ambiguity and address the questions above (Silva *et al.*, 2004).

The identity of the CUG codon was redefined from leucine to serine by the ser-tRNA<sub>CAG</sub>, which appeared  $272 \pm 25$  million years (My) ago (Figure 15), as shown by rRNA and tRNA molecular phylogeny studies (Massey *et al.*, 2003). Therefore, during the early stages of the evolutionary pathway of CUG reassignment, the appearance of the ser-tRNA<sub>CAG</sub> introduced low-level ambiguity, due to CUG misreading as serine. At these early stages, the presence of G<sub>33</sub> decreased the decoding efficiency of the new tRNA, allowing for its tolerance by the cell. The levels of serine misincorporation at CUG codons gradually increased while translation of CUG as leucine decreased, until the gene(s) for the leu-tRNA that decoded the CUG codon disappeared from the genome (Santos *et al.*, 1997). Indeed, high-level serine-CUG decoding is  $171 \pm 27$  My, coincident with the divergence of the genera *Candida* and *Saccharomyces*, suggesting that the *Saccharomyces* ancestor was ambiguous for the CUG codon during at least 100 My (Figure 15). The serine-tRNA<sub>CAG</sub> was maintained in the lineage that originated the genus *Candida* but lost in the lineage leading to the genus *Saccharomyces*, since the cognate leucine-tRNA<sub>CAG</sub> is absent from *S. cerevisiae* (Massey *et al.*, 2003).

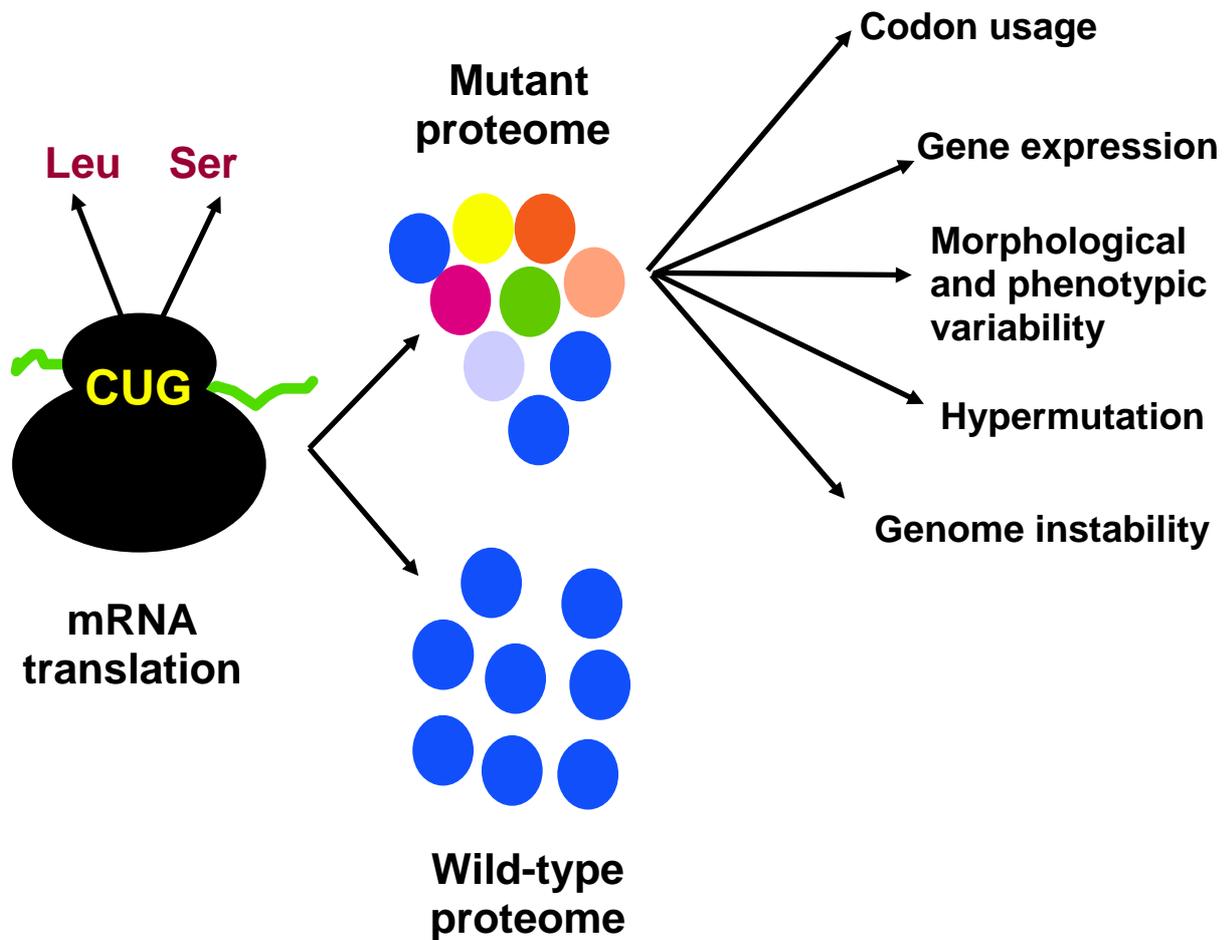


**Figure 15 – Evolutionary pathway of CUG reassignment from leucine to serine in *Candida* spp and its reconstruction in *Saccharomyces cerevisiae*.**

The standard leucine CUG codon has been reassigned to serine by a novel tRNA, which appeared approximately 272 million years (My) ago. However, the divergence of the genera *Candida* and *Saccharomyces* occurred approximately 170 My, suggesting that the *Saccharomyces* ancestor was ambiguous for at least 100 My. The ser-tRNA<sub>CAG</sub> was maintained in the lineage that originated the genus *Candida* but lost in the lineage leading to the genus *Saccharomyces* (Massey *et al.*, 2003). In order to reconstruct the evolutionary pathway of CUG reassignment, the *C. albicans* ser-tRNA<sub>CAG</sub> was reintroduced in *S. cerevisiae* using a single copy plasmid.

The consequences of genetic code ambiguity are represented on Figure 16. Decoding of the leucine-CUG codon as serine results in the production of mutant proteins. Part of this mutant proteome may be degraded, while the other part may accumulate forming aggregates that can be toxic to the cell. To counteract these deleterious effects, the cell triggers the stress response, with the consequent increase in the expression of molecular chaperones. This scenario raises important questions about the adaptation and evolution of organisms under such conditions. First, “how does the organism cope with the permanent production of altered proteins?”, and “what will be the impact of a mutant proteome to the species that redefine codon identity?”.

To tackle these questions, the *Candida* genetic code change was reconstructed by expressing the *C. albicans* Ser-tRNA<sub>CAG</sub> in *S. cerevisiae* (Figure 15). *C. albicans* and *S. cerevisiae* are closely related yeast species, but the CUG codon is translated as leucine in *S. cerevisiae*. Both species have similar GC content, 40 and 36% G+C respectively, similar codon usage and the CUN codon family is used at low frequencies in their mRNAs (Lloyd and Sharp, 1992), making *S. cerevisiae* the ideal experimental model (Santos *et al.*, 1996; Silva *et al.*, 2004). Additionally, the *Saccharomyces* ancestor was ambiguous for a period of about 100 My, before the ser-tRNA<sub>CAG</sub> was lost in the lineage that originated the genus *Saccharomyces* (Massey *et al.*, 2003). The expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* results in ambiguous decoding of the CUG codon, due to competition between the endogenous leu-CUG decoder tRNA and the *C. albicans* ser-tRNA<sub>CAG</sub>, which decodes the same codon as serine. Therefore, the reconstruction of the *C. albicans* CUG reassignment pathway from leucine to serine in *S. cerevisiae* provides important insight on the impact of ambiguous codon decoding to cell physiology and evolution.



**Figure 16 – Working model highlighting possible consequences of genetic code ambiguity.** Ambiguity of the CUG codon causes proteome destabilization, with the synthesis of mutant proteins. This may result in changes in gene expression, triggering of the stress response, increasing morphological variability, and may also result in a hypermutagenic state that decreases genome stability.

Previous studies have shown that a structural change in the Ser-tRNA<sub>CAG</sub> was fundamental for the reassignment of the CUG codon from leucine to serine in *C. albicans* (Santos *et al.*, 1996). The presence of a purine at position 33 (G<sub>33</sub>) that replaces a conserved pyrimidine (generally U) in other tRNAs hints on the role of such alteration in CUG decoding (see chapter I.3.1. above). Indeed, this is the only known tRNA with a guanosine at position 33 of the anticodon loop (Steinberg *et al.*, 1993), which lowers the decoding efficiency of the tRNA and allows cell survival under low-level serine CUG decoding (Santos *et al.*, 1996).

To assess the effect of the mutation at position 33 on this genetic code alteration, variants of the *C. albicans* Ser-tRNA<sub>CAG</sub> gene were produced and expressed in *S. cerevisiae* (Santos *et al.*, 1996). Similarly, the *C. albicans* Ser-tRNA<sub>CAG</sub> gene variants with G and T at position 33 were expressed in diploid strains of *S. cerevisiae*, in the present work. These tRNA variants were chosen to represent the situations of low and high CUG ambiguity levels, respectively, since the replacement of G<sub>33</sub> by pyrimidines greatly enhances leucylation of the ser-tRNA<sub>CAG</sub> (Suzuki *et al.*, 1997). In other words, expression of the *C. albicans* ser-tRNA<sub>CAG</sub>-G<sub>33</sub> in *S. cerevisiae* likely introduces low-level ambiguity, whereas the ser-tRNA<sub>CAG</sub>-T<sub>33</sub> introduces high-level ambiguity in *S. cerevisiae*, mimicking distinct evolutionary stages of the CUG reassignment pathway. On both cases, expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* results in CUG ambiguity, since the CUG codon is translated both as leucine and serine. The direct consequence of CUG ambiguity is the production of aberrant proteins, which will have a major impact on the proteome, and ultimately affect genome stability, as highlighted above (Figure 16).

### **3.5. Objectives of this thesis**

The aim of this study was to re-construct the early events of the *C. albicans* genetic code alteration in *S. cerevisiae*, in order to determine the effect of ambiguous CUG decoding on gene expression, and also to shed new light on the impact of codon reassignment on cell physiology, and ultimately on the evolution of the genetic code. For this, the *C. albicans* ser-tRNA<sub>CAG</sub> was expressed in *S. cerevisiae* and the transcriptome, proteome and physiology of the engineered *S. cerevisiae* cell lines were characterized. The characterization of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> also provided a model for studying mRNA mistranslation in general.

## **II. Materials and Methods**

1. Strains and growth conditions
2. Transcriptome analysis
3. Proteome analysis
4. Physiological characterization

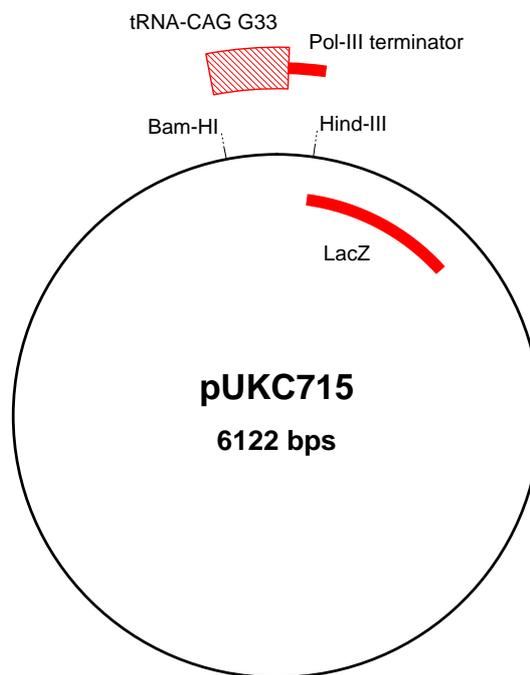


## 1. Strains and growth conditions

All experiments were performed using *Saccharomyces cerevisiae* CEN-PK2 (*MAT a / α, ura3-52 / ura3-52, trp1-289 / trp1-289, leu2-3, 112 / leu2-3, 112, his3Δ1 / his3Δ1*), except for proteasome purification (see chapter II.3.4.). Control cells were transformed with the single copy vector pRS315 (vector alone), and mutant strains bear the plasmids pUKC715 (Figure 17), containing the *Candida albicans* Ser-tRNA<sub>CAG</sub> G<sub>33</sub>, and pUKC716 (Figure 17) containing the *C. albicans* Ser-tRNA<sub>CAG</sub> T<sub>33</sub> (described in Santos *et al.*, 1996).

Transformation of *S. cerevisiae* CEN-PK2 was carried out using the lithium acetate method (Gietz and Woods, 1994). Briefly, 0.5 ml of stationary phase cells were inoculated into 10 ml (for 10 transformations) of fresh YEPD, pre-warmed at 30°C, and incubated at 30°C for approximately 4.5 hours until the OD<sub>600nm</sub> was 0.5-0.6. Cells were harvested at 4000 rpm for 3 minutes and the pellet resuspended in 5 ml of lithium acetate solution (10 mM Tris-Cl pH 7.5, 1 mM EDTA, and 0.1 M lithium acetate). After centrifugation, the supernatant was discarded and cells resuspended in 500 μl of lithium acetate solution and transferred to a microfuge tube. 100 μl of carrier DNA (sheared salmon sperm DNA at 5 mg/ml) were added into the solution, mixed and 60 μl of cells were aliquoted in 10 microfuge tubes. 0.1-1 μg of plasmid DNA was added to each tube and after addition of 0.5 ml of 40% PEG solution (40 % PEG 3300 in lithium acetate solution), mixing well, cells were incubated at 30°C for 45 min. Then cells were heat shocked at 42°C for 15 min and 100-200 μl of the cell transformation mix were plated directly onto a selective plate and incubated at 30°C for 2 days.

*S. cerevisiae* CEN-PK2 transformed cells were grown at 25°C, unless otherwise stated, in minimal medium containing 0.67 % yeast nitrogen base without amino acids, buffered at pH 5.8 with 1 % succinate and 0.6 % NaOH, and 2 % glucose. The medium was supplemented with 100 μg / ml of each of the required amino acids (uracil, tryptophan and histidine, as leucine is the plasmid marker).



**Figure 17 – Plasmid pUKC715 containing the *Candida albicans* Ser-tRNA<sub>CAG</sub> G<sub>33</sub>.** The plasmid was based on the single copy vector pRS315. pUKC716 is identical to pUKC715 with the exception that the Ser-tRNA<sub>CAG</sub> gene has a T at the N<sub>33</sub> position.

## 2. Transcriptome analysis

Transcriptome analysis was performed using DNA-microarray technologies (Schena *et al.*, 1995; Brown and Botstein, 1999). All protocols used are MIAME (minimum information about a microarray experiment)-compliant (Brazma *et al.*, 2001), and are available at [http://www.genomics.med.uu.nl/pub/jvp/ext\\_controls](http://www.genomics.med.uu.nl/pub/jvp/ext_controls) (UMC Utrecht Genomics Laboratory website).

### 2.1. Total RNA extraction

50 ml of exponentially growing cells ( $OD_{600nm} = 0.5$ ) were harvested by brief centrifugation at 4000 rpm for 3 minutes at room temperature. After a quick removal of the supernatant, tubes were immediately immersed on liquid nitrogen and frozen at  $-80^{\circ}C$ .

Total RNA was isolated from *S. cerevisiae* using hot phenol (Schmitt *et al.*, 1990). Frozen pellets were resuspended in Acid Phenol Chlorophorm (Sigma, 5:1, pH 4.7) kept at  $65^{\circ}C$  (500  $\mu$ l per 25 OD units of cells). The same volume of TES-buffer (10 mM Tris pH 7.5, 10 mM EDTA, 0.5 % SDS) was added and the tubes were vortexed for 20 seconds to resuspend the cell pellet. After 1 hour incubation in a water bath at  $65^{\circ}C$ , with 20 seconds vortexing every 10 minutes, the tube content was transferred to 1.5 ml Eppendorf tubes and centrifuged for 20 minutes at 14000 rpm at  $4^{\circ}C$ . The water-phase was added to a new Eppendorf tube, filled with 500  $\mu$ l Acid Phenol Chloroform (Sigma, 5:1, pH 4.7), vortexed for 20 seconds and centrifuged for 10 minutes at 14000 rpm at  $4^{\circ}C$ . The water-phase from this step was added to a new Eppendorf tube filled with 500  $\mu$ l Chloroform:Isoamyl-alcohol (Sigma, 25:1), vortexed for 20 seconds and centrifuged for 10 minutes at 14000 rpm at  $4^{\circ}C$ . Again, the water phase was transferred to a new Eppendorf tube with 50  $\mu$ l sodium acetate (3 M, pH 5.2), the tube was filled with ethanol (100 %, kept at  $-20^{\circ}C$ ) and incubated at  $-20^{\circ}C$  for 45 minutes to 1 hour. After RNA precipitation, tubes were centrifuged for 7 minutes at room temperature, 14000 rpm. The fluid

was removed carefully to avoid touching the RNA-pellet. The pellet was washed with 500  $\mu\text{l}$  ethanol (80 %, kept at  $-20^{\circ}\text{C}$ ) and centrifuged for 3 minutes at room temperature, 14000 rpm. After removal of all traces of ethanol, the RNA pellet was air dried for 1 minute and dissolved in sterile (mQ) water to a concentration of approximately 10  $\mu\text{g} / \mu\text{l}$  (100  $\mu\text{l}$  per 25 mid-log OD units, corresponding to about 1 mg total RNA). Samples were frozen using liquid nitrogen and kept at  $-80^{\circ}\text{C}$ .

## **2.2. mRNA enrichment (poly A<sup>+</sup> RNA isolation)**

mRNA was isolated using Oligotex (Qiagen), which consists of a suspension of beads covalently linked to dT oligonucleotides. The affinity of the oligo-dT for polyadenylic (poly A<sup>+</sup>) acid sequences allows the retrieval of RNA with poly-A tails, therefore enabling the purification of mRNAs from total RNA.

Total RNA samples were thawed and kept on ice. 1 mg from each sample was transferred to a fresh tube and Internal Hybridization Standards (external controls consisting of *Bacillus subtilis* RNA sequences to allow better image analysis, as well as normalisation for intra and inter-chip comparisons) were added as a diluted mixture of 10  $\mu\text{l}$  to 1 mg of total RNA. RNase-free water was added to a total volume of 575  $\mu\text{l}$ , and an equal volume of OBB buffer (20 mM Tris.HCl pH 7.5, 1 M NaCl, 2 mM EDTA, 0.2 % SDS) was added. After mixing the Oligotex Bead Suspension to assure homogenization, 70  $\mu\text{l}$  were added to each tube and samples were mixed gently by pipeting up and down. Samples were incubated in a water bath at  $65^{\circ}\text{C}$  for a maximum of 3 minutes, mixed again and incubated at room temperature for 10 minutes with occasional mixing. After this period, samples were centrifuged for 2 minutes at 14000 rpm, room temperature, and the supernatant was carefully removed (to avoid the removal of any beads, approximately 50  $\mu\text{l}$  of supernatant was left in the tubes). Beads were resuspended in 1 ml of OW2 buffer (10 mM Tris.HCl pH 7.5, 150 mM NaCl, 1 mM EDTA) and the tubes centrifuged for 2 minutes at 14000 rpm, room temperature. The supernatant was carefully removed (again, approximately 50  $\mu\text{l}$  of supernatant

was left in the tubes) and beads were washed again with 1 ml of OW2 buffer. After centrifugation for 2 minutes at 14000 rpm, room temperature, all traces of supernatant were removed with the pipette. Beads were resuspended in 175  $\mu$ l of OEB (5 mM Tris.HCl pH 7.5), incubated in a water bath at 65°C for 1 minute and centrifuged for 1 minute at 14000 rpm, room temperature. 155  $\mu$ l of eluate were carefully transferred to a fresh tube, to avoid the removal of any beads. Beads (kept in the first tube) were resuspended with 175  $\mu$ l of OEB, incubated in a water bath at 65°C for 1 minute and centrifuged for 1 minute at 14000 rpm, room temperature. 175  $\mu$ l of eluate was pooled with the eluate obtained with the previous OEB wash, giving a total volume of eluate of 330  $\mu$ l. To make sure this eluate did not contain any beads (they can interfere with the downstream reverse transcription reaction), it was further centrifuged for 5 minutes at 14000 rpm, room temperature, and 300  $\mu$ l of eluate were kept.

Purified mRNA was quantified from 10  $\mu$ l of eluate as described below. To the remaining 300  $\mu$ l of eluate, 1 / 10 of the volume (approx. 30  $\mu$ l) of 3 M NaOAc pH5.2 and 675  $\mu$ l of 100 % ethanol ( $-20^{\circ}\text{C}$ ) were added and incubated for 30 minutes at  $-80^{\circ}\text{C}$  to precipitate. Samples were then centrifuged for 30 minutes at 14000 rpm at 4°C. The supernatant was removed with a pipette and the pellet was washed with 500  $\mu$ l of 80 % ethanol ( $-20^{\circ}\text{C}$ ). After a 5 minute spindown at 14000 rpm, 4°C, the wash was completely removed with a pipette. The pellet was dried at room temperature and resuspended with an appropriate volume of mQ water to give a final concentration of 1  $\mu\text{g} / \mu\text{l}$ . RNA was snap frozen and stored at  $-80^{\circ}\text{C}$ .

## **2.3. cDNA synthesis**

### **2.3.1. RT reaction**

Complementary DNA synthesis was carried out in the presence of 2-aminoallyl-dUTP (aa-dUTP, SIGMA), and for each sample a negative control reaction was always prepared without reverse transcriptase. The RNA/primer mix was prepared on ice with 3 µg of yeast mRNA (3 µl of 1 µg/µl mRNA from previous step) and 1.5 µg Oligo dT<sub>12-18</sub> primer (6 µl of 0.25 µg/µl stock, Amersham), diluted in 4 µl water and incubated at 70 °C for 10 minutes. During this incubation the labelling mix was prepared, on ice, by adding in the following order: 3 µl dGAC-mix (dGTP, dATP and dCTP, 1 mM each), 0.9 µl dTTP (1 mM), 2.1 µl aa-dUTP (1 mM, SIGMA), 6 µl 5x first strand buffer (250 mM Tris-Cl pH 8.3, 375 mM KCl, 15 mM MgCl<sub>2</sub>) and 3 µl DTT (0.1M). After a short spindown of the RNA/primer mixture, samples were chilled on ice for 5 minutes and the labelling mix was added. The tubes were incubated at room temperature for 2 minutes and 2 µl of SuperScript™ II Reverse Transcriptase (200 U / µl, Gibco) were added. The reaction was allowed to proceed at 42°C for 60 minutes.

The RNA template was then removed by hydrolysis. After a 2-minute incubation at 95°C, tubes were chilled on ice immediately. 10 µl of NaOH (1M) and 10 µl of EDTA (0.5 M) were mixed and added to each cDNA reaction, incubating in a water bath at 65°C for 15 minutes. After a short spindown, 25 µl of HEPES buffer (1M, pH 7.5) was mixed with the samples to neutralize the reaction.

### **2.3.2. Cleanup with Microcon-30 concentrators**

In order to couple the amino-allyl dyes, the Tris buffer must be removed from the reaction to prevent the monofunctional NHS-ester Cy-dyes coupling to free amine groups in solution. For this, samples were purified using Microcon-30 columns (Millipore). Each column was filled with 450 µl of mQ water, and the neutralized

reaction from the previous step was added. After spinning for 10 minutes at 10000 rpm, room temperature, the flow-through was discarded and this process repeated 3 more times, refilling the original filter with 500  $\mu$ l of mQ water. Finally, samples were eluted by placing the filter upside down in a new tube and spinning at 3000 rpm for 3 minutes. To determine the amount of cDNA produced, 1  $\mu$ l of eluate was taken for quantification as described below (chapter II.2.5.).

## **2.4. cDNA labelling**

### **2.4.1. Coupling monofunctional NHS-ester Cy-dyes**

Purified cDNA samples were concentrated to 8  $\mu$ l by using a SpeedVac (if samples were below 8  $\mu$ l, mQ water was added as required), and 1  $\mu$ l of Na<sub>2</sub>CO<sub>3</sub> buffer (0.5 M, pH 9) was added. Monofunctional NHS-ester Cy3 and Cy5 dyes (Amersham) were resuspended in DMSO, and 1.25  $\mu$ l were added per sample (Cy3 or Cy5, according to the reactions previewed). After quick mixing, samples were incubated in the dark at room temperature for 1 hour.

Before combining Cy3 and Cy5 samples for hybridization the reactions were quenched, to prevent cross coupling, by addition of 4.5  $\mu$ l of 4 M hydroxylamine (SIGMA) and incubated in the dark at room temperature for 15 minutes.

### **2.4.2. Cleanup with Chromaspin-30 columns**

Before hybridization, free dyes were removed using Chromaspin-30 columns (DEPC version, Clontech). The gel matrix was completely resuspended by inverting each column several times before use. The end of the spin column was placed into a 2 ml microcentrifuge tube and centrifuged at 1500 rpm for 5 minutes (in an Eppendorf 5810 centrifuge). The collection tube (containing the column equilibration buffer) was discarded and the spin column was placed into another 2 ml microcentrifuge tube. Each sample was slowly and carefully applied to the

center of the gel bed's surface of a column, drop by drop, and then centrifuged at 1500 rpm for 5 minutes. The spin columns were detached from the collection tubes, where the purified samples remain, and the tubes were immediately stored on ice and dark. 1  $\mu$ l of eluate was kept to determine the amount of cDNA produced and frequency of dye incorporation (see below).

## 2.5. RNA quantification

All measurements were done in a Shimadzu UV1240 mini spectrophotometer. For total RNA quantification, 5  $\mu$ l of total RNA were added to 45  $\mu$ l of mQ water (1:10), and 5  $\mu$ l of this dilution were added to another tube with 495  $\mu$ l of mQ water (1:1000).  $OD_{260nm}$  was measured and the amount of RNA was calculated considering that 1  $OD_{260nm}$  unit corresponds to 40  $\mu$ g/ml. To quantify the purified mRNA,  $OD_{260nm}$  was measured in 10  $\mu$ l of eluate diluted with 60  $\mu$ l mQ water (1:7) using a 70  $\mu$ l cuvette and the amount of mRNA calculated as above.

The amount of cDNA produced was determined before and after coupling the dyes, using a 5  $\mu$ l cuvette. Before coupling, 1  $\mu$ l of eluate was taken and diluted in 4  $\mu$ l mQ water and the OD was measured from 190 to 330 nm. After coupling, 1  $\mu$ l of eluate was diluted in 4  $\mu$ l mQ water and used for spectrophotometry (190 to 750 nm range) to determine the amount of cDNA generated and the frequency of dye incorporation.

Considering that 1  $OD_{260nm}$  unit of ssDNA = 37 ng/ $\mu$ l, the amount of cDNA synthesized was calculated as follows:  $OD_{260nm} \times 37 \times \text{total volume of probe } (\mu\text{l}) = \text{ng of probe}$ , and the negative control was subtracted from this result.

Taking into account the molar extinction coefficients of Cy3<sup>TM</sup> and Cy5<sup>TM</sup> (150 000 / M cm at  $OD_{550nm}$  and 250 000 / M cm at  $OD_{649nm}$  respectively), the quantity of dye incorporated was calculated as follows:  $OD_{550nm} \times \text{volume of probe } (\mu\text{l}) / 0.15 = \text{pmol of Cy3}^{\text{TM}}$  dye incorporated, and  $OD_{649nm} \times \text{volume of probe } (\mu\text{l}) / 0.25 = \text{pmol of Cy5}^{\text{TM}}$  dye incorporated. The negative controls were subtracted from these results.

The frequency of dye incorporation (number of dye-labelled nucleotides per 1000 nucleotides) was calculated as follows:  $\text{pmol of dye incorporated} \times 324.5$  (average molecular weight of one kb of DNA in g/mol) / ng of probe.

## **2.6. Microarray hybridization**

### **2.6.1. Enhanced prehybridization**

100 ml of borohydride buffer containing 2 x SSC (0.15 M NaCl and 0.015 M sodium citrate, pH 7.0), 0.05 % SDS and 0.25 % sodium borohydride were prepared by mixing only once to prevent foaming. The buffer was heated to 42°C in a 100 ml Coplin Jar without lid (to prevent SDS overflow). Meanwhile, 50 ml Falcon tubes were filled with a buffer containing 2x SSC and 0.05 % SDS, at room temperature, and each slide was washed in one tube by rigorous shaking for about 1 minute, to remove excess unbound DNA, thus preventing formation of comet tails. The slides were incubated in the borohydride buffer for 30 minutes at 42°C, and prehybridization buffer was prepared (100 ml for a maximum of 4 slides) containing 5x SSC, 25 % formamide, 0.1 % SDS and 1 % BSA. This solution was filtered through a 0.22 micron syringe filter and heated to 42°C in a Coplin Jar. The slides from the borohydride incubation were washed by dipping 5 times in room temperature mQ water, then placed into the pre-heated prehybridization buffer and incubated for 45 minutes at 42°C. After prehybridization, slides were washed by dipping 5 times in room temperature mQ water and blow-dried using compressed nitrogen (air gun). Slides were immediately used following this procedure.

### **2.6.2. Hybridization**

5 ml of 2x hybridization buffer (50 % formamide, 10x SSC and 0.2 % SDS) were prepared and immediately filtered through a 0.22 micron filter to prevent SDS precipitation. 250 µl of the 2x hybridization buffer (enough for 6 slides) were taken and 5 µl of Herring sperm DNA (stock 10 µg / µl, sheared) were added, for a final concentration in 2x hybridization buffer of 200 µg/ml. The solution was then preheated to 42°C to overcome SDS precipitation. 40 µl of target (combined Cy3 and Cy5 samples, consisting of 300 ng of cDNA from each sample) and 40 µl of the pre-heated 2x hybridization buffer were mixed and heated at 95°C for 5 minutes, followed by centrifugation at 12000 rpm for 2 minutes. Cleaned LifterSlips

(Eyrie Scientific) were blow-dried using compressed nitrogen and placed with the Teflon side down over the array area of the slides. With the pipette tip placed along an open edge of the LifterSlip, the 80  $\mu$ l of the target were slowly pipeted out of the tip and by capillary action the probe mix was drawn under the slip. The slide was placed inside a Corning Hybridization Chamber (Corning), 20  $\mu$ l of water were added to each well and the chamber was closed with clamps. The hybridization chamber was then gently placed into a 42°C water bath for 16-20 hours.

### **2.6.3. Washing hybridized arrays**

Each slide was carefully removed from the hybridization chamber, without disturbing the coverslip and washed as follows. A low-stringency wash was performed by placing the slides in a Coplin Jar containing 100 ml of 1x SSC and 0.2 % SDS. The coverslips were gently removed while the slides were in solution. LifterSlips were washed as described below, and the slides were incubated for 4 minutes at room temperature in washing buffer. After this step, a high-stringency wash was performed by placing the slides in a Coplin Jar containing 100 ml of 0.1x SSC and 0.2 % SDS, and incubated for 4 minutes at room temperature. Finally, slides were washed in 100 ml of 0.1x SSC to remove particles of SDS and incubated for 4 minutes at room temperature. This last wash was repeated 2 times, and then the slides were blow-dried using compressed nitrogen.

### **2.7. LifterSlip cleaning**

LifterSlips were cleaned before and after use. Before use, two Falcon tubes were filled with 50 ml mQ water and 50 ml of 100 % ethanol, respectively. Each lifter slip was firmly held with flat-headed forceps, dipped five times in the water and immediately dipped five times in ethanol. After washing, the lifterslips were blow-dried using compressed nitrogen. After using the LifterSlips, four 50-ml Falcon tubes, two containing distilled water with a little soap and two filled with distilled water, and a beaker with mQ water were prepared. Each lifter slip was firmly

helded with flat-headed forceps, dipped five times in the first soap tube, five times in the second soap tube, five times in the first water tube, and then five times in the second water tube. Finally the lifterslip was placed in the beaker containing mQ water. After washing all the lifterslips, they were rinsed and dried as described above in “Before use” and stored in a lifterslip box.

## **2.8. Microarray production**

C6-amino-linked oligonucleotides (70 nucleotides in length) and the Yeast Genome ArrayReady were purchased from Qiagen. Printing was performed on Corning UltraGAPS slides with a MicroGrid II (Apogent Discoveries) using 48-quill pins (Microspot2500, Apogent Discoveries), in 3x SSC at 50 % humidity and 18°C. After printing, slides were processed by ultraviolet crosslinking (2,400 millijoules, 10 minutes) with a Stratalinker2400 (Stratagene).

## **2.9. Data analysis**

Slides were scanned and image analysis was carried out using Imagene 4.0 (Biodiscovery). This program was used for spot detection, as it offers the possibility to select and flag the spots that should not be used in posterior analysis, and also quantifies each spot. Quantified data was pre-normalized in QQCC (quick quality check for chips), a program from the Genomics lab in the UMC Utrecht. Normalization is required to correct differences in labelling efficiency, hybridizations and fluorescence intensities. QQCC is based on Lowess (locally weighted least squares regression) print-tip normalization (Yang and Speed, 2002; Yang *et al.*, 2002), which defines windows of spots and performs linear regressions within these windows. Algorithms were altered for the import of Imagene 4.0 files, flagging of control spots, Lowess line calculation on subsets of spots and extrapolation to all spots in the subgrid. The method of normalization chosen was based on the expression levels of endogenous genes, assuming that

the overall change is balanced (there are as many genes induced as repressed), and uses gene spots to calculate the Lowess line for each subgrid and then applies these lines to all spots. As a quality control, the QQCC output displays MA plots [ $M = \log_2(R/G)$  and  $A = 1/2 \log_2(RG)$ , where R and G are the red and green channel intensities, respectively] for the raw data and normalized data based on genes only, genes and external controls and controls only, and also graphics with the signal-to-noise ratio and the saturated spots.

Data was then viewed and analyzed using GeneSpring, where it was normalized for both intra and inter-chip comparisons. The mode of analysis chosen was log of the ratios, as the log transformation assures a symmetric distribution of the up and down-regulated genes and with a normal distribution it is possible to perform statistical tests. The statistical significance of the results was obtained with SAM (significance analysis for microarrays). This method was developed to overcome the excessive number of false positives obtained when conventional t-tests are applied to microarray experiments, where for example a significant value of  $p=0.01$  in 10 000 genes would identify 100 genes by chance (Tusher *et al.*, 2001). To avoid this situation, instead of calculating a p-value for each gene the SAM method defines a single value that is verified by all genes, and this error probability is the false discovery rate (FDR). To identify potentially significant genes, the observed and the expected relative difference in expression is plotted for each gene. If the distance between the two values is greater than a threshold  $\Delta$  the gene is considered significant, and the bigger this distance the more significant a gene is, for the same error probability (Tusher *et al.*, 2001).

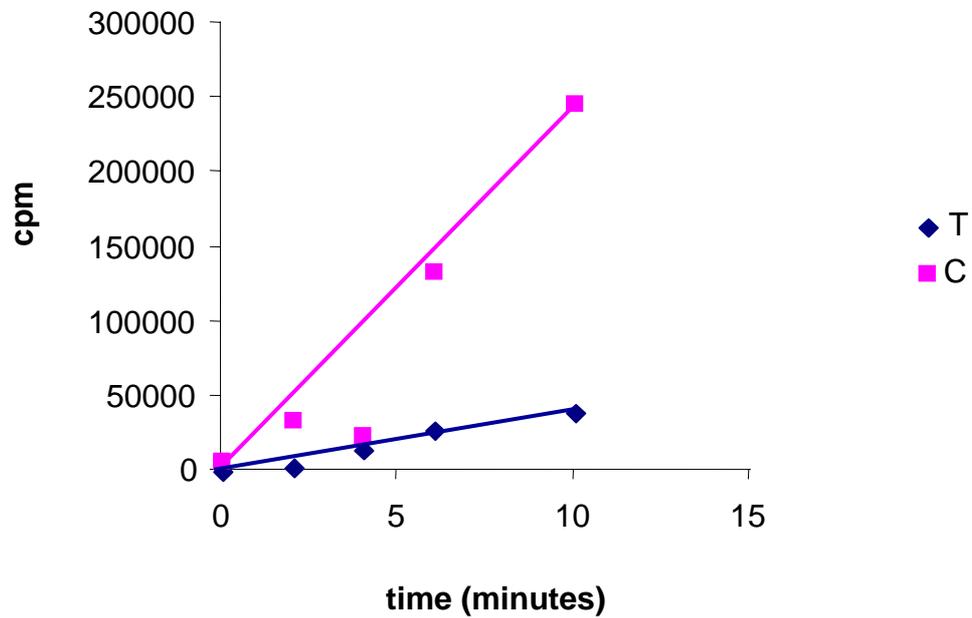
### 3. Proteome analysis

#### 3.1. Kinetics of <sup>35</sup>S-Methionine incorporation

A kinetic assay of <sup>35</sup>S-methionine incorporation was performed, in order to determine the optimal time for *in vivo* cell labelling. For this, samples of 0.5 ml of each culture with an OD<sub>600nm</sub> of 0.5 were taken into a tube containing 60 x 10<sup>-8</sup> M methionine (24 µl of 1.25 x 10<sup>-5</sup> M cold methionine plus 3 µl of radioactive methionine) and were carefully vortexed. 20 µl of cells were taken to new ice-cold tubes, 2.5 ml of 5 % TCA with 1 g / l of cold methionine were added and the tubes were vortexed and placed on ice. This process was repeated at times 2, 4, 6 and 10 minutes. After 30 minutes on ice, tubes were placed on a water bath at 85°C for 10 minutes and then allowed to cool on ice.

A vacuum filtering system was assembled, placing on each channel a glass microfibre filter (GF/C, Whatman) soaked in the TCA / Met solution. Channels were washed with TCA before applying the samples. Each tube was carefully vortexed and its content loaded into one TCA washed channel. The tubes were washed with 2.5 ml TCA / Met, vortexed and loaded into the respective channel. After filtering, TCA / Met was added again to wash the channels, and a little quantity of ethanol was added to dry the filters more quickly. Channels were disassembled and the filters were placed on a hot plate to dry. After drying, each filter was placed in 5 ml scintillation liquid (Ready value, Beckman) and the radioactivity incorporation counted in a scintillation counter (Monribot and Boucherie, 2000).

The time determined for the *in vivo* labelling of cells corresponds to 10 minutes. This time is required for the maximum incorporation of <sup>35</sup>S-methionine while keeping a linear uptake of radioactivity (Figure 18).



**Fig 18 - Kinetics of  $^{35}\text{S}$ -methionine incorporation.** Samples of exponentially growing ( $\text{OD}_{600\text{nm}}$  of 0.5) cells were taken into a tube containing radioactive methionine and processed as described above. The plot shows the radioactivity incorporation counted in a scintillation counter corresponding to 0, 2, 4, 6 and 10 minutes of labelling. The results for *S. cerevisiae* control cells are shown in pink, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> are displayed in blue.

## **3.2. Radioactive labelling**

**Growth at 25°C** - Samples of 2 ml of each culture with an OD<sub>600nm</sub> of 0.5 were transferred to tubes containing 300 or 600 µCi of <sup>35</sup>S-Methionine (Amersham) and labelled for 10 minutes. Cells were harvested by centrifugation at 13000 rpm for 2.5 minutes, and the supernatant was discarded. The cell pellet was washed twice with ice-cold mQ water, by centrifugation at 13000 rpm for 2.5 minutes, and resuspended in a minimum volume of ice-cold mQ water (approximately 50 µl). The tubes were closed with a perforated lid and stored at – 80°C (Boucherie *et al.*, 1995).

**Growth at 37°C** - *S. cerevisiae* cells were grown at 37°C with shaking to an OD<sub>600nm</sub> of 0.5. Samples of 2 ml culture were transferred to a new tube, pre-heated at 37°C, and labelled as described above.

**Heat Shock** - *S. cerevisiae* cells were grown at 25 °C with shaking to an OD<sub>600nm</sub> of 0.5. Samples of 2 ml culture were transferred to a new tube, pre-heated at 37°C. After 20 minutes in a water bath at 37°C, cells were labelled for 15 minutes as described above.

## **3.3. 2D-PAGE**

### **3.3.1. Protein extraction**

After lyophilization of the frozen pellets for 4 hours, cells were disrupted on a MiniBeadBeater (Biospec Products) in the presence of 50 mg of glass beads, shaking the tubes for 5 cycles of 20 seconds followed by 20 seconds on ice. Proteins were then solubilized by successively adding 60 µl of extraction buffer (0.1 M Tris-Cl pH 8.0, 0.3 % SDS), 1.65 µl β-mercaptoethanol (final concentration 2.5 %, v/v), 110 mg of urea and 33 µl of sample buffer (4.75 M urea, 4 % CHAPS, 1 % Pharmalytes 3-10 and 5 % β-mercaptoethanol). Protein samples were kept at room temperature for 5 minutes to allow protein dissolution, then centrifuged for 3 minutes at 13000 rpm and the supernatant was distributed in 30 µl aliquots (2 µl

were kept for radioactivity incorporation measurements, as described below). Protein extracts were stored at  $-80^{\circ}\text{C}$  or immediately loaded on first-dimension gels (Boucherie *et al.*, 1995; Monribot and Boucherie, 2000).

### **3.3.2. Radioactivity counting**

2  $\mu\text{l}$  of protein extract were applied on a glass microfibre filter (GF/C, Whatman) and allowed to dry at room temperature. Filters were soaked for 10 minutes in 2.5 ml of 5 % TCA (to precipitate proteins) containing 1 g / L of cold methionine (the same amino acid used for labelling, to remove excess radioactive methionine not incorporated into protein), dried at room temperature and placed in a counting vial containing 5 ml of scintillation liquid (Ready value, Beckman). Counting was performed using a scintillation counter (Monribot and Boucherie, 2000).

### **3.3.3. First dimension**

Proteins were separated in the first dimension by isoelectric focusing (IEF) as described (Boucherie *et al.*, 1995; Monribot and Boucherie, 2000). Isoelectric focusing was carried out on gel rods (24 cm long and 1 mm diameter), and the proteins were separated along a pH gradient ranging from 4.5 to 7. First-dimensional acrylamide gel solution was prepared in a corex tube (final composition 3.4 % acrylamide, 0.17 % bisacrylamide, 9.5 M urea, 3.6 % CHAPS, 4 % ampholytes). After urea dissolution by gentle mixing, the solution was deaerated under vacuum for 3 minutes. Polymerization was initiated by adding 20  $\mu\text{l}$  of 10 % APS (TEMED was not required for polymerization). Glass tubes, previously placed at  $26^{\circ}\text{C}$ , were filled individually using a pipette-pump, according to Monribot and Boucherie (2000).

Gels were loaded with 15  $\mu\text{l}$  of loading buffer (0.3 M Tris-Cl pH 8, 0.1 % SDS, 0.8 %  $\beta$ -mercaptoethanol, 0.65 % CHAPS, 0.3 % Pharmalytes 3-10, 9.5 M urea) and pre-run as follows: 15 minutes at 500 V and 45 minutes at 1000V. After pre-

running, 10  $\mu$ l of protein sample were loaded, overlaid with 15  $\mu$ l of overlay buffer (0.5 % Pharmalytes 3-10, 2 % CHAPS, 2.4 M urea) and gels were run for 15 minutes at 500 V, followed by 45 minutes at 1000V and 21 hours at 1600 V. Anode and cathode solutions were 0.08 M phosphoric acid and 0.1 M NaOH, respectively. Pre-focusing and focusing were performed at 26°C.

After focusing, gels were sorted out of the glass tubes with the help of a 2.5 ml syringe filled with mQ water and fitted with a yellow pipette tip. The water was introduced from the top end of the tube, between the gel and the interior wall of the tube, keeping a constant pressure on the syringe to allow the gel to sort from the bottom of the tube, slowly and carefully not to break the gel or stretch it. The gels were directly recovered to a piece of parafilm, wrapped in aluminium foil and stored at – 80°C until use.

#### **3.3.4. Second dimension**

The second dimension was carried out according to Boucherie et al. (1995). Gels were run on a vertical slab gel with 90 cm large, allowing the running of three first-dimension gels in parallel. Second-dimension gels contain 11 % acrylamide (ratio 29:1) and are devoid of SDS, resolving proteins with molecular weights that range from 180 to 17 kDa.

Gel solution (0.36 M Tris-Cl pH 8.5, 10.6 % acrylamide, 0.35 % bisacrylamide) was prepared in a vacuum flask and deaerated for 3 minutes before initializing polymerization, by adding 1 ml of 10 % APS and 136  $\mu$ l of TEMED. The gel solution was pumped into the assembled cassette with a 50 ml syringe, gently overlaid with mQ water and allowed to polymerize overnight at room temperature.

First dimension gels were not equilibrated prior to the second dimension and were directly applied on top of second dimension gels, without agarose solution. With

the help of a blunt-ended spatula and some drops of running buffer, the first dimension gel was pushed between the glass plates until it reached the top of the slab gel. Care was taken not to trap air bubbles between the IEF gel and the surface of the second dimension gel.

Electrophoresis was performed at 5W for 15 minutes and then at 25W for approximately 6 hours. Electrophoresis buffer was 25 mM Trizma base, 192 mM glycine and 0.2 % SDS. After electrophoresis gels were fixed by immersion in a solution of 50 % ethanol and 7.5 % acetic acid, overnight and with gentle agitation. Then gels were placed for about 1 hour in 25 % ethanol with 2.5 % acetic acid. After fixation, gels were dried (in a Savant stacked gel dryer) and exposed to phosphorous screens (MolecularDynamics). Radioactive protein spots were detected using a PhosphorImager (MolecularDynamics) (Boucherie *et al.*, 1995; Monribot and Boucherie, 2000).

### **3.3.5. Materials washing**

All plastic and glassware was washed or rinsed only with mQ water, before and after 2D electrophoresis, with the exception of the first dimension tubes. These were cleaned in sulphochromic acid for 30 minutes and then thoroughly washed with tap water. After, they were immersed in KOH for 30 minutes and washed with tap water under continuous flow for at least 30 minutes. Finally, tubes were rinsed for 2 hours with a continuous flow of mQ water and dried with compressed air.

### **3.3.6. Data analysis**

2D-PAGE data analysis was performed using ImageMaster software (Amersham Pharmacia Biotech). The best 3 gels from each condition were selected and the spots were detected, matched between all images and quantified. For normalization, each spot was divided by the total spot volume present in each gel, to compensate eventual loading or exposure differences, and only the spots with 6 matches, that is, present in all images analyzed, were considered.

For each protein spot, a medium expression level was calculated, as well as the ratio between each two given conditions. Variances were tested by performing an F-test, and when values were similar a student T-test was applied between the medium expression levels of the two conditions analyzed. If the variances were different, the Welch test was used. Only proteins with more than a 2-fold variation and whose medium expression level was considered significantly different in the statistical test applied were taken into account. Spots whose expression levels were too low were left out of the final reference maps due to high variability between gels.

Protein identification in the 2D-maps was based on pre-existent maps (Boucherie *et al.*, 1995; Shevchenko *et al.*, 1996; Sagliocco *et al.*, 1996; Perrot *et al.*, 1999), available at <http://www.ibgc.u-bordeaux2.fr/YPM/carte.htm>

### 3.4. Proteasome purification

Proteasome subunits were purified by the TAP-tag approach (Rigaut *et al.*, 1999; Puig *et al.*, 2001) using the *S. cerevisiae* strains SCO973 and SCO832 (purchased from EUROSCARF), tagged at the C-terminal of *SCL1* (20S) and *RPN5* (19S) genes, respectively. Cells were grown at 30 °C in YEPD (1 % yeast extract, 2 % peptone, 2 % glucose) and purification was done from 6 liters of yeast cultures as described below.

#### 3.4.1. Protein extraction

Total protein extracts were prepared from cells grown to an OD<sub>600nm</sub> of 1-1.5 and collected by centrifugation at 4000 rpm, for 15 minutes, at 4°C. The pellet was washed with ice-cold mQ water (5 ml per 250 ml culture), centrifuged again at 4000 rpm, for 10 minutes, at 4°C, and frozen at – 80°C. Extracts were prepared in lysis buffer (10 mM K-Hepes pH 7.9, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.5 mM DTT, 0.5 mM PMSF, 2 mM benzamidine, 1 μM leupeptin, 2 μM pepstatin A, 4 μM chymostatin, 2.6 μM aprotinin) in the presence of glass beads. Cells were disrupted by shaking the tubes in a vortex, for 5 cycles of 1 minute followed by 1 minute on ice. The extract was centrifuged at 25 000 x g for 30 minutes at 4°C, using the 80 Ti Rotor in a Beckman Ultracentrifuge Optima LE-80K, and the supernatant clarified at 100 000 x g for 1 hour. The protein extract (middle phase) was carefully recovered and dialyzed against dialysis buffer (20 mM K-Hepes pH 7.9, 50 mM KCl, 0.2 mM EDTA pH 8.0, 0.5 mM DTT, 20 % glycerol, 0.5 mM PMSF, 2 mM benzamidine), for 3 hours at 4°C. After dialysis, the extract was frozen and kept at – 80°C (Puig *et al.*, 2001).

### **3.4.2. Complex purification**

The proteasome macro-complex was purified by two consecutive affinity purification steps using IgG Sepharose (SIGMA) and Calmodulin beads (Pharmacia) as described (Puig *et al.*, 2001).

200  $\mu$ l of IgG Sepharose bead suspension (SIGMA), corresponding to 100  $\mu$ l of beads, were transferred into a 10 ml column (BioRad) and washed with 10 ml of washing buffer (10 mM Tris-Cl pH 8.0, 150 mM NaCl, 0.1 % NP-40). The composition of the extract buffer was adjusted to 10 mM Tris-Cl pH 8.0, 100 mM NaCl, and 0.1 % NP-40 (for a maximum final volume of 10 ml), loaded into the column containing the washed sepharose beads and rotated for 2 h at 4°C, to allow binding of the complexes. After elution of unbound proteins by gravity flow, the column beads were washed 3 times with 10 ml of washing buffer and once with cleavage buffer (10mM Tris-Cl pH 8.0, 150 mM NaCl, 0.1 % NP-40, 0.5 mM EDTA, 1 mM DTT). Cleavage was done in the same column by adding 1 ml of cleavage buffer and 100 units of TEV protease (Invitrogen). The beads were rotated for 2 hours at 16°C and the eluate (containing the complexes) recovered by gravity flow.

200  $\mu$ l of calmodulin bead suspension (Pharmacia), corresponding to 100  $\mu$ l of beads, were transferred into a new column and washed with 10 ml of binding buffer (10 mM Tris-Cl pH 8.0, 10 mM  $\beta$ -mercaptoethanol, 150 mM NaCl, 1 mM magnesium acetate, 1 mM imidazole, 0.1 % NP-40, 2 mM  $\text{CaCl}_2$ ). 3 ml of binding buffer and 3  $\mu$ l of 1M  $\text{CaCl}_2$  were added to 1 ml of eluate recovered after TEV cleavage. This solution was loaded into the column containing washed calmodulin beads and rotated for 1hour at 4°C. The beads were washed 3 times with 10 ml of binding buffer, and bound proteins were eluted with 1 ml of elution buffer (10 mM Tris-Cl pH 8.0, 10 mM 2-mercaptoethanol, 150 mM NaCl, 1 mM magnesium acetate, 1 mM imidazole, 0.1 % NP-40, 2 mM EGTA). Purified complexes were separated by 1D or 2D-gel electrophoresis as described above.

### 3.4.3. Column regeneration

After elution, the columns were left in washing or elution buffers to avoid drying the beads. Columns containing IgG beads were washed with 10 column volumes (100 ml) of IgG regeneration solution-1 (0.1 M glycine, 0.15 M NaCl, pH 2.4), then with 1 volume (10 ml) of IgG regeneration solution-2 (0.01 M sodium phosphate buffer, 0.5 M NaCl, pH 7.2) and stored with 1 ml of regeneration solution-2, at 4°C, wrapped in parafilm. Columns containing calmodulin beads were washed with 3 column volumes (30 ml) of calmodulin regeneration solution-1 (50 mM Tris-Cl pH 7.5, 2 mM EGTA, 1 M NaCl), then with 1 volume (10 ml) of calmodulin regeneration solution-2 (calmodulin binding buffer, 2 mM CaCl<sub>2</sub>) and stored with 1 ml of regeneration solution-2, at 4°C, wrapped in parafilm.

## **4. Physiology**

### **4.1. Cell viability assays**

Cells were grown at 30°C with shaking to stationary phase and kept in culture for a period of about 70 days with weekly sampling. At each time point, aliquots of the culture were taken and the cells were counted in a haemocytometer. The number of cells per ml was calculated and cells were plated at appropriate dilutions to form about 50 to 100 colonies per plate. After growth, colonies were counted and the colony forming units were calculated as a percentage of the cells plated.

### **4.2. Proteasome activity assay**

The proteasome chymotrypsin-like activity was assayed using the fluorogenic peptide succinyl-leucine-leucine-valine-tyrosine-MCA (s-LLVY-MCA) as substrate (Grune *et al.*, 1998; Demasi *et al.*, 2003). For this, 50 ml of cells were grown at 30°C to an OD<sub>600nm</sub> of 0.5 (exponential phase) or 3.5 (stationary phase) and harvested by centrifugation at 4000 rpm, 4°C, for 10 minutes. The cell pellet was washed with 1 ml ice-cold mQ water and frozen at –80°C. Cells were disrupted on a MiniBeadBeater (Biospec Products) as described above, in lysis buffer containing 10 mM K-Hepes, 10 mM KCl and 1.5 mM MgCl<sub>2</sub>. Protein extracts were kept at –80°C until further use.

For activity assays, 100 µg of protein extracts were incubated at 37°C in assay buffer (10 mM Tris-Cl, pH 8, 20 mM KCl 1M and 5 mM MgCl<sub>2</sub>), and the reaction was started by addition of s-LLVY-MCA (diluted in 10 % DMSO) to a 50 µM final concentration. After 60 minutes, fluorescence of the proteolytically released MCA was measured at 435 nm emission (excitation at 365 nm), in a Perkin Elmer Luminescence Spectrometer (LS 50B). Final activity was calculated as fluorescence emission at time 60 min subtracted from fluorescence emission at time 0 min (Grune *et al.*, 1998; Demasi *et al.*, 2003).

### 4.3. Trehalose and glycogen quantification

Trehalose and glycogen were quantified as described by Parrou and François (1997). Briefly, cells were grown at 30°C to an OD<sub>600nm</sub> of 0.5. For heat shock samples, cultures were grown at 25°C to an OD<sub>600nm</sub> of 0.5, transferred to a new tube, pre-heated at 37°C, and grown for further 30 minutes at 37°C. 2 ml samples from each culture were harvested by centrifugation at 5000 rpm, 4°C for 5 minutes. The supernatant was discarded and the cell pellet was resuspended in 250 µl Na<sub>2</sub>CO<sub>3</sub> (0.25 M), and placed at 95°C for 4 hours with agitation to ensure cell lysis. After this incubation, the pH was lowered by addition of 150 µl acetic acid 1 M, and 600 µl sodium acetate 0.2 M, pH 5.2. The final volume was aliquoted into 2 new eppendorf tubes, with approximately 500 µl each. For trehalose determination, 0.05 U / ml of trehalase (SIGMA) were added and the reaction was left overnight at 37°C, with agitation. For glycogen assay tubes were incubated overnight at 57°C, with agitation, in the presence of 1.2 U / ml amyloglucosidase (MERCK). Reactions were centrifuged for 3 minutes at 5000 rpm and glucose determination on the supernatant was carried out using a glucose oxidase/peroxidase kit from Sigma. Briefly, the glucose present on the sample was oxidized to gluconic acid and hydrogen peroxide in the presence of glucose oxidase. The hydrogen peroxide then reacted with o-dianisidine in the presence of peroxidase forming a coloured product. Finally, oxidized o-dianisidine was allowed to react with sulphuric acid to form a more stable coloured product whose colour intensity, which is proportional to the original glucose concentration on the samples, was determined. For this, 200 µl of an adequately diluted sample were mixed with 400 µl of the reagent mixture containing glucose oxidase/peroxidase and o-dianisidine. Tubes were incubated for 30 minutes in a water bath at 37°C and the reaction was stopped by addition of 400 µl H<sub>2</sub>SO<sub>4</sub> 12 N. Absorbance measurements were done at 540 nm in a Beckman DU530 spectrophotometer. The glucose content was determined based on a calibration curve using glucose as standard (Parrou and Francois, 1997).

## **4.4. SDS-PAGE**

### **4.4.1. SDS-PAGE**

Proteins in all extracts were quantified using the Bradford reagent from Bio-Rad, having bovine serum albumin as standard (Bradford, 1976).

Protein extracts were separated on a 12 % resolving acrylamide gel with a 4 % stacking gel (Laemmli, 1970). The resolving acrylamide gel was prepared by mixing 1.1 ml water, 50  $\mu$ l SDS 10 %, 1.875 ml Tris-Cl (1 M, pH 8.0), 2 ml Acrylamide 30 % (29:1), 25  $\mu$ l APS 10 % and 5  $\mu$ l TEMED. The stacking gel was prepared by mixing 1.65 ml water, 25  $\mu$ l SDS 10 %, 0.5 ml Tris-Cl (0.625 M, pH 6.8), 0.332 ml Acrylamide 30 % (29:1), 25  $\mu$ l APS 10 % and 5  $\mu$ l TEMED.

Samples were diluted in 2 or 3  $\mu$ l of 6x sample buffer (30 % glycerol, 10 % SDS, 0.6 M DTT and 0.012 % bromophenol blue in 0.5 M Tris-Cl / 0.4 % SDS, pH 6.8), to a final volume of 12 or 18  $\mu$ l, and boiled for 1 minute before loading on the gel. Low Molecular Weight (Amersham) and Pre-stained markers (SIGMA) were used for stained and blotted gels, respectively. Gels were run on a Bio-Rad mini-gel apparatus, at 50 V for about 1 hour (stacking) and then at 100-150 V for about 2 hours, until the front of the migration reached the bottom of the gel, in a electrophoresis buffer containing 25 mM Tris, 192 mM Glycine and 0.2 % SDS. After electrophoresis, gels were stained or blotted according to the protocols described below. The gel images were acquired in a densitometer and analyzed with the QuantityOne software (BioRad) for 1D gels.

#### **4.4.2. Coomassie staining**

Coomassie Blue stain was prepared as a solution of 0.25 % Brilliant Blue R in 50 % methanol and 10 % acetic acid. This solution was filtered before use. Gels were stained by immersion in the solution for 5 to 10 minutes, with slow agitation. After staining, gels were destained in 25 % methanol and 5 % acetic acid with slow agitation, until the protein bands (or spots) were visible, and stored in distilled water (Rabilloud and Charmont, 2000).

#### **4.4.3. Silver staining**

After electrophoresis, gels were fixed in a solution containing 50% ethanol and 5% acetic acid, for 3 periods of 30 minutes or overnight. After fixing, gels were washed in 50 % ethanol for 10 minutes, and in deionized (mQ) water for another 10 minutes. For the sensitization step, gels were placed in 0.02 % sodium tiosulphate for 1 minute, and after washing with mQ water for 2 times 20 seconds, staining was carried out in 0.15 % silver nitrate for 20 minutes. Again, gels were washed with mQ water 2 times for 15 seconds before development in a solution of 0.04 % formaline and 2 % sodium carbonate, for a variable period of time until the protein bands or spots were visible. The reaction was stopped by addition of 5 % acetic acid for 1 minute, and the resulting gels were stored in mQ water (Rabilloud and Charmont, 2000).

#### **4.5. Western Blot Analysis**

After electrophoresis, proteins were electroblotted onto a nitrocellulose membrane (Hybond ECL, Amersham) for immunodetection (Towbin *et al.*, 1979). For this, six sheets of 3MM paper (Whatman) and the blotting membrane were cut to the gel dimension. The membrane was pre-hydrated in distilled water and then hydrated in TGM (20 mM Tris-Cl, 150 mM glycine and 20 % methanol) for 10 minutes. The gel was also equilibrated in transfer buffer for 10 minutes. 3 sheets of 3MM paper

hydrated in TGM were placed on the anode of the transfer system, and the “sandwich” was assembled by laying down the membrane on top of the papers, followed by the gel and the remaining 3 sheets of 3MM paper hydrated with TGM. Air bubbles were avoided by rolling a glass pipette on the papers before placing the cathode to complete the setup. Protein transfer was carried out in a semi-dry blotter (BioRad), at 0.8 mA / cm<sup>2</sup> of gel (approximately 12V) for 20 minutes.

After protein transfer, the membrane was washed in PBS-T (140 mM NaCl, 1 mM KCl, 19 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM K<sub>2</sub>HPO<sub>4</sub> pH7.4, with 0.1 % (v/v) Tween-20) for 15 minutes and blocked at room temperature for 2 hours with 5 % (w/v) skimmed milk powder in PBS-T. Then the membrane was washed twice with PBS-T, for 5 minutes each time, and sealed inside a plastic bag for incubation with the primary antibody, diluted in 1 % skimmed milk powder in PBS-T, overnight at 4°C. The membrane was then washed 3 times for 20 minutes each time, with 1 % skimmed milk powder in PBS-T, and sealed inside a plastic bag for incubation with the secondary antibody, diluted in 1 % skimmed milk powder in PBS-T, for 1 hour at room temperature. Finally the membrane was washed 3 times with PBS-T, for 10 minutes each time.

Immunodetection was performed by chemiluminescence, using the ECL kit from Amersham, according to the manufacturer’s instructions. In a dark room, detection reagent A and detection reagent B from the ECL kit were mixed (1:40) and the solution applied on the membrane, placed in a layer of Saran wrap, ensuring that the entire surface is covered. After a 5-minute incubation at room temperature, the mixture was quickly removed with a pipette and the membrane covered with the Saran wrap, avoiding air bubbles. The membrane was then exposed to X-ray film (Kodak) for a suitable period of time and the film was developed and fixed using the appropriate reagents (Bini *et al.*, 2000).

#### 4.6. Protein carbonylation assays

Yeast protein extracts from cultures grown at 30°C to an OD<sub>600nm</sub> of 0.5 were prepared in 50 mM potassium phosphate buffer (pH 7.0) containing a mixture of protease inhibitors (30 µg / ml pepstatin, 30 µg / ml leupeptin, 6 µg / ml antipain and 6 mM EDTA). Cells were disrupted on a MiniBeadBeater (Biospec Products) in the presence of glass beads, by shaking the tubes for 5 cycles of 20 seconds followed by 20 seconds on ice. After centrifugation at 4000 rpm for 5 minutes at 4°C, the supernatant was transferred to a new tube and clarified by centrifugation at 13000 rpm for 15 minutes at 4°C. The extract was divided in aliquots and stored at – 80°C.

For 1D-PAGE analysis of oxidatively modified proteins, a volume corresponding to 40 µg of protein was mixed with an identical volume of 12 % SDS and centrifuged for 15 seconds. Derivatization of the carbonyl groups in the protein side chains to 2,4-dinitrophenylhydrazone was achieved by reaction with 2 volumes of 20 mM 2,4-dinitrophenylhydrazine in 10 % TFA. After addition of this solution, samples were centrifuged for 15 seconds and incubated for 30 minutes at room temperature, in the dark. The reaction was neutralized by adding 1.5 volumes of a solution containing 2 M Tris, 30 % glycerol, and 17 % β-mercaptoethanol (Levine *et al.*, 1994; Costa *et al.*, 2002).

For the SDS-PAGE analysis, a volume corresponding to 12 µg of protein was loaded on each gel lane. After electrophoresis, proteins were electroblotted as described above (II.4.5.). Membranes were incubated with the primary antibody, specific for the DNP moiety of the proteins (rabbit IgG anti-DNP, Dako), at a 1:5000 dilution, and subsequently with the secondary antibody, goat anti-rabbit IgG linked to horseradish peroxidase (Sigma), at a 1:5000 dilution. Immunodetection was performed by chemiluminescence, as described above (II.4.5.).

#### **4.7. Ubiquitylation assays**

Protein extracts used for determining the levels of protein ubiquitylation were identical to those used for the protein carbonylation assays (II.4.6.). A volume corresponding to 12  $\mu\text{g}$  of protein was loaded on each lane of the gel. After electrophoresis, proteins were electroblotted as described above (II.4.5.), membranes were incubated with the primary rabbit IgG anti-ubiquitin antibody at a 1:5000 dilution, and subsequently with the secondary goat anti-rabbit IgG antibody linked to horseradish peroxidase (Sigma), at a 1:5000 dilution. Immunodetection was performed by chemiluminescence, as described above (II.4.5.).

#### **4.8. HSF-1 analysis**

Cells were grown at 30°C to an  $\text{OD}_{600\text{nm}}$  of 0.5 and proteins were extracted as described above for the protein carbonylation assays (II.4.6.). A volume corresponding to 5  $\mu\text{g}$  of protein was loaded on each lane of the gel. After electrophoresis, proteins were electroblotted as described above (II.4.5.), membranes were incubated with the primary rabbit anti-human-HSF-1 polyclonal antibody (Stressgen), at a 1:5000 dilution, and subsequently with the secondary goat anti-rabbit IgG antibody linked to horseradish peroxidase (Sigma), at a 1:5000 dilution. Immunodetection was performed by chemiluminescence, as described above (II.4.5.).

#### **4.9. PhosphoSerine analysis**

Cells were grown at 30°C to an  $\text{OD}_{600\text{nm}}$  of 0.5 and protein extracts were obtained as described above for the protein carbonylation assays (II.4.6.). A volume corresponding to 5  $\mu\text{g}$  of protein was loaded on each lane of the gel. After electrophoresis, proteins were electroblotted as described above (II.4.5.), membranes were incubated with the primary mouse monoclonal anti-PhosphoSerine antibody (Qiagen), at a 1:1000 dilution, and subsequently with the secondary anti-mouse IgG antibody linked to horseradish peroxidase (Amersham),

at a 1:5000 dilution. Immunodetection was performed by chemiluminescence, as described above (II.4.5.).

#### 4.10. Karyotype analysis

DNA preparation for PFGE was essentially carried out as described by Chu and colleagues (1993). Briefly, cells were grown overnight at 30°C and a maximum of 2 ml overnight culture were harvested at 13000 rpm, at room temperature, and the supernatant was discarded. The cell pellet was washed once in 1 ml of 50 mM EDTA (pH 7.5). Cells were resuspended in 160 µl of 50 mM Tris-HCl (pH 7.5), 60 µl of 2.5 mg / ml Zymolyase 100T solution (Seikagaku Corp.) and 600 µl of 1.5 % low-melting agarose (LM-MP, Boehringer Mannheim) in 0.125 M EDTA (pH 7.5), kept at 50°C. The mixture was then transferred into a 1ml syringe used as a mould and allowed to solidify. Upon solidification, the syringe body was broken and the agarose block was cut in slices, transferred into a 15 ml Falcon tube and incubated at 37°C, in 3 ml of LET buffer containing 0.5 M EDTA, 10 mM Tris-HCl pH 7.5, 1% 2-mercaptoethanol, for 1 day. The LET buffer was replaced by 3 ml of NDS buffer containing 0.5 M EDTA, 10 mM Tris-HCl pH 7.5, 1 % N-lauroyl sarcosine, 0.2 mg / ml of proteinase K (Boehringer Mannheim) and incubated at 50°C for 2 days. The plugs were then washed twice with 0.5 M EDTA (pH 9.0) and stored at 4°C (Chu *et al.*, 1993).

Chromosomes were separated in 0.6% agarose gels (Pulsed Field Certified Agarose, BioRad), at 14°C, in 0.5 x TBE (0.045 M Tris, 0.045 M Boric acid, 1 mM EDTA) using the BioRad CHEF (contour-clamped homogeneous electric field) DRII system. The separation conditions were 120 to 300 s for 24 h, 420 to 900 s for 48 h at 80 V.

Karyotype gels were stained in a 0.1 mg / ml ethidium bromide solution for 1 hour and then washed in mQ water, to remove the excess of ethidium bromide, for 20 minutes. The gel image was acquired in the Molecular Imager FX Pro Plus Multilmager System (BioRad) using the Quantity One software (BioRad).



## **III. Results**

1. Transcriptome analysis of *S. cerevisiae* cells expressing CUG ambiguity
2. Proteome analysis of *S. cerevisiae* cells expressing CUG ambiguity
3. Physiological characterization of *S. cerevisiae* cells expressing CUG ambiguity



## 1. Transcriptome analysis of *S. cerevisiae* cells expressing CUG ambiguity

### 1.1. Introduction

The global impact of genetic code alterations on the physiology of an organism has not yet been studied. However, the characterization of an *E. coli* strain with an editing-deficient isoleucyl-tRNA synthetase (IleRS), over a wide-range of growth conditions, sheds new light on the consequences of living with an ambiguous genetic code (Bacher *et al.*, 2005). Indeed, impairment of the editing activity of the IleRS results in the production of statistical proteins, due to the misincorporation of amino acids that are very similar to isoleucine, namely valine, at Ile codons. These errors in translation have generalized effects on protein function, and consequently decrease cellular growth under the several temperatures and media tested (Bacher *et al.*, 2005). Although these results obtained in bacteria are important to understand the evolution of the genetic code, they do not provide a global picture of the effects of genetic code ambiguity on the physiology and gene expression in an eukaryotic organism. Moreover, the gene expression changes occurring in response to genetic code manipulations should be compared with the cellular response to a standard stress such as heat (Santos *et al.*, 1999; Hughes and Ellington, 2005). This underlines the importance of the work described in this and in the next chapters, on the response of yeast cells to genetic code ambiguity.

Previous results from *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> suggest that CUG ambiguity induces the general stress response, as indicated by increased thermotolerance, tolerance to oxidative and osmotic stress, and resistance to heavy metals and drugs (Santos *et al.*, 1996; Santos *et al.*, 1999). These pleiotropic effects caused by misreading of the CUG codon are similar to the response that has been reported for other conditions, such as overexpression of the Yap1p transcription factor (Wu *et al.*, 1993; Gounalaki and Thireos, 1994; Wemmie *et al.*, 1994). Nevertheless, the response triggered by CUG mistranslation is distinct from that induced by heat shock and other stress conditions, since it consists of an internal and permanent stress rather than a

transient and external alteration of the environment (Santos *et al.*, 1999). The response of yeast cells to standard stress conditions is well studied, but the consequences of long-term stress exposure have not yet been characterized. Therefore, understanding the response induced by CUG ambiguity at the molecular level is crucial to elucidate the evolution of genetic code alterations mediated by ambiguous decoding, and also to better understand the cellular mechanisms of adaptation to permanent stress conditions.

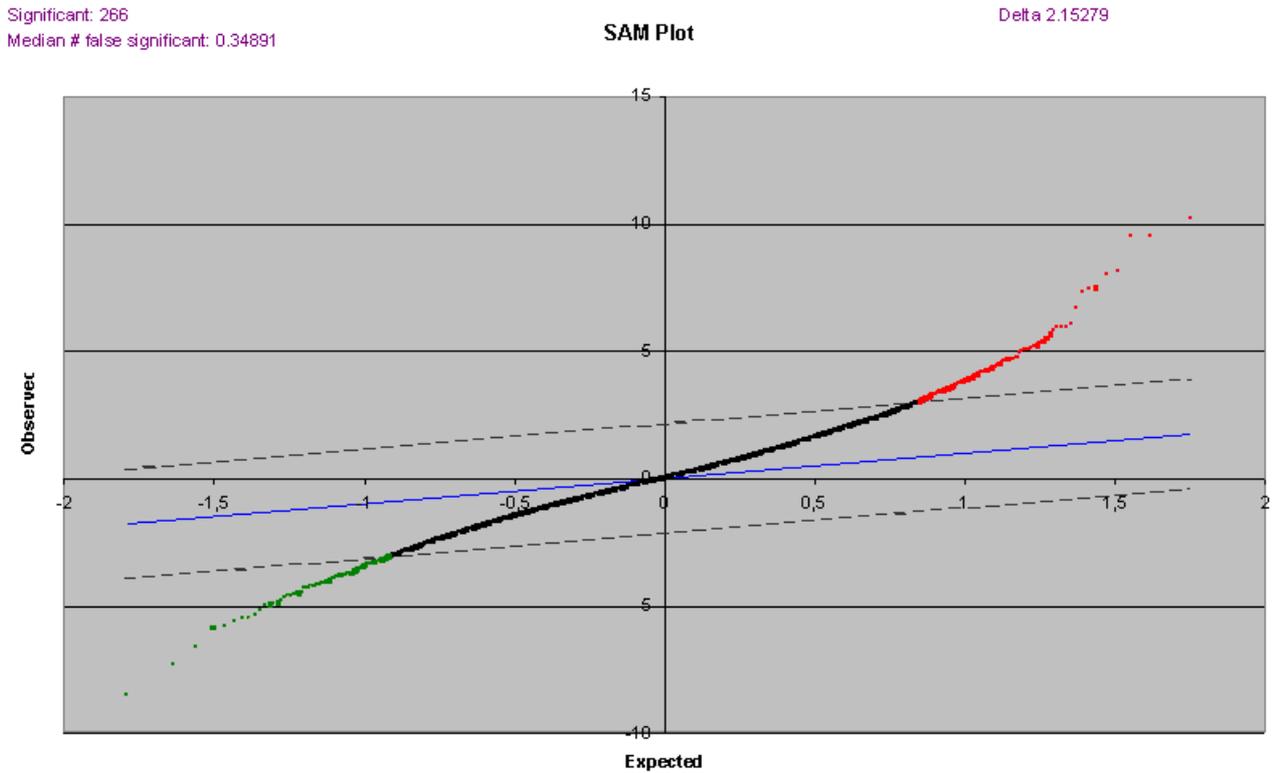
It is well established that yeast cells adapt to stress challenges by reprogramming gene expression (Gasch *et al.*, 2000). DNA microarrays are a powerful tool to characterize the transcriptome of cells under a given condition, as they allow the measurement of the relative transcript abundance for all genes. Indeed, several studies have focused on the global gene expression profiling of yeast cells under several stresses (reviewed by Gasch and Werner-Washburne, 2002). In this study we used these methodologies to uncover the gene expression alterations induced by CUG ambiguity.

## 1.2. Transcriptome of *S. cerevisiae* cells expressing CUG ambiguity

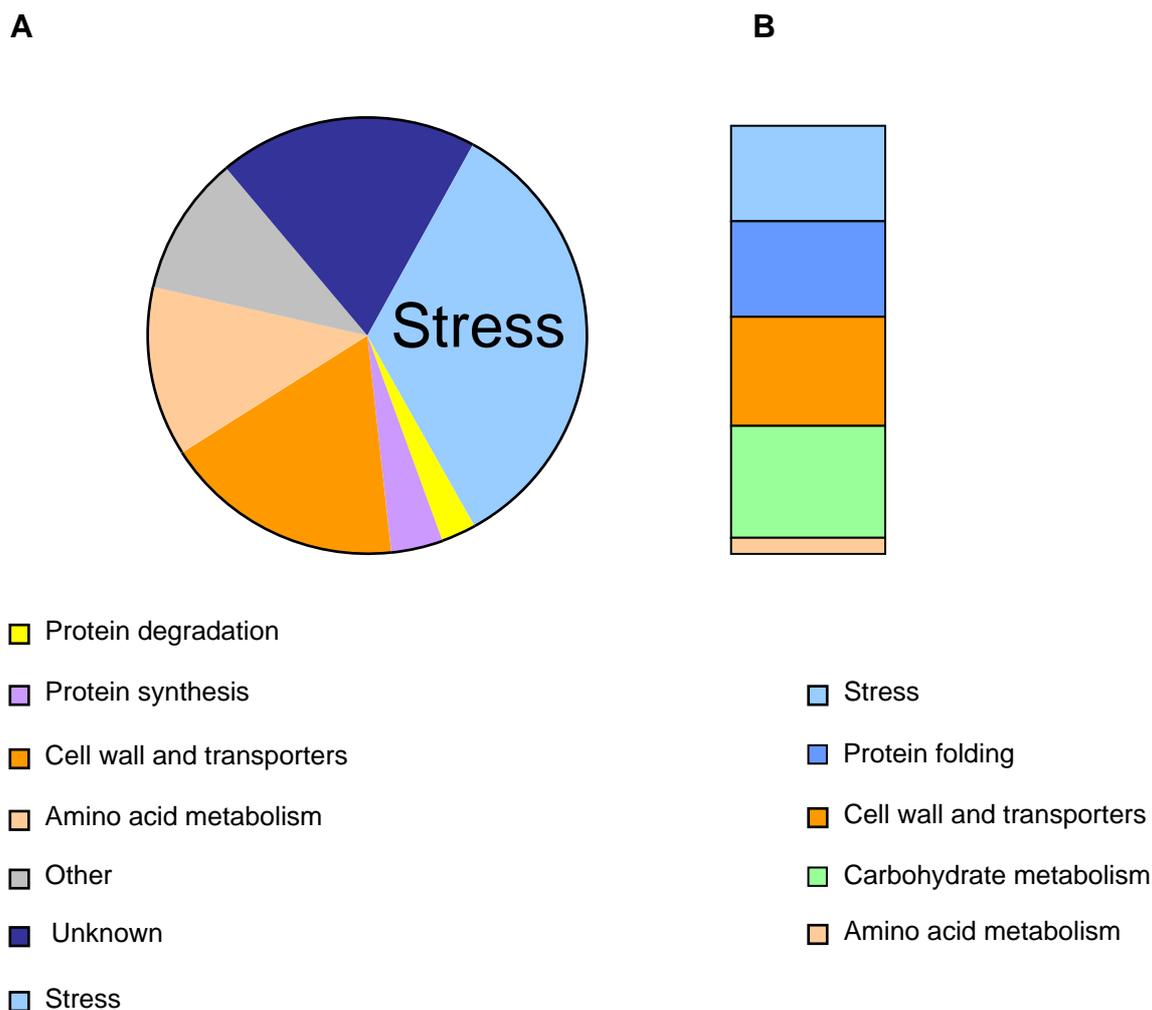
Characterization of the transcriptome of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> was done based on 12 microarray slides, corresponding to 10 independent cultures (5 for the tRNA G<sub>33</sub> and 5 for the T<sub>33</sub>), and performed with dye swap. Transcriptome alterations of both *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub> were taken together and, for this reason, a general response to CUG mistranslation will be considered.

*S. cerevisiae* control cells and cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (T<sub>33</sub> and G<sub>33</sub>) were grown at 25°C with shaking until the cultures reach an OD<sub>600nm</sub> of 0.5. Total RNA was isolated from *S. cerevisiae* using hot phenol and mRNA was recovered using Oligotex (Qiagen). After complementary DNA synthesis, samples were coupled to Cy3 and Cy5 fluorophores and hybridized for 16 to 20 hours at 42°C. Image analysis was carried out as described in methods. After pre-normalization with QQCC, data was analyzed in GeneSpring and SAM (see methods). The data obtained with GeneSpring and SAM was compared and the most significant results will be presented.

SAM analysis with a  $\Delta$  value set to 2.15 resulted in a false discovery rate (FDR) of 0.1 % in 266 significant genes (Figure 19). FDR represents the statistical significance that is, therefore, 0.001 for all genes (see methods). Comparison of the SAM list with the GeneSpring data resulted in a common set of 170 genes with altered expression levels in response to genetic code ambiguity. From these, 81 genes were selected based mainly on their fold variation (Annex, Tables A and B).



**Figure 19 – SAM plot of the gene expression data obtained from *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub>.** The graphic shows the observed vs. expected mRNA levels, with the  $\Delta$  value set to 2.15, resulting in a false discovery rate of 0.1 % in 266 genes. Genes with significantly altered expression levels coloured in red are up regulated, and in green are down regulated.



**Figure 20 – Genes whose expression is altered by CUG ambiguous decoding distributed by functional categories.** A - genes whose expression levels are both up and down regulated by genetic code ambiguity were grouped according to their functions. The graphic on B represents the genes that were included in the stress group from the graphic displayed on A.

The main functional groups of genes (functions were assigned according to the SGD web page, <http://www.yeastgenome.org>) whose expression is altered in *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub> are shown in Figure 20A. The biggest impact of CUG mistranslation was on the stress response, which includes genes belonging to protein folding, carbohydrate metabolism and other stress-responsive genes (Figure 20B). All genes within this group have increased expression levels in CUG ambiguous cells. Other functional categories that are affected by CUG ambiguity are cell wall and transporters and amino acid metabolism, which include up and down regulated genes. Protein synthesis and protein degradation are also affected, with the first group being repressed and the second induced.

### 1.2.1. CUG ambiguity induces the expression of stress genes

Genes that encode heat shock proteins (Table 2) and other stress-responsive proteins (Table 3) are induced in *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, and may therefore account for the stress tolerance phenotypes previously described (Santos *et al.*, 1999). Some genes can be included in more than one functional category, for instance, some cell wall, transporters and carbohydrate metabolism genes are also implicated in the stress response. Among the molecular chaperones induced by CUG mistranslation, it is noteworthy the absence of some members of the Ssa subfamily of the Hsp70 family. The mRNA levels from *SSA1*, *SSA2* and *SSA3* were below the 2 fold variation threshold and hence were not selected as significant, although their role in assisting protein folding in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> cannot be ruled out (see chapter III.2.). Hsp70 molecular chaperones work together with Hsp104 to refold non-native proteins found in aggregates (Glover and Lindquist, 1998; Lum *et al.*, 2004) and are, therefore, important for clearing protein aggregates that result from misfolded proteins produced during ambiguous CUG decoding.

**Table 2 – Molecular chaperone genes whose expression is induced in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>. The respective fold variation is indicated (FDR=0.001).**

Gene name	Fold induction
HSP12	12.4
HSP26	5.8
SSA4	3.9
HSP42	3.9
HSP30	3.7
SSE2	2.7
HSP104	2.4

**Table 3 – Stress genes whose expression is induced in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>. The respective fold variation is indicated (FDR=0.001).**

Gene name	Fold induction
DDR2	9.0
CRS5	3.0
CUP1	2.9
PNC1	2.1
GRE2	2.1

Cells with genetic code ambiguity show increased expression levels of genes involved in copper resistance, such as *CUP1-1*, *CUP1-2* and *CRS5*, but also up regulate stress-protective genes like *PNC1*, *GRE2*, and cell wall genes (Table 4).

### 1.2.2. CUG ambiguity alters the expression of cell wall genes

Many cell wall genes show increased expression levels in response to CUG ambiguity (Table 4). In this group, there are genes whose products have a role in maintaining the structure of the cell wall, such as *CWP2*, but also several transporters like *MCH2*, *PDR5* and *PTR2*. *TIP1* and *TIR1* are stress-induced genes, *SED1* is related with the maintenance of cell wall integrity and stress tolerance, and *PDR5* is a drug-efflux pump involved in drug resistance. Some genes that code for transporters are down regulated in response to genetic code ambiguity, such as *ATR1*, a member of the multidrug-resistance family (Goffeau *et al.*, 1997), and *VHT1*. *VHT1* encodes an H<sup>+</sup>-biotin symporter (Stolz *et al.*, 1999) and its repression correlates well with the decreased expression levels registered for biotin biosynthesis genes (Table 8).

**Table 4 – Cell wall and transporter genes whose expression is altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>.** Genes shown in red are induced and genes in green are repressed. The respective fold variation is indicated (FDR=0.001).

Gene name	Fold variation
YGP1	8.8
PIR3	7.5
TIP1	6.2
CWP2	4.0
SED1	2.8
MCH2	2.7
PDR5	2.2
PTR2	2.0
TIR1	2.0
ATR1	2.1
VHT1	2.0

The cell wall structure and composition seem to be major targets of CUG mistranslation, suggesting a role for CUG ambiguity in *C. albicans* biology since the cell wall is determinant for maintaining shape and integrity in yeast cells. Apart from being remodelled during growth, morphogenesis and in response to environmental challenges, it represents the initial point of interaction between host and pathogen (Jung and Levin, 1999; Martinez-Lopez *et al.*, 2004). One might expect that the production of distinct cell wall proteins in different cells within the population could be important during infection. In *C. albicans*, CUG mistranslation replaces leucines for serines, and moreover, those serines can eventually be glycosylated, so it is tempting to speculate that the production of different antigens from different cells might be advantageous to assure that part of the population elude the immune system. Another possibility is that the increased expression in cell wall structural components would harden the cell wall and provide stress resistance.

### **1.2.3. CUG ambiguity induces the expression of phosphate metabolism genes**

Phosphate metabolism also shows important alterations in response to CUG ambiguity, with 6 genes up regulated in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (Table 5). *PHO84* and *PHO89* are high-affinity orthophosphate permeases, and *PHO5*, *PHO11* and *PHO12* are secreted acid phosphatases. *SPL2* is a homologue of *PHO81*, the inhibitor of the Pho80p-Pho85p cyclin-dependent protein kinase, therefore related with signal transduction of the PHO regulatory system (Pinson *et al.*, 2004).

Phosphate is an essential nutrient required for the biosynthesis of nucleic acids, proteins, lipids and sugars, and energy metabolism, and it is possible that two signalling pathways sense external and internal phosphate concentrations in order to assure homeostasis (Auesukaree *et al.*, 2004; Pinson *et al.*, 2004). When the phosphate concentration is low, the Pho81p inhibits the Pho80p-Pho85p kinase

activity and abolishes the phosphorylation of the Pho4p. When hypophosphorylated, this transcriptional factor is located in the nucleus where it induces the expression of target genes, including all genes shown in Table 5 (Ogawa *et al.*, 2000).

**Table 5 – Phosphate metabolism genes whose expression is induced in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>.** The respective fold variation is indicated (FDR=0.001).

Gene name	Fold induction
PHO84	17.6
SPL2	6.9
PHO89	4.7
PHO12	3.4
PHO11	3.1
PHO5	2.9

Phosphate has also been implicated recently in the signal transduction pathway mediated by the protein kinase A (PKA), where *PHO84* would act as a sensor for rapid phosphate signalling (Giots *et al.*, 2003). Under phosphate starvation conditions PKA activity is low, resulting in a number of cellular alterations, namely storage of reserves such as trehalose and glycogen, increased stress resistance and low cell wall lyticase sensitivity, induction of STRE-controlled genes and repression of ribosomal protein genes, inhibition of pseudohyphal growth and stimulation of sporulation (Thevelein and de Winde, 1999; Thevelein *et al.*, 2000). Strikingly, the phenotypes indicative of low activity of the cAMP-PKA pathway are observed in yeast cells growing on non-fermentable carbon sources and stationary phase, and some of these phenotypes were observed in cells mistranslating the CUG codon, such as increased stress resistance (Santos *et al.*, 1996; Santos *et al.*, 1999), accumulation of trehalose and glycogen, induction of STRE-controlled

genes and repression of ribosomal protein genes (this work). The mRNA content for the known members of the cAMP-PKA pathway was not altered, although their activity might be regulated by phosphorylation (Giots *et al.*, 2003). An interesting result is the 3.8 and 2.4-fold induction of the *GPG1* and *BTN2* genes, respectively. The first encodes a protein that interacts with Gpa2p, a subunit of the G-protein coupled receptor mediating the PKA pathway. The second is involved in pH homeostasis, and PKA activity is triggered by cAMP synthesis, which is stimulated by intracellular acidification. They can be, therefore, good candidates to mediate the down-regulation of the PKA pathway even in the presence of glucose. Why are *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> starved for phosphate, and how is the PKA pathway repressed in these strains remains to be elucidated.

#### **1.2.4. CUG ambiguity induces the expression of carbohydrate metabolism genes**

Several carbohydrate metabolism genes showed increased expression levels in response to CUG mistranslation (Table 6). The induction of *HXK1* and *GLK1* might reflect an increase in the glucose flux for glycolysis, as more energy is required for the ATP-dependent folding and degradation processes (Gasch *et al.*, 2000). *GPH1* releases glucose from glycogen, and again glycogen breakdown might be related with an increased flux of glucose towards glycolysis. This is not too surprising, however, since recycling of glycogen and trehalose is enhanced under stress conditions (Parrou *et al.*, 1997; Godon *et al.*, 1998). Nevertheless, accumulation of these reserve carbohydrates was observed in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (this will be discussed in chapter III.3.4.). This is in agreement with the induction of *TSL1*, *TPS1* and *PGM2*, which suggests an increase in trehalose synthesis that may be necessary to stabilize mistranslated proteins (Kaushik and Bhat, 2003).

**Table 6 – Carbohydrate metabolism genes whose expression is induced in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>.** The respective fold variation is indicated (FDR=0.001).

Gene name	Fold induction
HXK1	6.2
GPH1	4.7
TSL1	3.2
PGM2	2.4
GLK1	2.3
TPS1	2.1

### 1.2.5. CUG ambiguity represses the expression of protein synthesis genes

As reported for other stress conditions, protein synthesis is repressed in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (Table 7).

The down regulation of the genes that code for the ribosomal proteins and other translation factors, including aminoacyl-tRNA synthetases is well known (reviewed by Gasch and Werner-Washburne, 2002), although the expression levels determined for this group of genes in response to CUG ambiguity are not as reduced as described by others that observe, for instance, a 20-fold decrease in the expression of ribosomal proteins (Trotter *et al.*, 2002). This suggests a rather small impact of CUG ambiguity on the rate of protein synthesis.

**Table 7 – Protein synthesis genes whose expression is repressed in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>.** The respective fold variation is indicated (FDR=0.001).

Gene name	Fold repression
RPL22B	2.1
YDR341C	2.1
RPL9B	2.0

### 1.2.6. CUG ambiguity alters the expression of amino acid metabolism genes

The amino acid metabolism is yet another functional group that is affected in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (Table 8).

**Table 8 – Amino acid metabolism genes whose expression is altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>.** Genes shown in red are induced and genes in green are repressed, and the respective fold variation is indicated (FDR=0.001).

Gene name	Fold variation
PUT1	3.7
SFA1	3.2
GAD1	2.6
ALD3	2.4
LYS9	2.4
SAM4	2.0
BIO5	2.0
BIO3	2.0

*ALD3* and *PUT1* are induced by CUG ambiguity, and also show increased expression in the presence of other stresses (Trotter *et al.*, 2002). *GAD1* is a glutamate decarboxylase that catalyzes the formation of  $\gamma$ -aminobutyrate (GABA) from glutamate. Although GABA is well known as an intercellular signalling molecule, the role of this amino acid metabolite in intracellular metabolism is not yet clear. However, increasing the copy number of the *GAD1* gene increases tolerance to oxidants such as H<sub>2</sub>O<sub>2</sub> and diamide (Coleman *et al.*, 2001). Finally, *SAM4* is involved in sulfur amino acid metabolism and its expression decreases during misfolding stress (Trotter *et al.*, 2002). The repression of *BIO5* and *BIO3* points towards a down-regulation of the biotin biosynthesis pathway.

### **1.2.7. Effect of CUG ambiguity on additional cellular functions**

The induction of *TFS1* and *RPN4* suggests an increase in protein degradation in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>. *RPN4* is a transcription factor that regulates the expression of many proteasome subunits, as well as other genes involved in the ubiquitin-proteasome pathway (Mannhaupt *et al.*, 1999). However, the absence of increased levels of the mRNAs encoding other proteasome subunits, besides *RPN4*, together with the increased levels of a number of proteasome subunits (chapter III.2.2.3.), indicates that expression of proteasomal subunits is regulated at the translational level. The induction of the *RPN4* gene could also serve other functions, as it connects the ubiquitin-proteasome system with different pathways. For example, it plays a role in the unfolded protein response (UPR), a signalling pathway that senses misfolded proteins in the endoplasmic reticulum (ER) lumen and induces the production of ER chaperones and proteins related to the secretory pathway (Ng *et al.*, 2000). Proteins that cannot be rescued after UPR activation are degraded, by a pathway named ER-associated protein degradation (ERAD) that involves translocation of the misfolded proteins back to the cytosol (retrotranslocation) with consequent degradation by the proteasome (Ng *et al.*, 2000; Kostova and Wolf, 2003). *RPN4* and other proteasomal subunit genes are not regulated by the UPR, but are co-

ordinately induced in UPR-deficient cells after treatment with tunicamycin or DTT (Travers *et al.*, 2000), suggesting the existence of an UPR-independent signalling pathway in which *RPN4* could participate. The fact that some UPR target genes contain the PACE (proteasome-associated control element) sequence in their promoter indicates a possible regulation by *RPN4* (Ng *et al.*, 2000).

*RPN4* can also regulate transcription of some DNA excision repair genes, namely *MAG1* that contains in its promoter a sequence similar to PACE (Jelinsky *et al.*, 2000). Additionally, the *RPN4* promoter contains sequences that bind Pdr1p, Pdr3p and Yap1p, which are transcription factors responsible for expression of the membrane transporters that mediate multiple drug resistance and the oxidative stress response, respectively. Thus, Rpn4p also links multiple drug resistance to the ubiquitin-proteasome system (Owsianik *et al.*, 2002).

Some genes altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> have other functions, namely mating and meiosis, or vesicle transport. One interesting gene is *PEX21*, encoding a peroxisomal biogenesis protein involved in peroxisomal protein targeting, which also enhances the aminoacylation activity of seryl-tRNA synthetase (Ses1p) *in vitro* (Rocak *et al.*, 2002). Although the decrease on *PEX21* expression is only 1.6-fold, the interaction with Ses1p may reduce serylation of the ser-tRNA<sub>CAG</sub> and, consequently, increase CUG decoding as leucine.

### **1.2.8. The response to CUG ambiguity is different from other stress responses**

Previous stress studies performed with yeast cells using DNA-microarrays (Gasch *et al.*, 2000) describe an environmental stress response (ESR) as a set of genes induced and repressed in all conditions tested that behave similarly. Therefore, the ESR is a general adaptive response to suboptimal environmental conditions. The two main groups of genes repressed in the ESR are involved in RNA metabolism, namely RNA processing, translation and tRNA synthesis and processing, and other growth-related processes, as well as ribosomal protein genes. The repression of genes belonging to this later group was also observed under other stresses (Gasch and Werner-Washburne, 2002). Considering the groups induced in the ESR, these authors identified genes related to protein folding and degradation, detoxification of reactive oxygen species and redox reactions, carbohydrate metabolism, cell wall modification and transport, among others. Many of those genes have been implicated in the defence against stressful environments, namely during exposure to oxidative stress, heat shock, osmotic shock and starvation. Interestingly, these global gene expression changes were transient, and transcript levels achieved new steady states as cells adapted over time to the stressful environments. The nature of the stress imposed by the expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* cells is different from other described environmental stresses, as it consists of an internal and permanent stress rather than an external and transient unfavourable condition. Despite these differences, the functional groups altered by genetic code ambiguity consist mainly of those altered in the ESR, with the repression of ribosomal protein genes and the induction of molecular chaperones, carbohydrate metabolism, cell wall and transport genes.

There are other stress conditions that induce protein misfolding, namely ethanol or the proline analogue azetidine-2-carboxylic acid (AZC), and they also activate heat shock factor-regulated genes but not the stress response element (STRE) regulon (Trotter *et al.*, 2002).

Transcriptome analysis of AZC-treated cells showed that this analogue causes most, but not all, of the gene expression changes observed in response to temperature upshift, and that it does not activate the endoplasmic reticulum unfolded protein response (Trotter *et al.*, 2002). Among the genes induced by AZC but not by temperature upshift, *CUP1-1* and *CUP1-2* also showed increased expression in response to genetic code ambiguity. Interestingly, many gene expression alterations reported in AZC-treated cells, specially induced genes, were also found in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, although the fold variation is systematically lower in CUG ambiguous cells with only few exceptions. One possible explanation for these differences may be that CUG ambiguity consists of a permanent stress, whereas treatment with AZC is transient.

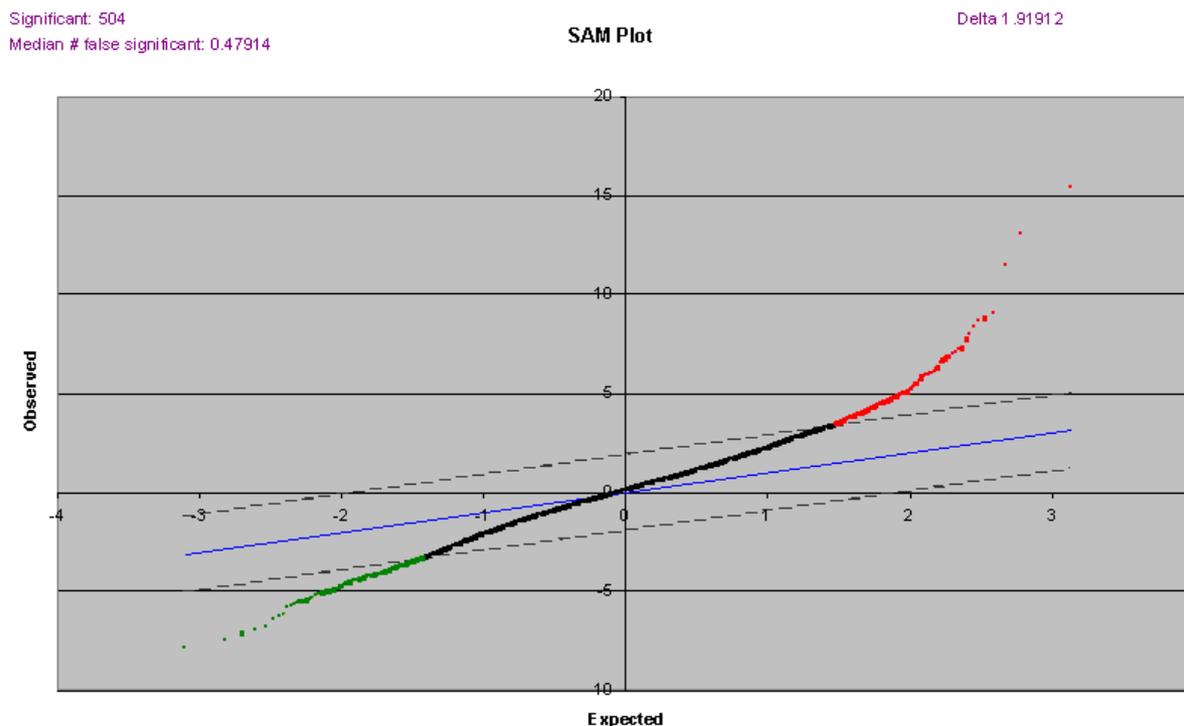
Curiously, most carbohydrate genes whose expression is increased in response to CUG mistranslation (Table 5) do not appear to be altered by AZC, namely the genes related to the trehalose and glycogen metabolism. As the expression of these genes is dependent on Msn2p or Msn4p (Boy-Marcotte *et al.*, 1999) and AZC exposure does not induce the STRE regulon, the results suggest that the cellular response to CUG mistranslation is different from that induced by AZC, although both stress conditions involve protein misfolding. Indeed, ambiguity of the CUG codon seems to induce both STRE and HSE-driven gene expression, as deduced from the increased levels of carbohydrate metabolism genes and molecular chaperones that are responsive to the Msn2/4p and Hsf1p transcription factors, respectively (Boy-Marcotte *et al.*, 1999). If the protein misfolding caused by CUG mistranslation elicits a different cellular response from the one described upon AZC exposure, it remains to be shown if the UPR is activated or impaired in our system, although the genes responsive to the UPR were not significantly altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>.

Cells submitted to AZC treatment show an increase in the mRNA levels of some proteasome subunits that are not altered by temperature upshift (Eisen *et al.*, 1998; Trotter *et al.*, 2002). Expression of *RPN4* increased 6-fold by AZC

treatment, while in CUG ambiguous cells it was only slightly induced (1.8-fold). Also noteworthy is the fact that *RPN4* is the only proteasome subunit gene whose expression changed. Nevertheless, the protein levels of some proteasome subunits are increased in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (chapter III.2.2.3.), suggesting the possibility of a translation regulatory mechanism.

Finally, considering the genes whose expression is downregulated, there is little overlap between the response to AZC and our dataset. More ribosomal genes are down regulated in response to AZC, suggesting a severe translational repression that is only partially observed in cells expressing genetic code ambiguity. Moreover, the expression levels of the amino acid and phosphate metabolism genes are not altered in AZC-treated cells, whereas in CUG ambiguous cells the amino acid biosynthetic pathways are repressed and phosphate metabolism genes are induced. Hence, the transcriptome changes that occur in cells that mistranslate the CUG codon are distinct from the alterations observed in response to other stresses such as temperature upshift, or agents that also induce misfolding like AZC.

### 1.3. Transcriptome of *S. cerevisiae* expressing the ser-tRNA<sub>CAG-T<sub>33</sub></sub>



**Figure 21 – SAM plot of the gene expression data obtained from *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG T<sub>33</sub></sub>.** The graphic shows the observed vs. expected mRNA levels, with the  $\Delta$  value set to 1.9 that results in a false discovery rate of 0.1 % in 504 genes. Genes with significantly increased or decreased expression levels are coloured in red and green, respectively.

For this chapter, only the transcriptome alterations of the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> mutant were considered. The ser-tRNA<sub>CAG</sub>-T<sub>33</sub> is more toxic than the ser-tRNA<sub>CAG</sub>-G<sub>33</sub>, as shown by a higher decrease in growth rate (Santos *et al.*, 1996), and, therefore, a specific response to high-level CUG ambiguity will be discussed. SAM analysis with a  $\Delta$  value set to 1.9 resulted in a false discovery rate of 0.1 % in 504 genes (Figure 21). Comparison of this list with the GeneSpring data resulted in a selection of genes displayed in Table 9.

**Table 9 – Genes whose expression is altered exclusively in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-T<sub>33</sub>.** Genes shown in red are induced and genes in green are repressed, and the respective fold variation in both tRNA mutants is indicated (FDR=0.001).

Gene name	Fold variation in T <sub>33</sub>	Fold variation in G <sub>33</sub>
ZRT1	4.9	1.4
FET4	2.8	1.0
PUT4	2.2	1.3
KAR2	2.2	1.2
LEU2	2.1	1.3

The metal metabolism seems more affected in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-T<sub>33</sub> than in the strain carrying the *C. albicans* ser-tRNA<sub>CAG</sub>-G<sub>33</sub>, in particular zinc metabolism. *ZRT1* encodes a high-affinity zinc transporter protein and *FET4* codes for a low-affinity iron transport protein that also functions in low-affinity copper and in zinc transport (Serrano *et al.*, 2004). The *ZPS1* gene was also found significantly induced in both mutants, however the fold increase is 35.3 and 4.3 for the T<sub>33</sub> and G<sub>33</sub> strains, respectively (Table 10), and its product is a zinc metalloproteinase. Other zinc metabolism genes were induced in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-T<sub>33</sub> but were not considered significant after SAM analysis, due to the high variability of their

expression levels between experiments. For example, *ZAP1*, the zinc-responsive transcriptional activator that regulates expression of genes involved in zinc uptake, whose expression level was increased 3.7 fold in the T<sub>33</sub> and 1.1 fold in the G<sub>33</sub> strain, and *ZRT3*, a vacuolar transporter that regulates zinc storage and mobilization in response to zinc deficiency, which is up-regulated 3.1 fold in the T<sub>33</sub> and 1.0 fold in the G<sub>33</sub> strain. Zinc is an essential nutrient and is a widespread structural or catalytic cofactor in yeast proteins (Lyons *et al.*, 2000). In zinc-limited cells the Zap1p transcriptional factor induces the expression of *ZRT1*, *ZRT2* and *ZRT3*, in order to increase zinc uptake and mobilize zinc reserves from the vacuole. It is somehow puzzling that the *ZRT2* gene is not induced in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub>, given the induction of *ZAP1*, *ZRT1* and *ZRT3*, but as a low-affinity zinc transporter its activity might be dispensable, since the high-affinity zinc transporter *ZRT1* has increased expression. *FET4* expression is also under the control of Zap1p (Lyons *et al.*, 2000).

The *KAR2* gene product is an ER-resident chaperone and one main target of the UPR, required for protein translocation across the ER membrane. A recent study showed that zinc deficiency in yeast and mammalian cells impairs ERAD and induces the UPR (Ellis *et al.*, 2004). This work demonstrated that zinc is required for normal ER function and that Msc2p is the protein that transports zinc from the cytosol to the ER lumen. Zrc1p and Cot1p transporters also play a role but are mainly responsible for zinc storage in the vacuole (MacDiarmid *et al.*, 2000). It is, therefore, interesting that none of these three intracellular zinc transporters were induced in our model, indicating that overexpression of *ZRT1* and *ZRT3* might be sufficient to cope with the zinc needs in these cells. The fact that binding of Rpn4p to PACE (proteasome-associated control element) is dependent on zinc (Mannhaupt *et al.*, 1999) broadens the implications of zinc deficiency to normal cellular function. Indeed, an increase in protein degradation (chapter III.3.3.) could prompt the cell to require additional zinc, hence the induction of zinc transporters. Overall, these observations suggest that cells expressing the *C. albicans* ser-

tRNA<sub>CAG</sub>-T<sub>33</sub> are zinc starved, or have higher zinc requirements, consistent with the possibility of UPR activation in this strain.

*PUT4* is a high affinity proline permease that also transports alanine and glycine, and is induced in starved cells. *PUT1* catalyses the first step in the synthesis of glutamate from proline, and its expression is increased in both *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (Table 8, previous chapter). *GAP1*, the general amino acid permease, is not induced (1.8 and 1.4 fold in T<sub>33</sub> and G<sub>33</sub> strains, respectively), nor are the SNZ genes that are stationary-phase induced and also markers of starved cells (reviewed by Gasch and Werner-Washburne, 2002). Furthermore, the amino acid biosynthesis pathways are repressed in *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (Table 8, previous chapter, and chapter III.2.2.4.). This raises the question of whether the T<sub>33</sub> cells are starved for nitrogen or amino acids? Starvation for an essential nutrient represses the PKA pathway, even in the presence of glucose, triggering stress resistance, accumulation of trehalose and glycogen, and expression of STRE-controlled genes (Thevelein *et al.*, 2000). These phenotypes were observed in cells expressing CUG ambiguity, although the down-regulation of amino acid biosynthesis in these cells does not support the hypothesis of starvation for nitrogen or amino acids. Alternatively, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> might be starved for other nutrients, namely phosphate, as discussed in the previous chapter. It is also possible that ambiguous cells are not starved, instead CUG ambiguity might interfere with the signalling pathways that control nutrient availability (see also chapter III.2.2.4), resulting in the same phenotypes as nutrient deficiency. This should be further studied, as nitrogen starvation triggers the expression of genes involved in sporulation, if cells are growing on a poor carbon source, and pseudohyphal growth, when cells are grown on a good carbon source such as glucose (Gasch and Werner-Washburne, 2002).

Apart from the genes described above, others were also found significantly altered in both *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, however the fold variation was much higher in T<sub>33</sub> than in G<sub>33</sub> cells (Table 10).

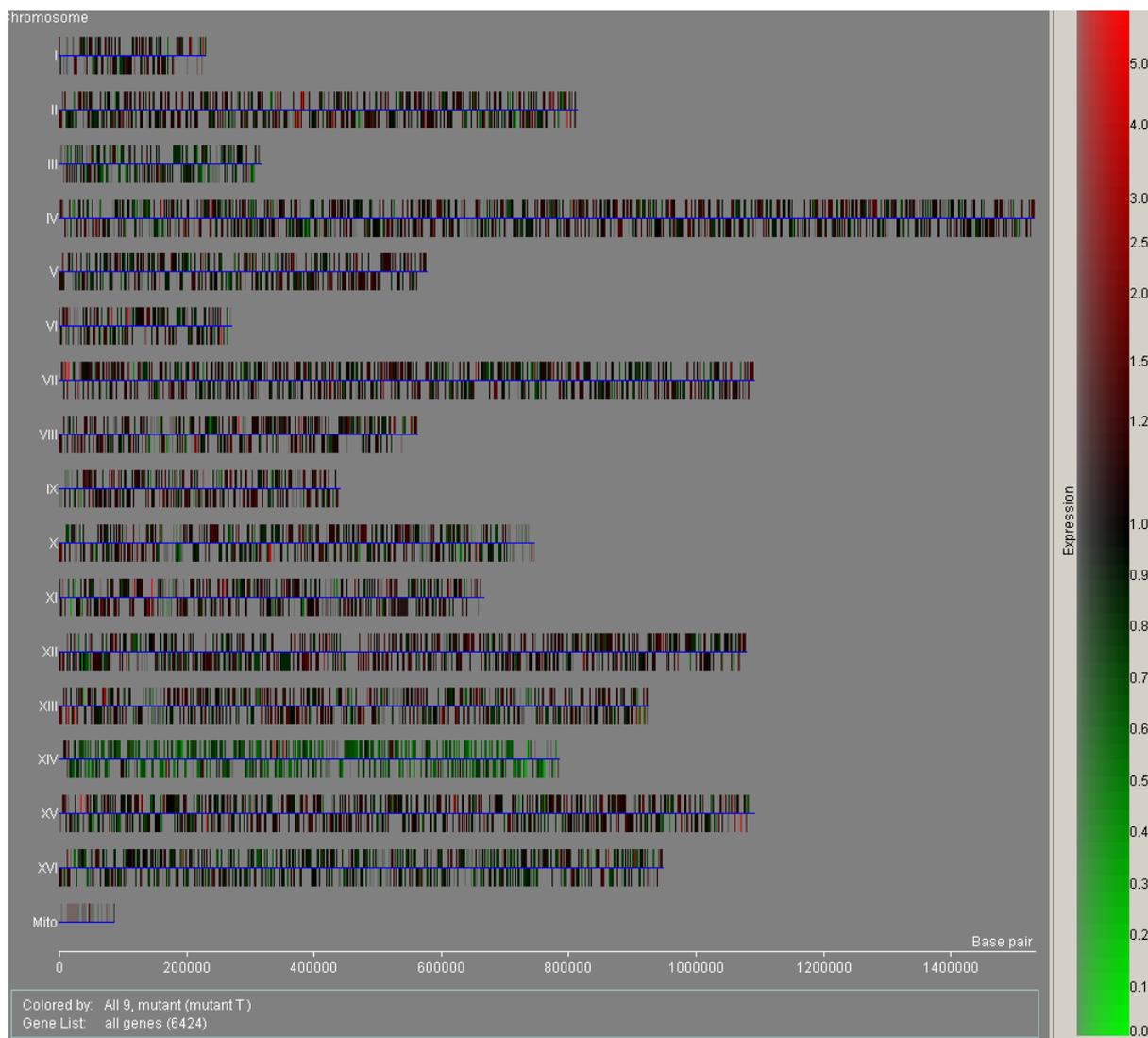
**Table 10 – Genes whose expression is highly induced in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-T<sub>33</sub>.** The respective fold variation is indicated (FDR=0.001).

Gene name	Fold variation T	Fold G
ZPS1	35.3	4.3
SPL2	8.2	5.6
PHO89	5.9	3.4
PHO12	4.4	2.3
PHO11	4.0	2.2

Phosphate metabolism is altered in both *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (discussed in the previous chapter), and it is remarkable that some phosphate metabolism genes appear more induced in the T<sub>33</sub> mutant, including the phosphate transporter *PHO89*. However, the G<sub>33</sub> cells might use an additional alternative phosphate source, as hinted by the 2.6-fold induction of the *GIT1* gene, whose expression was increased only 1.9-fold in the T<sub>33</sub> strain. The Git1p transports glycerophosphoinositol that can act as the sole phosphate source for the cell (Almaguer *et al.*, 2004), which could explain the lower induction of other phosphate transporters in G<sub>33</sub> cells. Globally, the data indicates that zinc and phosphate metabolism are more altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-T<sub>33</sub> than in the G<sub>33</sub> strain, however the significance of these findings remains unclear.

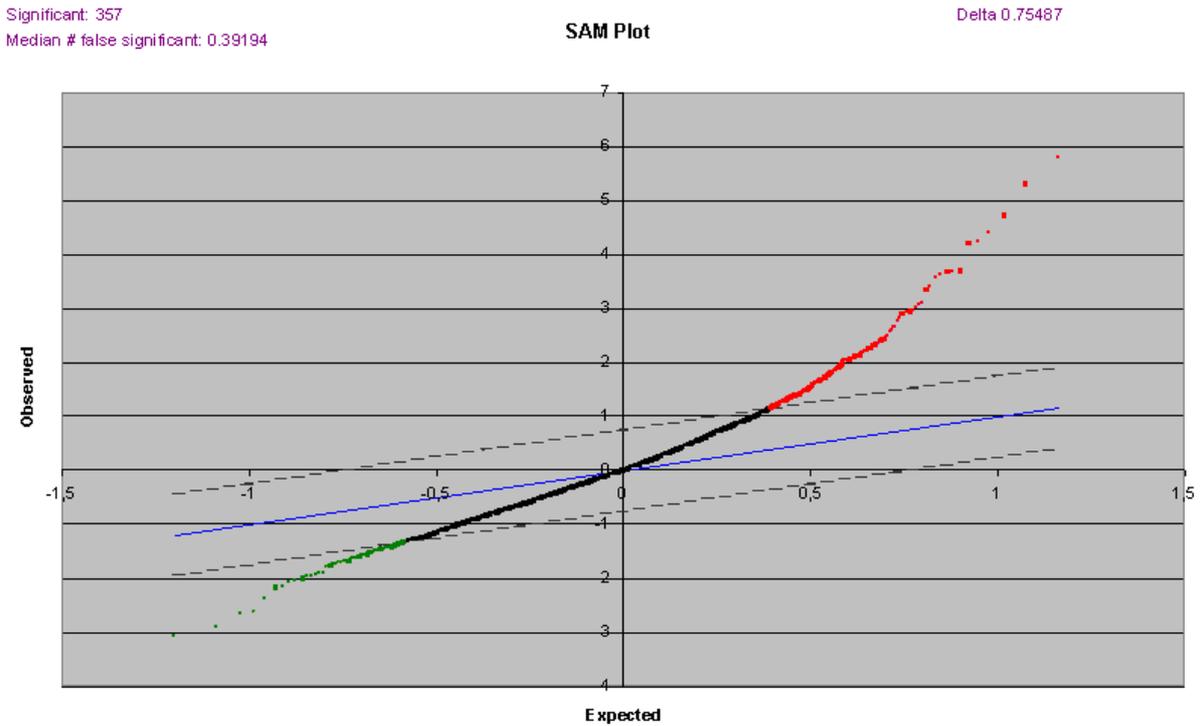
The genes whose expression levels are altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub>, both induced and repressed, were distributed according to their chromosomal localization using GeneSpring (Figure 22). Surprisingly, chromosome XIV shows an accumulation of down-regulated genes, instead of the normal trend observed for the other chromosomes that display a random mixture of genes with increased, decreased and unaltered expression levels. This could reflect chromosomal abnormalities, in this case a deletion of one copy of chromosome XIV. It has been previously determined that chromosome-wide expression biases in aneuploid cells can lead to erroneous results on gene expression profiling (Hughes *et al.*, 2000). Cells with genomic instability showed increased or decreased expression levels that could be explained solely based on duplication or deletion events. As the mRNA abundance of most genes in trisomic or monosomic chromosomes was altered, it was suggested that yeast cells do not have dosage-compensation mechanisms to normalize expression from each gene or chromosome (Hughes *et al.*, 2000).

The possible deletion of one chromosome in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> points to an unstable genome in this strain, thus, are the observed expression changes biased for the repressed genes only in chromosome XIV? Or are the results for the induced genes also altered due to an indirect effect of the partial loss of some genes? A possible approach to determine chromosomal alterations would be to isolate genomic DNA from both control and mutant strains, label and hybridize to DNA microarrays (Hughes *et al.*, 2000). Alternatively, karyotype and flow cytometry analyses were undertaken and will be discussed later (chapter III.3.7.).



**Figure 22 – Distribution of the genes whose expression is altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub>, according to their chromosomal localization.</sub>** Induced genes are represented in red and repressed genes in green. Genes whose expression levels are unaltered are shown in black. In chromosome XIV there is an accumulation of down-regulated genes, which could reflect the deletion of one copy of this chromosome.

### 1.4. Transcriptome of *S. cerevisiae* expressing the ser-tRNA<sub>CAG-G<sub>33</sub></sub>



**Figure 23 – SAM plot of the gene expression data obtained from *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG-G<sub>33</sub></sub>.** The graphic shows the observed vs. expected mRNA levels, with the  $\Delta$  value set to 0.75, resulting in a false discovery rate of 0.1 % in 357 genes. Genes with significantly increased or decreased expression levels are coloured in red and green, respectively.

For this chapter, only the transcriptome alterations of the *C. albicans* ser-tRNA<sub>CAG</sub> G<sub>33</sub> mutant were considered and, therefore, a specific response to low-level CUG ambiguity will be discussed. SAM analysis with a  $\Delta$  value set to 0.75 resulted in a false discovery rate of 0.1 % in 357 genes (Figure 23). Comparison of this list with the GeneSpring data resulted in a selection of genes shown in Table 11.

**Table 11 – Genes whose expression is altered exclusively in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-G<sub>33</sub>.** Genes shown in red are induced and genes in green are repressed, and the respective fold variation is indicated (FDR=0.001).

Gene name	Fold variation T	Fold G
CTS1	1.4	2.9
CAR2	1.5	2.6
GSY2	1.4	2.5
CRR1	1.2	2.4
MAL32	1.4	2.4
MAL31	1.6	2.3
MPC54	1.6	2.1
DIT1	1.4	2.1
SPO22	1.1	2.0
ARG3	1.6	2.9
MET1	1.4	2.5
ARG8	1.3	2.2
MET16	1.4	2.0

Carbohydrate and amino acid metabolism, cell wall and sporulation are the functional categories that seem altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-G<sub>33</sub> but not in the strain carrying the *C. albicans* ser-tRNA<sub>CAG</sub>-T<sub>33</sub>.

*CTS1* encodes an enzyme that cleaves the cell wall components glucan and chitin to promote mother-daughter cell separation, thereby impairing pseudohyphal growth (Pan *et al.*, 2000). *CRR1*, *MPC54* and *DIT1* are required for meiosis or sporulation.

*CAR2* acts in the degradation of arginine, citrulline and ornithine. The increase in their catabolism might be related with the decrease in amino acid biosynthetic pathways, as pointed by the repression of *MET* and *ARG* genes, since the availability of arginine would result in both *CAR2* activation and repression of the *ARG* genes (Park *et al.*, 1999).

*MAL31* is a high affinity maltose permease, and *MAL32* is a maltase, suggesting that these cells could be using alternative carbon sources, similarly to heat-shocked or stationary-phase cells (Gasch *et al.*, 2000).

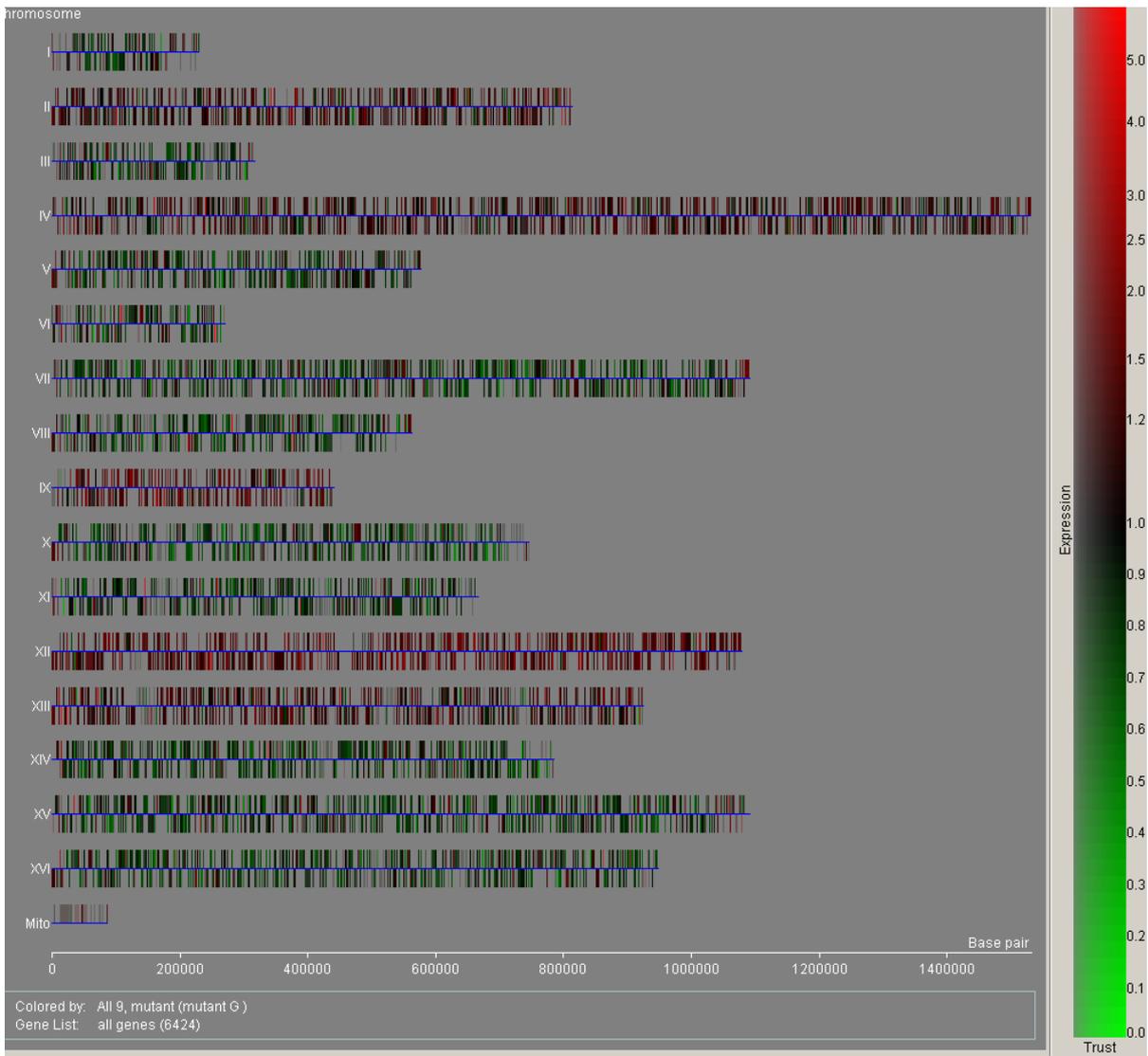
*GSY2* is one of the isoforms of the glycogen synthase. The other isoform is encoded by *GSY1* that shows unaltered expression levels in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (1.5 fold in the G<sub>33</sub> and 1.2 fold in the T<sub>33</sub> strain). *GLC3* is a glycogen branching enzyme necessary for glycogen synthesis and, although not significant in the analysis performed, also has increased expression levels in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> G<sub>33</sub> (2.4 fold) but not in the T<sub>33</sub> strain (1.5 fold). This trend is supported by further observations in which the G<sub>33</sub> cells accumulate glycogen at higher levels than the T<sub>33</sub> strain (chapter III.3.4.). However, the induction of *GPH1* (Table 5 of the previous chapter), which releases glucose from glycogen, indicates an enhanced recycling of this reserve due to simultaneous up-regulation of enzymes involved in its synthesis and degradation, as previously suggested (Parrou *et al.*, 1997).

Other genes were found significantly altered in both *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, however the fold variation was more pronounced in the G<sub>33</sub> than in the T<sub>33</sub> cells (Table 12).

**Table 12 – Genes whose expression is induced more significantly in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-G<sub>33</sub>.** The respective fold variation is indicated (FDR=0.001).

Gene name	Fold variation T	Fold G
PUT1	2.2	5.3
MSC1	2.6	4.5

*MSC1* affects meiotic homologous chromatid recombination, supporting the previously described increase in the expression levels of genes related to meiosis and sporulation (Table 11) in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-G<sub>33</sub>, which was not observed in the T<sub>33</sub> strain. The induction of *PUT1* could reflect nitrogen starvation, but again this observation is in disagreement with the repression of amino acid biosynthesis (Table 11).



**Figure 24 – Distribution of genes whose expression is altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-G33</sub>, according to their chromosomal location.** Induced genes are represented in red and repressed genes in green. Genes whose expression levels are unaltered are shown in black. The chromosome-wide expression biases are highlighted, with some chromosomes accumulating induced genes and others repressed ones. In particular, chromosomes IX, XII and XIII appear duplicated while chromosomes X, XI, XIV, XV and XVI appear deleted.

Like for the *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-T<sub>33</sub> data, the genes whose expression levels are altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-G<sub>33</sub> were mapped according to their chromosomal location (Figure 24).

The chromosome-wide expression biases are very pronounced in this strain, with about half the chromosomes accumulating induced genes and the other half repressed ones. In particular, chromosomes IX, XII and XIII appear duplicated while chromosomes X, XI, XIV, XV and XVI appear deleted. If chromosomal duplications and deletions affect significantly mRNA abundance levels, it would be expected that all the genes located on the same chromosome would show a similar fold increase or decrease, respectively. Nevertheless, there is a handful of examples that do not follow this trend. *SRY1* shows a 3-fold decrease in the G<sub>33</sub> strain and is located on chromosome XI that might have suffered a deletion. However, this gene is also down regulated in the T<sub>33</sub> strain that did not lose a copy of this chromosome. Moreover, many other genes located on chromosome XI are induced in both strains, namely *PIR3* and *CWP2*. *ATR1* shows a 2-fold repression in both G<sub>33</sub> and T<sub>33</sub> mutants and is located on chromosome XIII that is also supposedly deleted on the G<sub>33</sub> but not on the T<sub>33</sub> strain. *YGP1* is located on chromosome XIV and is induced 9-fold in both G<sub>33</sub> and T<sub>33</sub> mutants, even after apparent loss of one copy of the chromosome in both strains. Even admitting that chromosomal alterations could influence drastically gene expression levels, and hence the fold variations determined would not correspond to the real effect of the ser-tRNA<sub>CAG</sub>, the most important target genes will eventually be kept.

It is likely that these results are the additive consequences of CUG mistranslation and aneuploidy, but they were both induced by the presence of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* cells. The probability of some errors on gene expression data cannot be discarded, however the impact of CUG ambiguity on the genome is undoubtedly strong.

## 1.5. Discussion

The 81 genes whose expression is altered in response to CUG mistranslation are much lower than the numbers observed with other stress conditions, like the 458 genes induced in zinc-limited yeast cells (Lyons *et al.*, 2000). Upon AZC exposure, 217 genes were induced and 293 genes were repressed, and the fold variation was between 3 and 30 fold (Trotter *et al.*, 2002). It has been shown that transcript levels behave transiently, and after a short period of big changes in gene expression the mRNA abundance of most genes returns to new steady-state levels, close to the normal values, as cells adapt to the stressful condition imposed (Gasch *et al.*, 2000). Because the expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* cells is a permanent and internal stress, and not a transient environmental condition, the peak of gene expression changes might have been missed in our experimental model. Additionally, the selection of target genes for CUG ambiguity was based on too stringent criteria, by combining the results of two distinct analyses. Only the common genes showing significantly altered expression levels in both GeneSpring and SAM analysis were considered, and were further selected using their fold-variation. Nevertheless, the important final targets of CUG ambiguity were retained and allow a global characterization of the response triggered by the expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* cells.

The transcriptional profile of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> showed that the stress response is the functional group of genes most affected. This group includes genes belonging to protein folding, carbohydrate metabolism and other stress-responsive genes that show increased expression levels in CUG ambiguous cells. Genes related to protein degradation, cell wall and transporters, and phosphate metabolism were found to be mainly up-regulated, with a few genes down-regulated. Protein synthesis was repressed and amino acid metabolism includes genes that were both induced and repressed. The results also indicate that the structure of the cell wall is altered, with many genes showing increased expression levels. It would be interesting to assess if this altered structure induces stress resistance or if the changes are mainly structural,

for instance by determining the sensitivity of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> to compounds that interfere with the polymerization of the cell wall, such as Calcofluor white, Congo red or hygromycin B (Martinez-Lopez *et al.*, 2004). Other approach would be to determine if the regulators of the cell integrity signalling pathway are induced (Jung and Levin, 1999). Another puzzling question is the repression of the PKA pathway even in the presence of glucose, which suggests that CUG ambiguity could be mimicking nutrient starvation by interfering with signalling pathways. It is, therefore, possible that the constitutive activation of the PKA pathway would result in the loss of stress resistance, trehalose and glycogen accumulation and other events associated with elevated PKA activity, therefore decreasing viability of these cells.

In an attempt to find a condition that would mimic the cellular response elicited by genetic code ambiguity, *S. cerevisiae* cells were exposed to canavanine, an arginine analogue (data not shown). It turned out that canavanine does not mimic CUG ambiguity and it does not mimic AZC-treatment as well (Trotter *et al.*, 2002). AZC or ethanol treatment, as the changes induced by ethanol are similar to those described with AZC, could be better positive controls. However, there is some evidence supporting the hypothesis that CUG mistranslation prompts a unique stress response, different from temperature upshift or AZC-induced protein misfolding. One question remaining is whether or not there is activation of the unfolded protein response in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>. UPR induces protein degradation even in the absence of folding stress, but AZC treatment induces misfolding without activating the UPR and increases the expression levels of proteasome subunit genes (Trotter *et al.*, 2002). Treatment of cells with the drug tunicamycin that inhibits N-linked glycosylation, or the reducing agent dithiothreitol (DTT), which prevents disulfide bond formation, induces the UPR by impairing normal ER function without disrupting protein folding outside the secretory pathway (Travers *et al.*, 2000). This could be a good strategy to assess if there is UPR activation in response to CUG ambiguity.

The distribution, by chromosome, of the genes altered by genetic code ambiguity was visualized in GeneSpring, and pointed to a bias of genes with increased expression levels located on the same chromosomes. The same was verified with the genes with decreased expression levels, suggesting that there might have been deletions and duplications in these strains. These observations suggest that *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> might suffer genomic instability, and flow cytometry analysis of the DNA content confirmed the suspected aneuploidy (Rita Rocha, unpublished results). These results will be discussed later (chapter III.3.7.).

There were not many differences between the gene expression programmes of the mutant T<sub>33</sub> and wild type G<sub>33</sub> ser-tRNA<sub>CAG</sub>. The global response to CUG ambiguity is the same, and the main gene targets are common between cells expressing high and low genetic code ambiguity. It cannot be ruled out that some differences observed in gene expression levels might be due to chromosomal abnormalities and not real differences between the two tRNA strains. However, it has been observed that the toxicity of a certain agent might not relate directly to the quantitative transcriptional response of the cell, predicted by the number of genes altered (Jelinsky *et al.*, 2000). A treatment that results in less cell survival, and so the most toxic, alters the expression of fewer genes than a less toxic one (Jelinsky *et al.*, 2000). It is tempting to speculate that, under this reasoning, the ser-tRNA<sub>CAG</sub> T<sub>33</sub> is more toxic than the G<sub>33</sub>, an hypothesis that is supported by the higher decrease in growth rate induced by the T<sub>33</sub> mutant tRNA (Santos *et al.*, 1996).

The appearance of the ser-tRNA<sub>CAG</sub> on the ancestor produced drastic alterations on gene expression, as observed in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>. The impact of this event was global, with changes in the expression of structural genes from the cell wall, genes involved in metabolic pathways, namely carbohydrate, amino acid and phosphate metabolism, and alterations in the stress response. Such gene expression changes that result from the ambiguous translation of the CUG codon have implications for the adaptation,

physiology and evolution of organisms, for example, by enabling cells to explore novel ecological niches. Furthermore, expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* cells had a dramatic effect on the structure of the genome, suggesting that the genetic code change triggered by the ser-tRNA<sub>CAG</sub> produces not only proteome but also genome instability.

*S. cerevisiae* cells expressing CUG ambiguity are permanently under stress. This new type of stress is distinct from other forms of stress due to its constitutive nature, which results in proteome and genome destabilization. Therefore, will gene expression be regulated at the translational level, as observed in other stress responses?

## **2. Proteome analysis of *S. cerevisiae* cells expressing CUG ambiguity**

### **2.1. Introduction**

Integrating data from both transcriptomics and proteomics is the best approach for the comprehensive characterization of a biological system, as they provide complementary information (Ideker *et al.*, 2001; Griffin *et al.*, 2002). Indeed, measuring the expression levels for all proteins, as it can be accomplished by DNA-microarrays for the expression levels of all genes, would be the ideal experiment for global and complete characterization of a cell under a given condition, although such a system is not yet available.

In our experimental model, the introduction of the *C. albicans* ser-tRNA<sub>CAG</sub> on *S. cerevisiae* imposes a permanent and internal stress on these cells, due to proteome destabilization caused by mistranslation of the CUG codon. Indeed, the induction of a permanent stress response by CUG ambiguity may reprogramme the translational machinery, as observed in other cells submitted to stress conditions. It has been demonstrated for *S. cerevisiae* cells that in many cases mRNA abundance does not correlate with protein abundance (Gygi *et al.*, 1999; Griffin *et al.*, 2002), due to translational regulatory events. Such translational control of gene expression occurs in many organisms (Clemens, 1996; Day and Tuite, 1998), namely under stress conditions such as nutrient limitation (Hinnebusch, 1996), heat shock (Duncan, 1996) and viral infection (Katze, 1996; Matthews, 1996; Schneider, 1996), resulting from the reversible phosphorylation of several initiation factors like eIF2A or eIF4E, among others. In fact, the initiation of translation in eukaryotes is a complex process that involves many factors and the rate of initiation often limits translation (Clemens, 1996). The first step in translation initiation consists on the recognition and binding of the initiator tRNA by the initiation factor eIF2A, which forms a ternary complex with the Met-tRNA<sub>i</sub> and GTP. The hydrolysis of GTP occurs during ribosomal subunit joining, releasing the eIF2A bound to GDP, which is regenerated by eIF2B producing an eIF2A-GTP available for a new initiation (Trachsel, 1996). Phosphorylation of the A-subunit of

the initiation factor eIF2 impairs the exchange of GDP for GTP, inactivating its recycling capability, resulting in translation inhibition (Trachsel, 1996). This decreased rate of protein synthesis limits energy expenditure under adverse growth conditions, however, in order to promote cell survival, several specific mRNAs have to be translated, depending on the nature of the stress. For example, in *S. cerevisiae* growing under nutrient deprivation the protein kinase responsible for the phosphorylation of eIF2A is GCN2 (Hinnebusch, 1996). GCN2 is involved in the translational regulation of GCN4 synthesis during nitrogen depletion, as GCN4 is the transcription factor that controls the expression of amino acid biosynthetic genes. The mRNA of GCN4 has four short upstream open reading frames (uORFs), whose function is to prevent translation initiation when nutrients are present. But under starvation conditions, GCN2 phosphorylates eIF2A and the rate of protein synthesis decreases, reducing the ribosomes that will reinitiate after translation of uORF1. These ribosomes will bypass uORFs 2-4 enabling translation of GCN4 instead, which will activate transcription of the genes belonging to amino acid biosynthesis pathways, to allow cell survival (Hinnebusch, 1996).

Translational control also occurs on cells submitted to endoplasmic reticulum stress that activates the unfolded protein response (UPR) (reviewed in Patil and Walter, 2001; Ma and Hendershot, 2001). In yeast, the UPR depends on only three gene products: Ire1p, Rlg1p and Hac1p. *HAC1* mRNA is constitutively expressed but the Hac1 protein is only translated after ER stress (Cox and Walter, 1996). It was suggested that regulation of *HAC1* at the translational level could allow a rapid production of the required amount of protein from a stable mRNA pool that is not translated unless needed (Patil and Walter, 2001). ER stress leads to the oligomerization and activation of the kinase and endoribonuclease Ire1p, which in turn cleaves the intron from the *HAC1* mRNA. The two exons are rejoined by Rlg1p, which allows the production of the protein Hac1. Hac1p binds to the UPRE elements present in the promoters of ER chaperones (*KAR2*, *PDI1*, and *FKB2*), activating their transcription and inducing the UPR (Mori *et al.*, 1996). In mammals and *C. elegans*, three signalling pathways seem to mediate the

activation of the UPR. First, the activation of the PERK kinase leads to an increased eIF2A phosphorylation that inhibits translation during ER stress (Ma and Hendershot, 2001). Second, cleavage of the ER-transmembrane protein Atf6 results in the transport of its transactivation domain to the nucleus where transcription of ER chaperones is induced. Third, Ire1p deletes a fragment of the *XBP-1* mRNA, which causes a frame-shift producing a “chimeric” Xbp-1 that upregulates the expression of ER chaperones. While higher eukaryotes inhibit protein synthesis, yeast increases protein degradation to minimize cellular damage in response to the UPR (reviewed in Patil and Walter, 2001; Ma and Hendershot, 2001).

Heat shock is another stress system that reprogrammes translation, mainly by repressing general protein synthesis and translating preferentially the mRNAs that encode heat shock proteins that enable cell survival (reviewed by Duncan, 1996). Again, phosphorylation of initiation factors plays an important role in decreasing translation but other mechanisms might also be at work for selective translation during heat shock. As an example, the molecular chaperone Hsp70 regulates its own translation, and after accumulating to a certain threshold initiates its down-regulation (DiDomenico *et al.*, 1982a; DiDomenico *et al.*, 1982b).

Transcriptome analysis of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (chapter III.1.) did not detect increased expression of the mRNAs of the transcriptional factors normally implicated on the stress responses. However, their downstream targets are induced, namely the genes that code for amino acid transport, molecular chaperones or genes whose products have anti-oxidant properties. This may be related to the fact that most transcriptional factors are regulated by post-translational modification. For instance, Hsf1p is regulated by phosphorylation in response to a temperature rise (Bonner *et al.*, 2000) and Yap1p is relocalized to the nucleus by oxidative stress, without altering their expression levels (Owsianik *et al.*, 2002). Because many stress conditions decrease overall translation, the cells have to find alternative pathways to increase expression or activity of the proteins needed to face the adverse conditions. The possibility of an

---

additional control at the translational level increases the complexity of gene expression regulation, but enables a more direct and quick response from the cell to environmental changes than transcriptional regulation, because there is no need to synthesize, process or transport new molecules of mRNA. Additionally, as most translational regulatory mechanisms rely on protein phosphorylation, the reversibility of the process allows cells to do a better management of the pool of existing mRNAs for their profit, with energy economy when conditions are not favourable.

Expression of the *C. albicans* ser-tRNA<sub>CAG</sub> on *S. cerevisiae* cells permanently alters gene expression, as showed in the previous chapter. However, the facts that CUG ambiguity destabilizes the proteome and that translational regulation occurs under stress raise the questions of “Are the changes at the mRNA level representative of the alterations occurring in the cell at the proteome level?”, “Is microarray data reflecting the impact of this permanent stress on the cells?” or “Are these cells regulating gene expression at the translational level?”. In order to elucidate these questions, the proteome of *S. cerevisiae* expressing the *C. albicans* ser-tRNA<sub>CAG</sub> was characterized using quantitative proteomics methodologies.

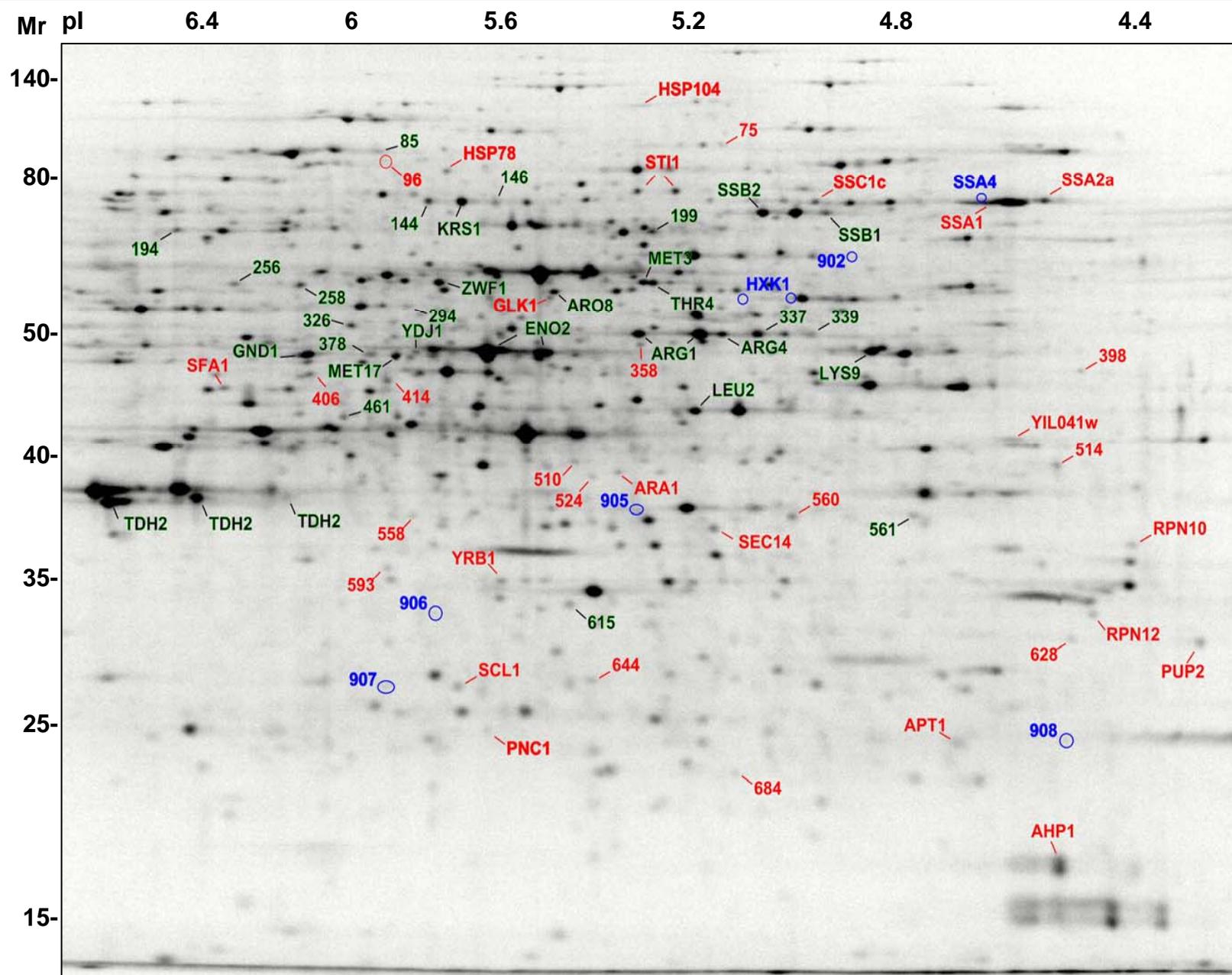
## 2.2. Proteome of *S. cerevisiae* cells expressing CUG ambiguity

For proteome characterization, *S. cerevisiae* control cells and cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (T<sub>33</sub>) were grown at 25°C with shaking until the cultures reached an OD<sub>600nm</sub> of 0.5. After *in vivo* metabolic labelling with <sup>35</sup>S-Methionine, proteins were extracted and separated by 2D-PAGE and gels were analysed as described in methods (chapter II.3.3.).

The proteome map obtained from *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (T<sub>33</sub>) is presented in Figure 25. Because some proteins are represented on the 2D map by more than 1 spot, the global number of spots altered in response to CUG ambiguity and the respective combination of identified and unidentified proteins is not the same. Overall, 77 protein spots showed altered expression levels in cells that mistranslate the CUG codon. Of these, 43 spots have increased expression levels corresponding to 19 identified proteins and 15 unidentified spots, plus 8 new spots that are not expressed in the control cells grown at 25 °C. From these new spots, 3 correspond to previously identified proteins (Annex, Table I). The 34 spots showing decreased expression levels in cells expressing genetic code ambiguity, correspond to 15 identified proteins and 14 unidentified spots. Although expressed in the control condition, 2 of these repressed proteins are not present in the ambiguous cells, and were, therefore, considered as disappeared from the 2D-gel map (Annex, Table I).

### **Figure 25 – Proteome alterations induced by CUG ambiguity in *S. cerevisiae*.**

The map shows in black the spots corresponding to the control cells and indicates in colour the proteins whose expression level is altered by at least 2 fold in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub>: in red, proteins that are induced; in green, proteins that are repressed; and in blue, proteins that are expressed uniquely in cells that decode the CUG codon ambiguously. Proteins were labelled *in vivo* with [<sup>35</sup>S-Methionine].

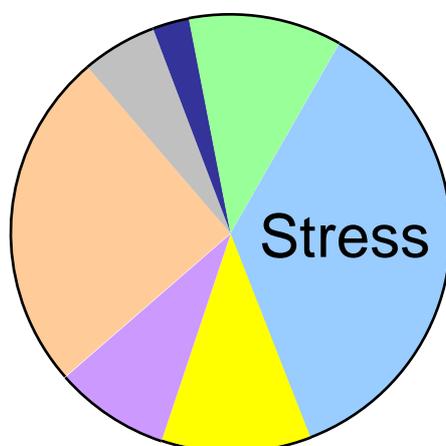


The functional distribution of proteins altered on *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> (as indicated on the SGD web page, <http://www.yeastgenome.org>), showed that the group most affected is the stress response (Figure 26A). Moreover, these proteins have increased expression levels in response to CUG ambiguity, with only one exception (see below). This stress category includes proteins associated with protein folding and carbohydrate metabolism, as well as proteins responsive to general stress (Figure 26B). Other major functional group showing expression alteration is the group of proteins belonging to amino acid metabolism. All these proteins and several proteins belonging to the protein synthesis group show decreased expression levels. Figure 26A also highlights protein degradation and carbohydrate metabolism as important functions altered by CUG ambiguity, since the members of the protein degradation group are up regulated, while members of the carbohydrate metabolism group are either induced or repressed.

### **2.2.1. CUG ambiguity induces the expression of stress proteins**

*S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> show increased levels of stress proteins (Table 13) such as the molecular chaperones Hsp104, Hsp78 and members of the Hsp70 family (namely the Ssa subfamily) that are involved in protein folding. The expression of the Ssa1 gene is high at physiological temperature and is stimulated under stress conditions, such as temperature shift to 37°C, whereas the Ssa2 gene is expressed at the same level at all temperatures. The Ssa3 and Ssa4 genes are expressed at very low levels at physiological temperature (25°C) but the amount of their mRNAs increase several fold after temperature upshift (Werner-Washburne *et al.*, 1987). Proteome analysis of yeast cells submitted to heat shock also showed increased expression levels of the Ssa1-4 proteins (Boy-Marcotte *et al.*, 1999).

A



B



■ Protein degradation

■ Protein synthesis

■ Amino acid metabolism

■ Other

■ Unknown

■ Carbohydrate metabolism

■ Stress

■ Stress

■ Protein folding

■ Carbohydrate metabolism

**Figure 26 – Proteins whose expression is altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> distributed by functional categories.** A - proteins whose expression is both up and down regulated by genetic code ambiguity were grouped according to their functions. The graphic on B represents the proteins that were included in the stress group from graphic displayed on A.

On the other hand, oxidative stress results in the induction of the Ssa1 and Ssa3 proteins following H<sub>2</sub>O<sub>2</sub> treatment (Godon *et al.*, 1998), and Ssa1, Ssa2 and Ssa4 proteins after cadmium exposure (Vido *et al.*, 2001). According to the proteome results obtained for heat shock and oxidative stress conditions, cells expressing genetic code ambiguity showed an increase in the expression of the Ssa1, Ssa2 and Ssa4 proteins (Table 13).

**Table 13 – Stress proteins and chaperones whose expression is altered by genetic code ambiguity, indicating the respective fold variation and statistical significance.** In red, proteins whose expression is increased and in green, proteins whose expression is decreased. Proteins that are not expressed in the control condition are considered new (n) spots and proteins that are not expressed in the ambiguous cells are considered disappeared (d) from the proteome and, therefore, their fold variation cannot be accurately determined. An average fold represents proteins that are present in the 2D gel by more than one spot. Nd stands for not determined.

Protein	Fold	P-value
Pnc1	29.5	0.0154
Hsp104	13.1	nd
Ssa1	4.7	0.0007
Ahp1	3.4	0.0021
Ssc1	3.2	0.0015
Sti1	3.0	0.0008
Hsp78	2.7	0.00002
Ssa2	2.5	0.0091
Ssa4	n	nd
Ydj1	d	nd

Although the nature of the stresses is different, as both the temperature upshift and the oxidative stress consist of a transient and external change in the environment, while the introduction of the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> in *S. cerevisiae* cells triggers a permanent and internal stress, all these conditions result on the accumulation of unfolded and altered proteins. Therefore, the cell

counteracts the negative impact of the production of aberrant proteins by inducing the molecular chaperones to assist protein folding.

The Sti1 protein showed a 3-fold induction in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> (Table 13). This protein has increased expression levels under other stress conditions, namely after heat shock and cadmium exposure (Boy-Marcotte *et al.*, 1999; Vido *et al.*, 2001) and activates the Ssa proteins (Wegele *et al.*, 2003). Indeed, the only members of the Hsp70 family that bind to Sti1p are the Ssa proteins (Ssa1-4) and Sti1p can increase the ATPase activity of Ssa1p up to 200-fold, whereas other stimulation factors such as Ydj1p show no more than a 10-fold activation (Wegele *et al.*, 2003). It is worth noting that all stress proteins altered by CUG ambiguity have increased expression levels, with the exception of Ydj1p whose expression has diminished so much that it can no longer be detected on the 2D-gel. A possible explanation for this is that the function of Ydj1p as activator of the Ssa1 protein has been replaced by Sti1p.

Other proteins included in the stress group are Pnc1 and Ahp1 (Table 13). Ahp1 is a protein with antioxidant properties and its expression is also increased about 3-fold after exposure of yeast cells to H<sub>2</sub>O<sub>2</sub> (Godon *et al.*, 1998) or cadmium (Vido *et al.*, 2001). Pnc1 is involved in lifespan extension by calorie restriction (Lin and Guarente, 2003; Anderson *et al.*, 2003), and it has been shown that a yeast strain carrying additional copies of the *PNC1* gene lived longer than the wild type strain (Anderson *et al.*, 2003). The expression of Pnc1 increases in response to several stresses (Godon *et al.*, 1998; Boy-Marcotte *et al.*, 1999; Vido *et al.*, 2001; Trotter *et al.*, 2002), promoting viability under adverse growth conditions. As the protein that is most induced in *S. cerevisiae* strain expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub>, with a 29.5-fold increase in the expression level, Pnc1p contributes to maintain cell viability allowing for tolerance to mistranslation of the CUG codon (chapter III.3.2.).

### 2.2.2. CUG ambiguity alters the expression of carbohydrate metabolism enzymes

Carbohydrate metabolism proteins altered by genetic code ambiguity are shown in Table 14. Proteins involved in carbohydrate metabolism, namely Zwf1, Glk1, Hxk1, Hor2 and Tps1, are up-regulated in other stress conditions such as oxidative stress (Godon *et al.*, 1998; Vido *et al.*, 2001) or heat shock (Boy-Marcotte *et al.*, 1999). The increased expression of carbohydrate metabolism enzymes is thought to control the glycolytic flux in order to provide energy for the activity of molecular chaperones and the proteasome, since both folding and degradation processes consume ATP (Gasch *et al.*, 2000).

**Table 14 – Carbohydrate metabolism proteins whose expression is altered by genetic code ambiguity, indicating the respective fold variation and statistical significance.** In red are proteins whose expression level is increased and in green proteins whose expression level is decreased. Proteins that are not expressed in the control condition are considered new (n) spots and proteins that are not expressed in the ambiguous cells are considered disappeared (d) from the proteome and, therefore, their fold variation cannot be accurately determined. An average fold represents proteins that are present in the 2D gel by more than one spot. Nd stands for not determined.

Protein	Fold	P-value
Glk1	5.0	0.0091
Hxk1	n	nd
Hor2	n	nd
Zwf1	3.7	0.0032
Gnd1	2.3	0.0008

Despite a 2-fold increase on Tps1p expression level, this result was not considered significant due to the high variability of the proteomics data for this spot. Nevertheless, increased mRNA expression levels of the *TPS1* gene were detected and the accumulation of trehalose was observed in cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> (chapters III.1.2.4. and III.3.4.), indicating that the Tps1p expression data is valid and also that its activity is induced in CUG ambiguous cells. These enzymes could be regulated at the post-translational level, since some authors refer the increased expression of Tps1p after H<sub>2</sub>O<sub>2</sub> treatment without trehalose accumulation, suggesting an enhanced recycling of this reserve carbohydrate (Parrou *et al.*, 1997; Godon *et al.*, 1998). Trehalose is an important protein stabilizer (Kaushik and Bhat, 2003) and, therefore, the accumulation of this disaccharide could play a major role in the protection of cells from denatured proteins resulting from the mistranslation of the CUG codon (chapter III.3.4.).

The induction of Hor2 suggests that glycerol synthesis can be an important response to CUG ambiguity, since glycerol is a major osmolyte in *S. cerevisiae*, and has an important role in the protection from osmotic and even thermic stress (Siderius *et al.*, 2000).

Also important is the repression of Zwf1 and Gnd1 proteins that catalyse the first two steps of the pentose phosphate pathway, generating NADPH and ribulose 5-phosphate, a precursor of purine biosynthesis (Schaaff-Gerstenschlager and Zimmermann, 1993). Conversely, during oxidative stress the pentose phosphate pathway is induced, while glycolysis is repressed, mainly to produce NADPH, a cofactor of the enzymes that counteract oxidative damage (Godon *et al.*, 1998). Thus, the repression of Zwf1p and Gnd1p with the simultaneous induction of Glk1p and Hxk1p indicate that glycolysis, together with trehalose and glycerol synthesis, are the preferential pathways for the glucose flux in cells expressing genetic code ambiguity.

### 2.2.3. CUG ambiguity induces the expression of proteasome subunits

Protein degradation is another functional group induced by genetic code ambiguity, with several proteasome subunits showing increased expression levels between 3 and 5 fold (Table 15). The mitochondrial protease Hsp78p shows increased expression levels in response to CUG mistranslation as well (Table 13). Proteome characterization of conditions that cause oxidative stress also showed induction of proteasome subunits, for example, in the yeast response to cadmium the expression of these proteins increased by a factor that ranged from 3 to 7-fold (Vido *et al.*, 2001), and after H<sub>2</sub>O<sub>2</sub> treatment proteasome subunits were up-regulated between 2 and 3-fold (Godon *et al.*, 1998). One of the transcriptional factors involved in the yeast response to oxidative stress is Yap1p (Lee *et al.*, 1999). The presence of a regulatory sequence that binds Yap1p in the promoter of RPN4, which regulates the expression of proteasome subunits (Mannhaupt *et al.*, 1999), can account for the increased expression of proteasome subunits in response to oxidative compounds (Owsianik *et al.*, 2002). Although there may be oxidatively damaged proteins in cells expressing CUG ambiguity (chapter III.3.5.), the induction of the proteasome is most likely related with the increased production of unfolded mistranslated proteins as a consequence of CUG ambiguity.

**Table 15 – Proteasome subunits whose expression is induced by genetic code ambiguity.** The respective fold increase and statistical significance are indicated.

Protein	Fold	P-value
Rpn12	5.2	0.0230
Rpn10	3.6	0.0065
Pup2	3.6	0.0010
Scl1	3.6	0.0228

Microarray analysis of cells submitted to AZC treatment, a proline analogue that induces misfolding, showed an increase in the mRNA levels of both regulatory and

structural proteasome subunits, including *RPN12* (4.3-fold) and *SCL1* (3.6-fold) (Trotter *et al.*, 2002). On the other hand, some observations indicate that there were no changes in the expression levels of proteasomal genes upon temperature upshift (Eisen *et al.*, 1998; Trotter *et al.*, 2002), and these proteins were not overexpressed under heat shock conditions (Boy-Marcotte *et al.*, 1999). Although the induction of protein degradation was considered as part of the environmental stress response (Gasch *et al.*, 2000), this functional group included genes belonging to the ubiquitin-conjugation pathway and other proteases, but no proteasome subunits. Hsp78, on the other hand, shows increased mRNA and protein expression levels in many stress conditions (Godon *et al.*, 1998; Boy-Marcotte *et al.*, 1999; Gasch *et al.*, 2000; Vido *et al.*, 2001; Trotter *et al.*, 2002). These results suggest that proteasome induction could, therefore, be a hallmark of the type of stress caused by the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub>, related to severe protein misfolding and not part of a general stress response. Indeed, it has been shown that increased proteasome activity can replace the induction of heat shock proteins (Hsps) during heat shock, in the removal of toxic protein aggregates (Friant *et al.*, 2003). According to this hypothesis, the key role of the molecular chaperones that allows cell survival under stress is not their refolding activity to recover non-functional proteins, but instead preventing aggregation of unfolded proteins. So, a stress that causes medium levels of denatured proteins can be solved solely with the induction of Hsps, whereas a severe misfolding stress would have to be backed up with increased proteasome activity. As CUG mistranslation is constitutive and produces an excess of aberrant proteins that the cell must get rid of, the increase in expression of molecular chaperones might not be enough to correct all misfolded proteins and prevent their aggregation, and consequently the induction of protein degradation by the proteasome is required.

#### 2.2.4. CUG ambiguity represses the expression of amino acid metabolism enzymes

Amino acid metabolism pathways are repressed in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>, as shown in Table 16. The repression of the amino acid biosynthesis pathways had been observed before during the heat shock response (Boy-Marcotte *et al.*, 1999). The repression of the sulfur amino acid pathway has also been reported under oxidative and osmotic stress (Norbeck and Blomberg, 1997; Godon *et al.*, 1998; Vido *et al.*, 2001).

**Table 16 – Amino acid metabolism proteins whose expression is repressed by genetic code ambiguity.** The respective fold decrease and statistical significance are indicated.

Protein	Fold	P-value
Met17	5.3	0.0004
Aro8	5.1	0.000007
Arg4	4.4	0.05
Arg1	3.4	0.005
Met3	2.9	0.03
Lys9	2.8	0.006
Leu2	2.3	0.0001
Thr4	2.0	0.008

One could argue that the amino acid biosynthetic pathways should be induced due to the activation of the transcriptional regulator *GCN4*, as cells are grown in the absence of most amino acids (Hinnebusch and Natarajan, 2002). Nevertheless, *S. cerevisiae* has specific repression mechanisms regulated by the end-product of the pathway, to override the general mechanism controlled by *GCN4*, namely for the biosynthesis of arginine, lysine, leucine, isoleucine-valine, methionine and threonine. Yet, for the other proteins the conclusions are not so straightforward. As the folds calculated are expression levels relative to the control cells, and not absolute values, these proteins could be in fact induced in both strains in response

to starvation, but less increased in the tRNA expressing cells. Or, these cells might be impaired in their response to nutrient starvation, while the control cells are not, and hence the apparent down-regulation. The question then is “why do the amino acid biosynthetic pathways show less induction on *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>?”.

The activation of *GCN4* translation can occur by a GCN2-independent mechanism that requires increased PKA function (Hinnebusch, 1996). As discussed in the previous chapter, it is possible that PKA activity is low in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, and this could repress the expression of *GCN4* resulting in reduced expression levels of amino acid biosynthesis enzymes. Alternatively, perhaps the increased protein degradation, deduced from the up-regulation of the proteasome, is providing an internal source of amino acids, allowing the cells to bypass the *GCN4* control even growing in minimal media. This mechanism might share similarities to autophagy, which occurs in yeast under nutrient starvation, namely nitrogen, carbon, sulphate or phosphate depletion (Takeshige *et al.*, 1992). Autophagy targets proteins for degradation by the vacuole, in a process that also involves conjugation of ubiquitin-like proteins to their substrates (Ohsumi, 2001). Mutations in genes belonging to these conjugation processes prevent adaptation of yeast cells to nutrient starvation and impair cell survival, by blocking bulk protein degradation (Ohsumi, 2001). Microarray data showed a significant increase in the expression of the gene *AUT7*, whose product is required for delivery of autophagic vesicles to the vacuole, in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>. Thus, this result can hint on the induction of additional pathways that can function as recycling mechanisms, to provide the required nutrients and enhance cell survival under the stress caused by CUG ambiguity.

### 2.2.5. CUG ambiguity represses protein synthesis

Similarly to several stress conditions (Gasch *et al.*, 2000), protein synthesis enzymes are repressed by genetic code ambiguity (Table 17). The repression of the Ssb subfamily of the Hsp70 family of molecular chaperones has been reported in response to heat shock (Werner-Washburne *et al.*, 1989), H<sub>2</sub>O<sub>2</sub> treatment (Godon *et al.*, 1998) and osmotic stress (Norbeck and Blomberg, 1997), along with ribosomal proteins and translation factors.

Ssb1 and Ssb2 are ribosome-associated heat shock proteins that play a role in the folding of nascent polypeptides, and the fact that they bind to yeast HSF has been proposed as a mechanism to fine-tune HSF activity both during heat shock and normal growth conditions (Bonner *et al.*, 2000). Similarly, in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, these proteins may be important to regulate HSF activity and modulate the stress response.

**Table 17 – Protein synthesis enzymes whose expression is repressed by genetic code ambiguity.** The respective fold decrease and statistical significance are indicated.

Protein	Fold	P-value
Krs1	2.7	0.0011
Ssb1	2.1	0.0167
Ssb2	2.1	0.0052

### 2.2.6. CUG ambiguity regulates gene expression at the translational level

As described in the previous chapter, the environmental stress response (ESR) consists of a common set of genes induced and repressed by stress conditions (Gasch *et al.*, 2000). An interesting result concerning the ESR is the differential expression of isozymes and the induction of genes whose products have reciprocal functions. These mechanisms were suggested to allow the cell to fine-tune the stress response according to slightly different signals, and also open the possibility of a quicker and more precise cellular response controlled at the post-translational level (Gasch *et al.*, 2000). Gasch and colleagues have also shown that the observed global changes in gene expression are mainly transient and that cells adapt over time to the stress imposed, reaching new steady-state transcript levels. These authors propose a model in which the levels of mRNA and protein behave differently under stress. For example, while transcript induction is transient, protein levels instead increase in a stepwise manner until reaching the appropriate new level, as observed for the molecular chaperones in response to heat shock (Gasch *et al.*, 2000). Nevertheless, some authors demonstrated that, at least for the *TRR1* gene, the mRNA and protein induction factors are equal and transient (Godon *et al.*, 1998).

A different behaviour between transcripts and proteins can account for the differences observed in the expression levels of some mRNAs and corresponding proteins (Figure 27 and Table 18), suggesting that genetic code ambiguity regulates gene expression at the translational level, as described for other stress responses (Hinnebusch, 1996; Patil and Walter, 2001).

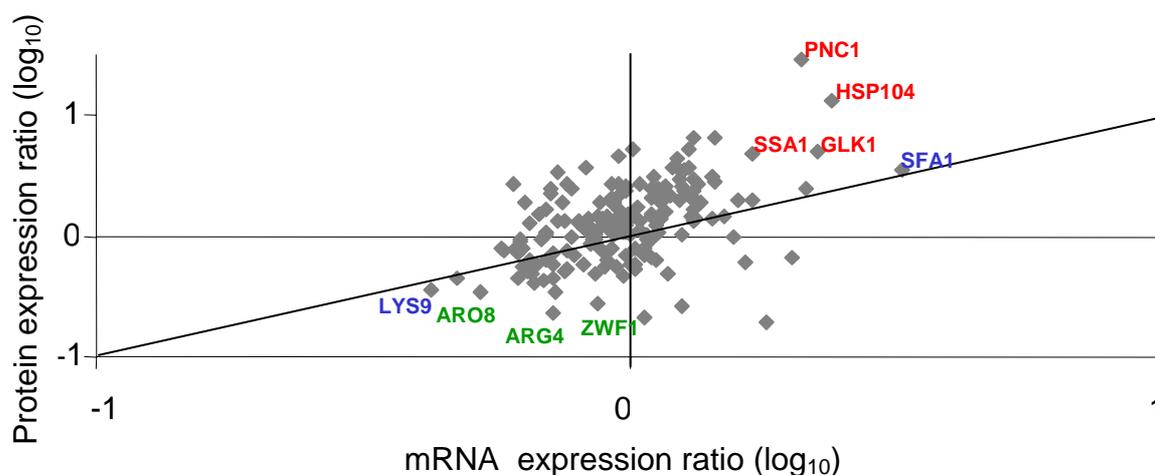
The selected genes shown in Table 18 are some examples of gene expression regulation at the translational level in cells expressing CUG ambiguity.

**Table 18 – Differences observed in gene and protein expression levels in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, as measured by DNA-microarrays and 2D-PAGE approaches. Genes shown in green are repressed and in red are induced.**

Function	Gene	Fold variation	
		mRNA	Protein
Metabolism	LYS9	2.4	2.8
	ARO8	1.9	5.1
	ARG4	1.4	4.4
	ZWF1	1.2	3.7
	SFA1	3.2	3.5
Stress	PNC1	2.1	29.5
	HSP104	2.4	13.1
	GLK1	2.3	5.0
	SSA1	1.7	4.7
	STI1	1.2	2.9
Proteasome	RPN12	1.3	5.2
	RPN10	1.3	3.6
	PUP2	1.3	3.6
	SCL1	1.2	3.6

*LYS9* and *SFA1* were included in Table 18 as genes with the same behaviour at the mRNA and protein levels. Some genes are induced both at the mRNA and protein levels, however for most of them the increase in expression is only detected for the corresponding protein. Among the genes listed, *PNC1* and *HSP104* are stress-inducible genes and have the biggest differences between mRNA and protein expression levels (Figure 27), being good candidates for translational regulation. Besides the genes that are stress-regulated, proteasome subunits can also be targets for translational control, as their expression levels are

unaltered at the mRNA level but the proteins are induced in response to CUG mistranslation, and additionally they might be subject of coordinated expression. The mechanism or mechanisms of translational control that might operate in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> are not yet known.



**Figure 27 - Correlation between gene and protein expression levels from *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>.** The genes prone to translational control are likely to be the ones that show higher differences between mRNA and protein levels, as shown by the deviation from the middle line. In blue, genes that have the same fold variation at the mRNA and protein level. In green and red, genes that have differences in mRNA and protein ratios. *PNC1* and *HSP104* are the genes that display less correlation between mRNA and protein levels, thus their expression is likely to be regulated at the translational level.

Another important conclusion arising from the data is that the impact of CUG ambiguity on the proteome is not directly related to the presence of CUG codons on the targets (Table 19). Again, the proteasome is a good example of that as some of its subunits show the same protein expression induction (3.6-fold) but the respective mRNAs contain 7 (*RPN10*), 2 (*SCL1*) or even no (*PUP2*) CUG codons. A comparison between *PNC1* and *HSP104* also shows that more CUG codons do not imply bigger fold increase, and the same applies to the repressed proteins. As an example, *ARO8* has a bigger fold decrease than *ZWF1*, however the number of CUG codons is less. The results indicate that the proteome, and not only the subset of proteins corresponding to the genes containing CUG codons, is globally affected by ambiguity.

**Table 19 - Selected genes altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, showing the number of CUGs encoded in the mRNA and respective protein fold variation.** In green are represented proteins that are repressed and in red proteins that are induced. d stands for a protein that is not represented on the gel from the T<sub>33</sub> strain when compared to the control condition.

Gene	N. ° CUG	Fold variation
<i>ARO8</i>	4	5.1
<i>ZWF1</i>	13	3.7
<i>YDJ1</i>	0	d
<i>PNC1</i>	2	29.5
<i>HSP104</i>	7	13.1
<i>RPN10</i>	7	3.6
<i>PUP2</i>	0	3.6
<i>SCL1</i>	2	3.6
<i>SFA1</i>	4	3.5

CUG mistranslation imposes a permanent and internal stress to the cell, rather than a transient and external change on the environmental conditions. This justifies the lower number of alterations observed, both in protein number and in magnitude of expression, when compared with other stress conditions (Godon *et al.*, 1998; Vido *et al.*, 2001). Therefore, these are the proteins that are relevant to the adaptation of the cell to genetic code ambiguity, since the general stress response has been turned off and the new steady-state was achieved due to the long term exposure of the cells to the stress imposed by CUG ambiguity.

In general, proteins whose expression is decreased by genetic code ambiguity belong to metabolic pathways such as carbohydrate metabolism, amino acid metabolism and protein synthesis. Inversely, proteins involved in the stress response and protein degradation are induced. The data clearly shows that increased expression of the molecular chaperones that assist folding of unfolded or misfolded proteins and resolve protein aggregates, and increased expression of proteasome subunits (the major protein degradation machinery), allows ambiguous cells to get rid of non-functional proteins and survive with a globally disrupted proteome.

### **2.3. Proteome alterations under additional stress conditions**

The previous chapters indicate that the *C. albicans* ser-tRNA<sub>CAG</sub> imposes a significant stress on *S. cerevisiae*, causing large transcriptome and proteome alterations. The results also suggest that the nature of this stress is different from other conditions previously tested, most likely due to the fact that it consists of a constitutive stress. Because genetic code ambiguity is a new type of stress that has never been studied before, the response of *S. cerevisiae* cells to the *C. albicans* ser-tRNA<sub>CAG</sub> needs to be fully and globally characterized.

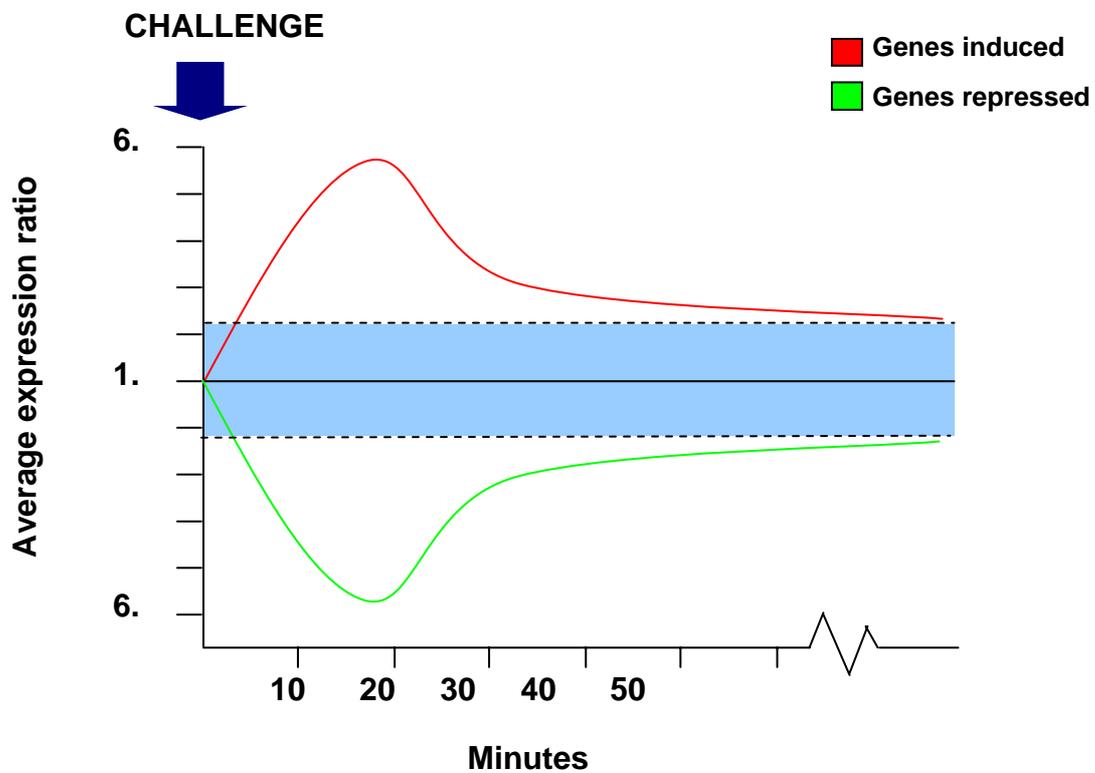
Many stress conditions result in pleiotropic effects, and it has been proposed that the cells detect simultaneously but independently different signals and respond to each environmental challenge. The gene expression programme generated then consists of the combination of the individual conditions (Gasch *et al.*, 2000; Gasch and Werner-Washburne, 2002). Moreover, it has been described that cells previously exposed to stress exhibit tolerance to a different type of stress (Hohmann and Mager, 1997). For example, treatment of cells with sodium arsenite pre-adapts them for subsequent heat shock challenge (Mizzen and Welch, 1988). Such cross-protection mechanism results from the induction of stress-protective proteins, namely heat shock proteins (Duncan, 1996). In response to the first stress condition, the expression levels of such proteins are increased, which enables cell survival under further environmental changes, without requirement for re-induction of Hsps.

The possibility that a cross-protection mechanism exists in *S. cerevisiae* cells expressing CUG ambiguity raises important questions concerning their response to additional stress conditions. For example, are these stressed cells capable of reacting to additional environmental challenges? Is their capacity to adapt to changes in the environment compromised by the presence of the *C. albicans* ser-tRNA<sub>CAG</sub>, or instead does CUG ambiguity protect them from further stresses? In order to answer these questions, proteome analysis of *S. cerevisiae* cells

---

expressing the *C. albicans* ser-tRNA<sub>CAG</sub> under additional stress conditions was performed.

It has been demonstrated that the transcriptional responses to stress are transient (Gasch *et al.*, 2000). Following an environmental challenge, most gene expression changes occur within the first 30 minutes (Figure 28), and after 1 hour the gene expression levels reach new steady states, as cells adapt to the stress condition, in which the values are closer to the basal expression levels. However, most stress experiments characterize the response of cells after a short period of time, and few studied the long-term consequences of exposure to stress conditions. Therefore, the transient and the permanent responses to stress ought to be differentiated. For this, two distinct stress experiments were performed in cells expressing genetic code ambiguity. The temperature upshift, from 25 to 37°C for 30 minutes, was chosen as a model of a transient stress. Indeed, the heat shock is a standard stress used in many studies with a well-characterized response in several organisms including yeast (Boy-Marcotte *et al.*, 1999; Gasch *et al.*, 2000; Causton *et al.*, 2001). With the aim of mimicking a permanent stress condition, growth at the sub-optimal temperature of 37°C was performed.



**Figure 28 - Generic transient changes in gene expression following an environmental challenge.** When exposed to stress, cells respond by altering gene expression. These changes occur within the first 30 minutes, and after 1 hour the gene expression levels return to values close to the basal expression levels. This new steady state reflects cellular adaptation to the stress condition (adapted from Gasch *et al.*, 2000).

### 2.3.1. Proteome alterations under growth at 37°C

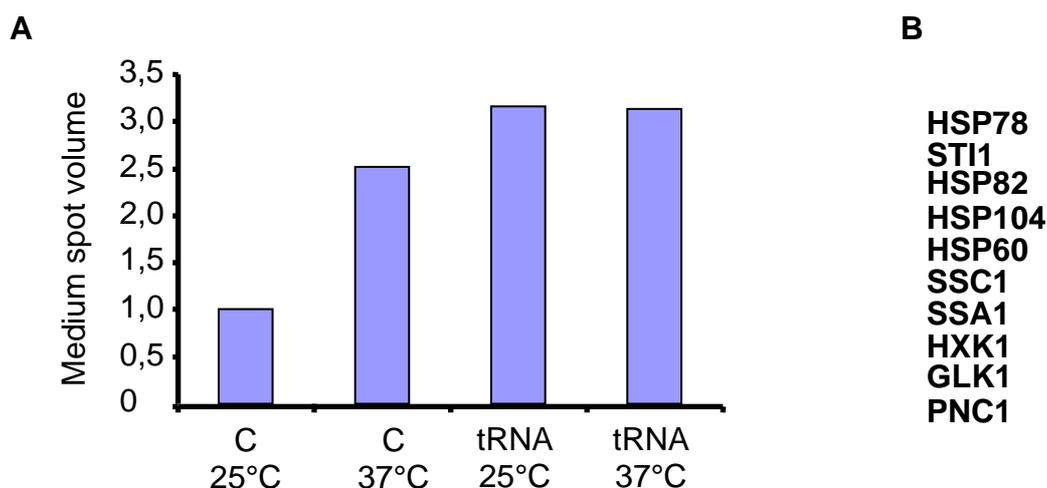
Proteome characterization of *S. cerevisiae* control cells grown at 37°C showed that 55 protein spots have altered expression levels (Annex, Map 1) when compared to the proteome of the same cells under normal growth temperature. Of these, 34 spots showed increased expression, which correspond to 20 identified proteins and 10 unidentified spots, plus 4 new spots that are not expressed in the control cells grown at 25°C. From these new spots, there are 2 that correspond to the previously identified proteins Hxk1p and Ugp1p (Annex, Table II). Among the proteins induced by this long-term stress imposed on *S. cerevisiae* control cells are Pnc1p, Hsp104p, Glk1p, Sti1p and Hsp78p. As discussed in the previous chapter for the tRNA T<sub>33</sub>-transformed strain, and coherent with what was expected, an important part of the cellular response to stress is involved in maintaining viability and providing energy for the functioning of the molecular chaperones that help in protein folding. Considering the 21 spots with decreased expression in control cells grown at 37°C, they correspond to 12 identified proteins, mainly amino acid metabolism enzymes, and 9 unidentified spots. Although expressed in the control cells grown at 25°C, one of these repressed proteins (Egd2p) is not expressed at 37°C and was, therefore, considered as disappeared from the 2D-gel map (Annex, Table II).

Proteome characterization of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> grown at 37°C showed 29 proteins with altered expression levels (Annex, Map 2), when compared to the same cells under normal growth temperature. 8 of these spots have increased expression levels in cells expressing CUG ambiguity, namely Adh1p, Oye2p, Met3p and 4 unidentified spots that are up-regulated, plus 1 new spot (Ugp1p) that is not expressed in cells grown at 25°C (Annex, Table III). Considering the 21 proteins with decreased expression levels in cells expressing genetic code ambiguity grown at 37°C, they correspond to 7 unidentified spots and 13 identified proteins, among which are Snz1p, Lys20p and Ssa2p (Annex, Table III). Consistent with the hypothesis that cells expressing genetic code ambiguity are permanently stressed, there are few proteome

alterations when these cells are grown at the sub-optimal temperature of 37°C, and the stress-responsive proteins are not induced, as seen for the control cells. This can be related to a cross-resistance mechanism already described (Hohmann and Mager, 1997), in which cells exposed to a mild stress become resistant to a second unrelated stress, even if the additional stress is a higher dose than what the cell could normally cope with. Indeed, such protection reflects the previous induction of stress-protective proteins, required for the response to both environmental changes. Therefore, increased expression levels of the heat shock proteins, induced by the first stress condition, allow the cell to tolerate further challenges and adapt to unfavourable conditions.

This result is highlighted in Figure 29, which shows the protein expression trend upon imposition of a permanent stress on *S. cerevisiae* cells, either by the presence of the *C. albicans* ser-tRNA<sub>CAG-T33</sub> or growth at the sub-optimal temperature of 37°C. The medium expression levels of the heat shock proteins and other stress-responsive proteins were calculated for the *S. cerevisiae* control strain and for the strain expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, both at 25°C and 37°C. In order to deduce general folds, the value obtained for the control was set to 1 and the others were then normalized to the control. Control cells grown under permanent stress show approximately a 2.5-fold induction of the stress proteins, whereas the *S. cerevisiae* strain expressing the *C. albicans* ser-tRNA<sub>CAG-T33</sub> shows about 3-fold up-regulation of the same proteins when grown either at 25°C or 37°C (Figure 29). Indeed, the stress proteins whose expression is increased in the control strain at 37°C are already expressed at higher levels in CUG ambiguous cells even when grown at 25°C, suggesting that cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> can grow at 37°C without further induction of those heat protective proteins (Figure 29). In other words, the presence of the ser-tRNA<sub>CAG</sub>, which induces CUG ambiguity, pre-adapts cells to long-term survival under stressful conditions, thus explaining why the *S. cerevisiae* strain expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, when grown at 37°C, shows little global proteome alteration (Annex, Table III). As described above, this is a cross-protection mechanism, where the expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* cells function

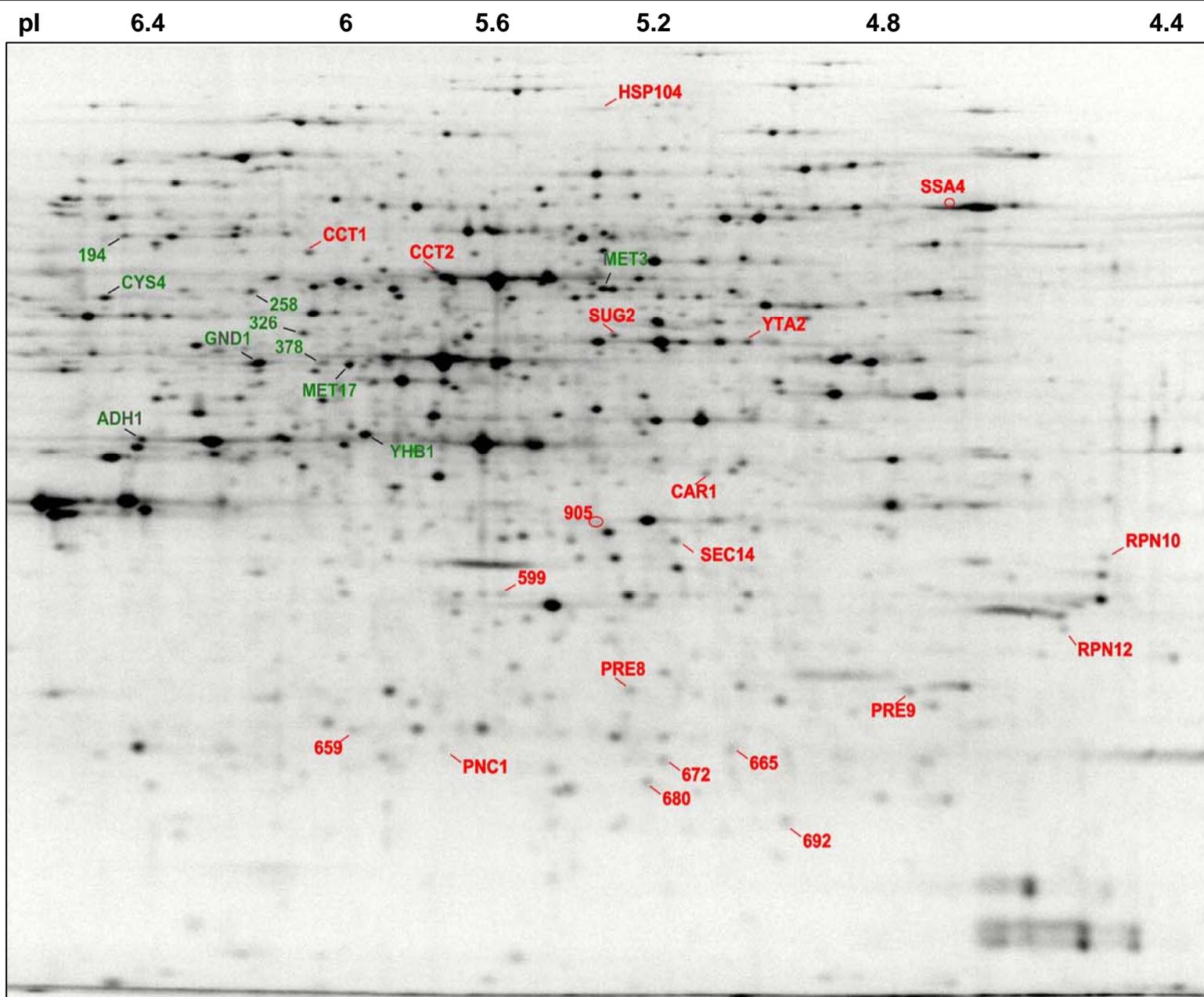
as a priming stress. By constitutively inducing a stress response with increased expression levels of heat shock proteins, CUG ambiguity would enable the cell to tolerate additional distinct stress conditions, as previously observed (Santos *et al.*, 1999).



**Figure 29 - CUG ambiguity creates a pre-adaptive condition that allows for tolerance to adverse environmental challenges.** A - The medium expression level of the stress proteins listed on B was calculated for each strain and conditions shown. The value obtained for the control was set to 1 and the others were then normalized to the control, in order to deduce general folds. From left to right, *S. cerevisiae* control cells grown at 25°C and 37°C, and *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> grown at 25°C and 37°C are shown. B – List of the stress-induced proteins whose expression levels were used to calculate the values shown on A.

Other important conclusion arises from the direct comparison of the proteome of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> grown at 37°C with the control cells grown in the same conditions. There are 20 spots that are increased in cells expressing CUG ambiguity at 37°C relative to the control cells at the same temperature (Figure 30). These spots correspond to 12 identified proteins, mainly stress proteins and proteasome subunits, and 6 unidentified spots, plus 2 new spots that are not expressed in the control cells grown at 25°C, corresponding to Ssa4p and 1 unidentified protein (Annex, Table IV). From the 10 proteins with decreased expression levels in cells expressing genetic code ambiguity grown at 37°C when compared to control cells grown at the same temperature, 6 correspond to identified proteins, namely Yhb1p, Met17p, Adh1p, Met3p, Gnd1p and Cys4p, and 4 unidentified spots (Annex, Table IV). Expression of proteasome subunits is increased. Since this was not detected by the previous analysis that compared *S. cerevisiae* control cells or cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> grown at 37°C (stress condition) with the corresponding strain grown at 25°C (regular growth temperature), the observation strengthens the idea that proteasome induction is responding to the stress imposed by CUG mistranslation rather than to a general stress mechanism. This issue will be reprised in the next chapter (Figure 33).

**Figure 30 – *S. cerevisiae* 2D-map of proteins from cells grown at 37°C.** Proteins were labelled with [<sup>35</sup>S]-methionine in cultures grown to mid-log phase (OD<sub>600nm</sub> = 0.5). The map shows in black the spots corresponding to the control cells and indicates in colour the proteins whose expression level is altered by at least 2 fold in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>: in red, proteins that are induced; and in green, proteins that are repressed in cells ambiguously decoding the CUG codon.



## **2.3.2. Proteome alterations under Heat-Shock**

### **2.3.2.1. Characterization of the response of control cells to heat shock**

Proteome characterization of *S. cerevisiae* control cells exposed to heat shock (temperature upshift from 25 to 37°C for 30 minutes) showed 162 protein spots whose expression levels were altered (Annex, Maps 3, 4 and 5). 117 spots are induced in heat-shocked control cells when compared to the same cells grown at 25°C (Annex, Map 3). These spots correspond to 39 identified proteins and 42 unidentified spots that are up-regulated (note that some proteins are represented on the 2D map by more than 1 spot), plus 31 new spots (Annex, Map 4) that are not expressed in the control cells grown at 25°C, from which there are 11 identified proteins (Annex, Table V). 45 spots showed decreased expression levels in control cells exposed to heat shock (Annex, Map 5), which correspond to 27 identified proteins and 10 unidentified spots. Although expressed in the control cells grown at 25°C, 4 of these repressed proteins are not expressed in heat-shocked control cells and were, therefore, considered as disappeared from the 2D-gel map (Annex, Table V).

As expected, heat shock proteins are among the most induced proteins, along with carbohydrate metabolism proteins, including enzymes related to trehalose synthesis. Under stress, energy requirements increase due to ATP-dependent folding (Gasch *et al.*, 2000) and trehalose may protect cellular components from denaturation (Causton *et al.*, 2001). As observed in the response to oxidative stress, heat shock up-regulates proteins with antioxidant functions, as well as proteins involved in glycerol metabolism (Godon *et al.*, 1998). Among the functions repressed by heat shock are sulfur, amino acid and purine metabolism enzymes, along with protein synthesis, which is in agreement with previous proteome alterations described in heat-shocked yeast cells (Boy-Marcotte *et al.*, 1999).

### 2.3.2.2. Characterization of the response of CUG ambiguous cells to heat shock

Proteome characterization of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> exposed to heat shock showed that 142 protein spots have altered expression levels in cells expressing CUG ambiguity (Annex, Maps 6, 7 and 8). 81 of those spots have increased expression levels (Annex, Map 6) and correspond to 21 identified proteins and 33 unidentified spots that are up-regulated, plus 18 new spots (Annex, Map 7) that are not expressed in the same cells grown at 25°C. From these new spots, 11 correspond to identified proteins (Annex, Table VI). Considering the spots with decreased expression levels in heat-shocked cells expressing genetic code ambiguity (Annex, Map 8), there are 61 spots that correspond to 42 identified proteins and 12 unidentified spots that are down-regulated. Although expressed in the CUG ambiguous cells grown at 25°C, 6 of these repressed proteins are not expressed in heat-shocked cells and were, therefore, considered as disappeared from the 2D-gel map (Annex, Table VI).

The functional groups of proteins altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> exposed to heat shock are mainly those observed for the control cells. There is an induction of heat shock proteins and carbohydrate metabolism proteins and the repression of sulfur, amino acid and purine metabolism enzymes, and protein synthesis.

A summary of the response of *S. cerevisiae* control cells and cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> to heat shock (transient stress) or growth at 37°C (permanent stress) is shown in Figure 31 and Table 20, where selected proteins whose expression is altered for each cell type or condition are represented. These results show that the proteins whose expression is altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> are mainly those responsive to heat shock, with the exception of proteasome subunits. Indeed, stress proteins, molecular chaperones and carbohydrate metabolism enzymes are induced in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>, when grown at 25°C, and are also induced in response to heat shock, both in control cells and CUG

ambiguous cells. However, very few of these proteins are induced when *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> are grown at 37°C, reflecting the increased capability of these cells to adapt to long-term stress exposure (Figure 29, previous chapter). On the other hand, these results also highlight that the permanent stress imposed by genetic code ambiguity does not impair the cells to mount a transient response to an additional, distinct type of stress, and can further increase the expression of stress-protective proteins (see also Figure 32 below). Similarly, the repression of protein synthesis is observed in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>, when grown at 25°C, and also in response to heat shock, both in control and CUG ambiguous cells. Nevertheless, this decreased expression of proteins involved in translation is not detected when cells were grown under a permanent stress condition, such as the sub-optimal temperature of 37°C, again reflecting the adaptability of cells to unfavourable environments. A distinct result was obtained for the amino acid and purine biosynthesis. The amino acid metabolism enzymes are repressed by ambiguity, as shown by the down-regulation of Arg1p and Arg4p in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>, when grown at 25°C or after heat shock, which was not observed for the control strain. On the other hand, proteins related with the purine metabolism, namely Ade13p and Gua1p, show decreased expression levels in response to heat shock only, both in control and ambiguous cell lines.

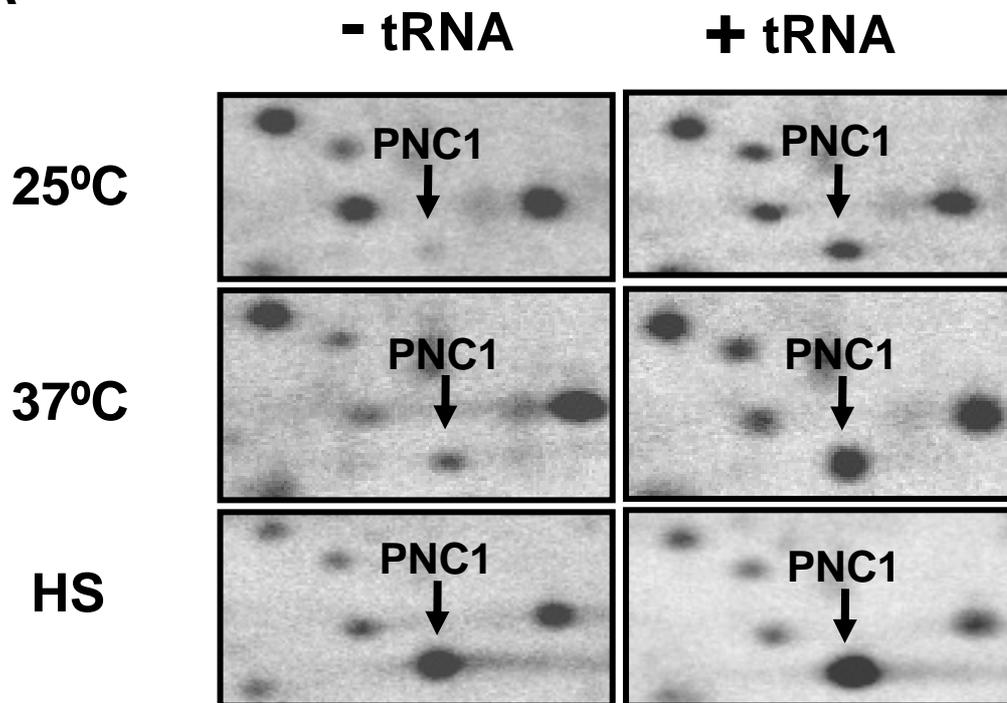
**Table 20 – Proteins whose expression is altered by genetic code ambiguity.**

The first two columns show the expression changes of ambiguous cells grown at 25°C and 37°C, compared to control cells under the same conditions. The last two columns indicate the alterations in response to heat shock, both in control or ambiguous cells compared to the growth at normal temperature (25°C). Proteins induced are shown in red, and proteins repressed are indicated in green. Proteins that are not expressed in the control condition are considered new (n) and proteins that are not expressed in the stress conditions are considered disappeared (d) (note that their real fold variation, therefore, might not be accurately determined).

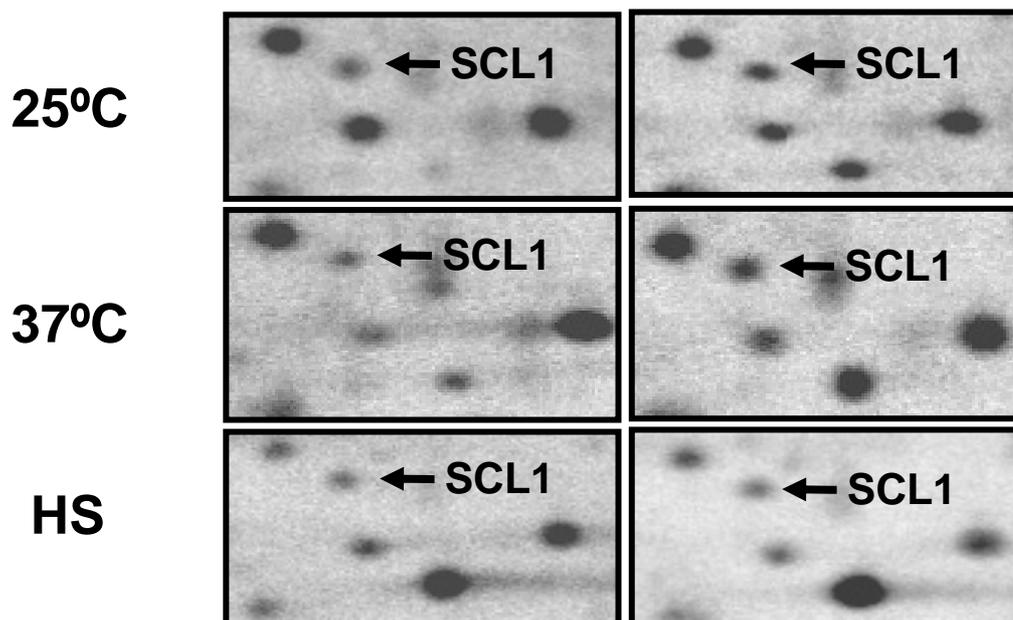
	tRNA 25°C	tRNA 37°C	control HS	tRNA HS
<b>Stress proteins and chaperones</b>				
Pnc1	29.5	2.2	135.4	8.1
Hsp104	13.1	2.6	112.3	10.1
Ssa1	4.7	-	5.1	1.9
Ssa4	n	2.3	18.8	19.4
Ssa3	-	-	2.6	6.3
<b>Carbohydrate metabolism</b>				
Hxk1	n	-	9.4	31.6
Glk1	5.0	-	45.1	11.5
Hor2	n	-	10.9	8.7
<b>Proteasome subunits</b>				
Rpn12	5.2	2.5	-	-
Rpn10	3.6	2.2	-	-
Pup2	3.6	-	-	-
Scl1	3.6	-	-	-
Pre8	-	3.5	-	-
Rpt3	-	2.1	-	-
Pre9	-	2.0	-	-
Rpt4	-	1.9	-	-
<b>Amino acid and purine metabolism</b>				
Met17	5.3	2.4	-	-
Arg1	3.4	-	-	4.1
Arg4	4.4	-	-	3.3
Ade13	-	-	10.4	5.7
Gua1	-	-	4.9	3.1
<b>Protein synthesis</b>				
Ssb1	2.1	-	8.6	5.0
Ssb2	2.1	-	5.5	2.8
Krs1	2.7	-	-	2.0
Tif1	-	-	5.6	2.9

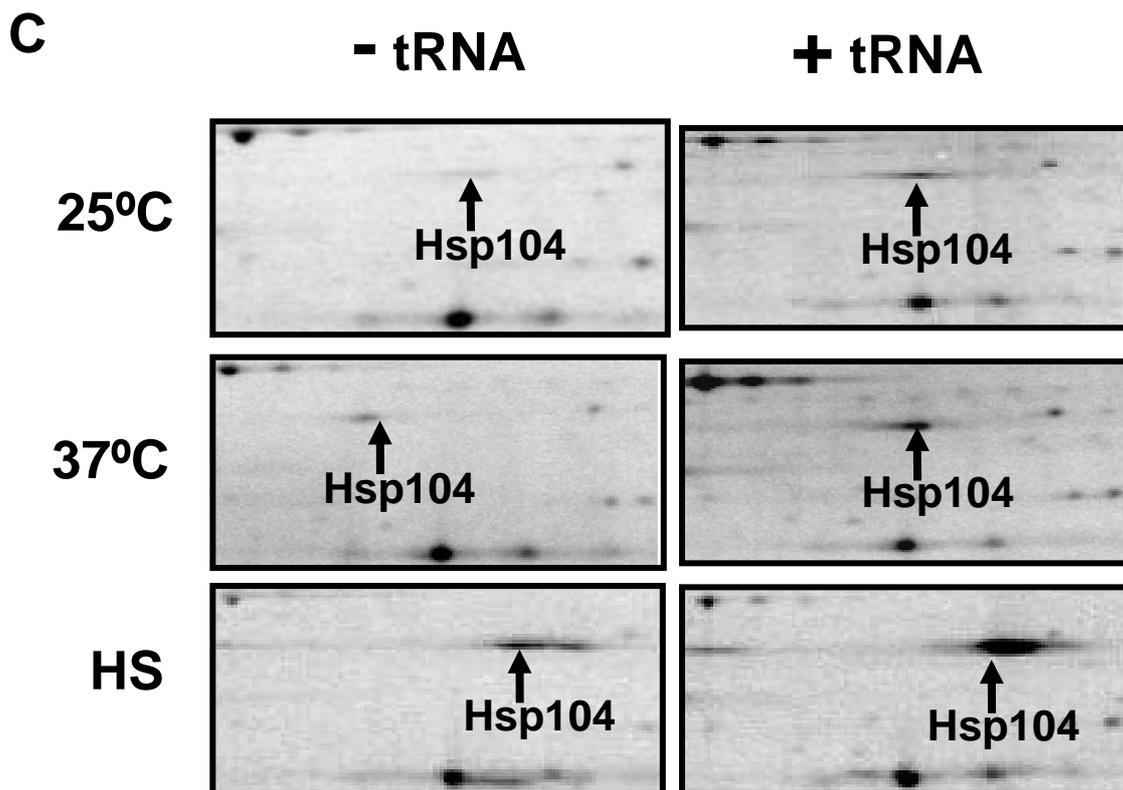
The proteasome is the single functional category that is responsive to CUG ambiguity both under normal and permanent stress conditions, but not after heat shock. Proteasome subunits are induced in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-T<sub>33</sub> when grown at 25°C and 37°C, however such induction was not detected for control cells (see also Figure 33 below).

**A**



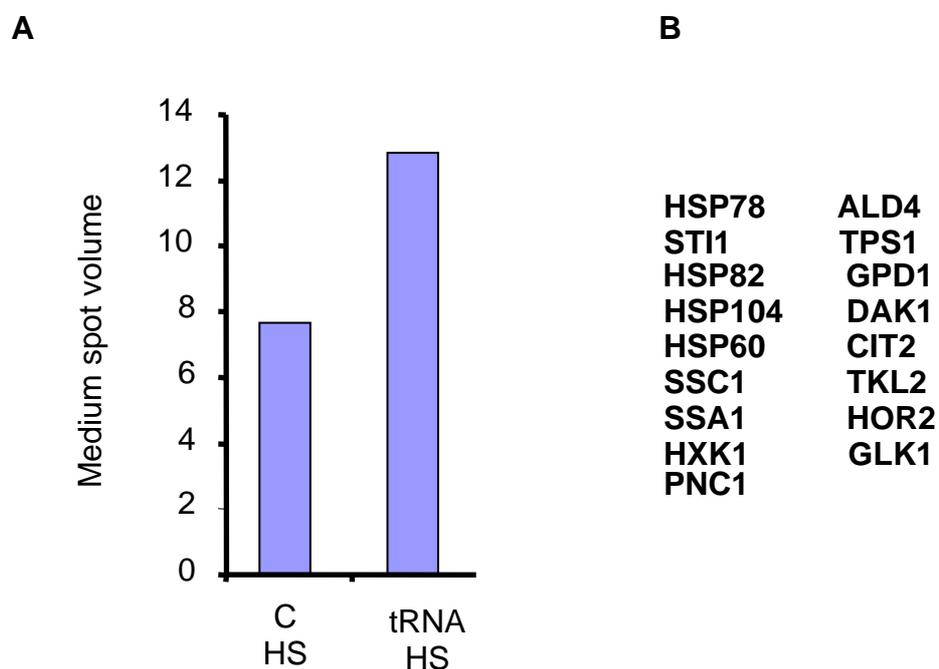
**B**





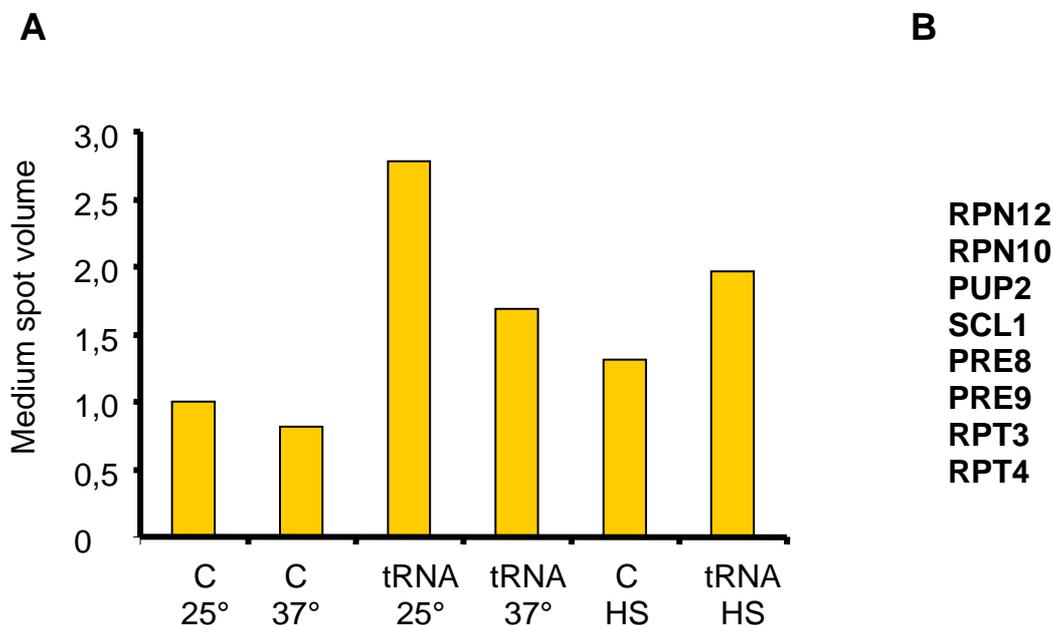
**Figure 31 – Details of the 2D-Map showing the proteins whose expression is most induced by genetic code ambiguity.** The portions of gels shown correspond to *S. cerevisiae* control cells or *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> (columns), grown at 25°C, 37°C, or under heat shock (lines). A – The stress-responsive protein Pnc1 is induced by all stress conditions and in both strains, however ambiguity and heat shock result in higher fold-increase. B – The proteasome subunits are induced only in ambiguous cells and only in response to permanent stress conditions, therefore, their expression is not increased in control cells neither after heat shock. C – The molecular chaperone Hsp104 is also induced by all stress conditions and in both strains, but mainly in cells expressing CUG ambiguity.

From the comparison of the responses of control and ambiguous cells to heat shock, it appears that the *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> induce stress proteins to a lower level, as the proteins induced in the latter strain submitted to heat shock are less, both in number and in fold (Annex, Tables V and VI). However, if we consider that these stress proteins are already expressed at higher levels in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>, when compared to the control cells (Figure 29, previous chapter), the response to heat shock might be similar between both strains. The medium expression levels of the heat shock proteins and other stress – responsive proteins were calculated for the control *S. cerevisiae* strain and for the strain expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> after heat shock. The value obtained for the control (*S. cerevisiae* cells grown at 25°C, Figure 29) was set to 1 and the others were then normalized to the control, in order to deduce general folds. After temperature upshift from 25°C to 37°C for 30 minutes, control cells show approximately a 8-fold induction of the stress proteins, whereas the *S. cerevisiae* strain expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> shows about 12-fold up-regulation of the same proteins when grown under the same stress condition (Figure 32). Note that these folds are relative to the control situation, that is *S. cerevisiae* control cells grown at 25°C. If the level of the stress proteins already expressed in the *S. cerevisiae* strain expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> at 25°C is taken into account, then the induction of stress proteins is only about 4-fold, which is half of the value observed for the control strain, strengthening the evidence from the previous chapter that these cells are pre-adapted to stress conditions. These results also indicate that *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> are able to further induce stress proteins when submitted to additional stress, suggesting that the stress imposed by mRNA decoding ambiguity does not compromise adaptation capacity.



**Figure 32 – Cells that ambiguously decode the CUG codon retain capacity to respond to additional stress.** A - The medium expression level of the stress proteins listed on B was calculated for each strain. The values shown are normalized to the control strain, grown at 25°C, in order to deduce general folds (Figure 26). Columns indicate *S. cerevisiae* control cells after temperature upshift from 25°C to 37°C for 30 minutes, and *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> after the same stress treatment. B – List of the stress-induced proteins whose expression levels were used to calculate the values shown on A.

Curiously, proteasome subunits were not induced in response to heat shock, although other proteases like Hsp78p and Prb1p show higher expression levels both in *S. cerevisiae* control cells and cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> after temperature upshift. Similar observations were made for these strains grown at 37°C, as seen in Table 20. The induction of the proteasome by CUG ambiguity but not in response to thermic stress, either with a transient temperature upshift or permanent growth at the sub-optimal temperature of 37°C, is highlighted in Figure 33. The medium expression level of the proteasome subunits was calculated for the control *S. cerevisiae* strain and for the strain expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> grown at 25°C, 37°C and after heat shock. As before, the value obtained for the control was set to one and the others were then normalized to the control, in order to deduce general folds. Growth at the sub-optimal temperature of 37°C (permanent stress) and temperature upshift from 25°C to 37°C for 30 minutes (transient stress) were unable to induce the proteasome both in control cells and on the *S. cerevisiae* strain expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> (Figure 33). Indeed, the proteasome subunits only show increased expression levels in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> grown at 25°C when compared to the control. Proteasome up-regulation also appears significant when the comparison of growth at 37°C is made between strains, therefore, supporting the hypothesis that the effect of the stress imposed by the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> on the proteasome is more specific than a general stress response. Also in favour of this idea is the fact that the stress response to the misfolding induced by AZC increases expression of the genes that encode proteasome subunits, whereas temperature upshift does not (Trotter *et al.*, 2002).



**Figure 33 – The proteasome is induced by CUG ambiguity but not by heat stress.** A - The medium protein expression levels of the proteasome subunits listed on B was calculated for the control *S. cerevisiae* strain and for the strain expressing the *C. albicans* ser-tRNA<sub>CAG-T33</sub>, grown at 25°C, 37°C and after heat shock. The value obtained for the *S. cerevisiae* control strain grown at 25°C was set to 1 and the others were then normalized to the control in order to deduce general folds. The results showed that the proteasome is induced in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, but not by heat shock or growth at 37°C. B – List of the proteasome subunits whose expression levels were used to calculate the values shown on A.

## **2.4. Proteasome subunit identification on 2D-maps**

The proteome maps of *S. cerevisiae* cells expressing the *C. albicans* Ser-tRNA<sub>CAG</sub>, grown at 25°C and 37°C (chapters III.2.2. and III.2.3.1.), have highlighted the proteasome subunits as an important group of proteins whose expression is increased by genetic code ambiguity. This result appears to be more relevant to these cells as the proteasome is not induced by heat shock (chapter III.2.3.2.). Other proteomic studies performed with yeast cells have found that some proteasome subunits have increased expression levels in response to cadmium (Vido *et al.*, 2001) or oxidative stress (Godon *et al.*, 1998), and an increase in the expression of proteasome genes was also detected in cells treated with the proline analog AZC that induces protein misfolding (Trotter *et al.*, 2002). However, microarray analysis failed to detect any induction of proteasome subunits in *S. cerevisiae* cells expressing the *C. albicans* Ser-tRNA<sub>CAG</sub> (chapter III.1.). These observations prompted the questions of “is the proteasome globally induced in response to CUG mistranslation?” or, “are there only some subunits responsive to genetic code ambiguity?” And if so, “does this putative partial induction have some regulatory effect?”

Regulation of cellular processes requires the control of both activation and inactivation pathways (Pickart and Cohen, 2004). Therefore, protein degradation is a very important negative regulatory mechanism, as it assures that the targeted proteins terminate their functions. To provide complete and selective protein degradation, organisms have evolved specialized proteases that recognize signals on the targeted substrates (Pickart and Cohen, 2004). These multi-subunit proteolytic enzymes include chaperones that unfold the substrates and translocate them to the active sites in an ATP-dependent manner. In eukaryotes, these machines are the 26S proteasomes that selectively degrade both aberrant and normal short-lived or regulatory proteins. The proteasome is, therefore, involved in many diverse cellular processes, such as regulation of cell cycle progression, signal transduction or antigen processing, besides protein quality control (for reviews see Kloetzel, 2001; Kostova and Wolf, 2003; Pickart and Cohen, 2004),

and more recently it has been implicated in transcription (reviewed by Muratani and Tansey, 2003).

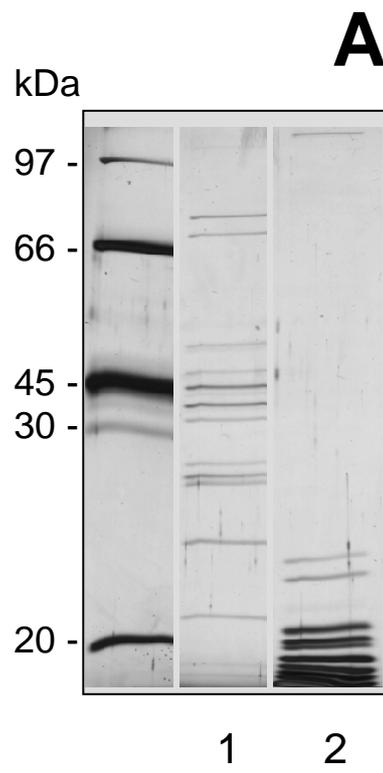
Proteins targeted for degradation by the 26S proteasome are covalently modified by addition of a 76-amino acid protein called ubiquitin (Thrower *et al.*, 2000). Ubiquitin is conjugated to lysine residues of the target substrates in a multistep pathway involving at least three types of enzymes, and is, therefore, a complex post-translational modification (Weissman, 2001). The ubiquitylation mechanism requires an ATP-dependent activation of ubiquitin by an enzyme named E1. After, an ubiquitin-conjugating enzyme (UBC or E2) accepts the activated ubiquitin from E1 and thirdly, an ubiquitin protein ligase (E3) transfers the ubiquitin from the E2 enzyme to the protein substrate. There are several E1, E2 and E3 enzymes, responsible for the specificity of the ubiquitylation process (Weissman, 2001). Conjugation of a polyubiquitin chain serves as signal for substrate recognition by the proteasome, whereas a single ubiquitin tag regulates structure, location and activity of the target proteins (reviewed in Hicke, 2001; Weissman, 2001). For example, histone monoubiquitylation is required for meiosis in yeast cells, and monoubiquitylation of membrane proteins is the signal required for their internalization into the endocytic pathway (Schnell and Hicke, 2003).

The *S. cerevisiae* 26S proteasome is a complex of about 36 different subunits assembled in a barrel-like structure (Baumeister *et al.*, 1998). The 26S proteasome results from the assembly of the 20S proteolytic complex and two 19S regulatory particles. The 20S proteasome is composed by 28 subunits arranged as four seven-membered rings, two central  $\beta$ -rings and two distal  $\alpha$ -rings. In the 20S proteolytic complex, only 3  $\beta$ -subunits contain the catalytic sites (*PUP1*, *PRE2* and *PRE3*), whereas the  $\alpha$ -subunits control the intake of substrates and the exit of degradation products of the proteasome. The 19S regulatory cap comprises the base that contains six ATPases (Rpt1 to Rpt6) plus two non-ATPase subunits (Rpn1 and Rpn2), and a lid of at least 9 distinct subunits (Rpn3 and Rpn5 to Rpn12) that includes the subunit responsible for the de-ubiquitylation (Rpn11) and

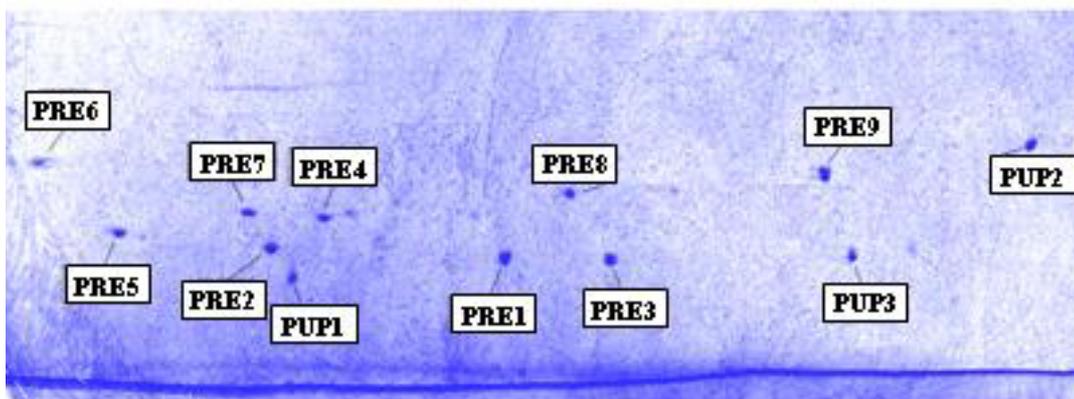
that is, therefore, fundamental in ubiquitin-dependent proteolysis (Pickart and Cohen, 2004).

The fact that the 2D-maps that were used for the proteome analysis only have the localization and identification of about one third of the proteasome subunits, and the possibility that genetic code ambiguity induces protein degradation by increasing the expression of the proteasome, make it crucial to have the identification of all proteasome subunits on the 2D-map. This would allow the comparison of their localization on the gel with the maps previously obtained for *S. cerevisiae* cells expressing CUG ambiguity, under normal and stressful growth conditions. The detection of more proteasome subunits with increased expression levels in ambiguous cells would provide important data to elucidate the question of which proteasome subunits are responsive to genetic code ambiguity, or if the proteasome is globally induced.

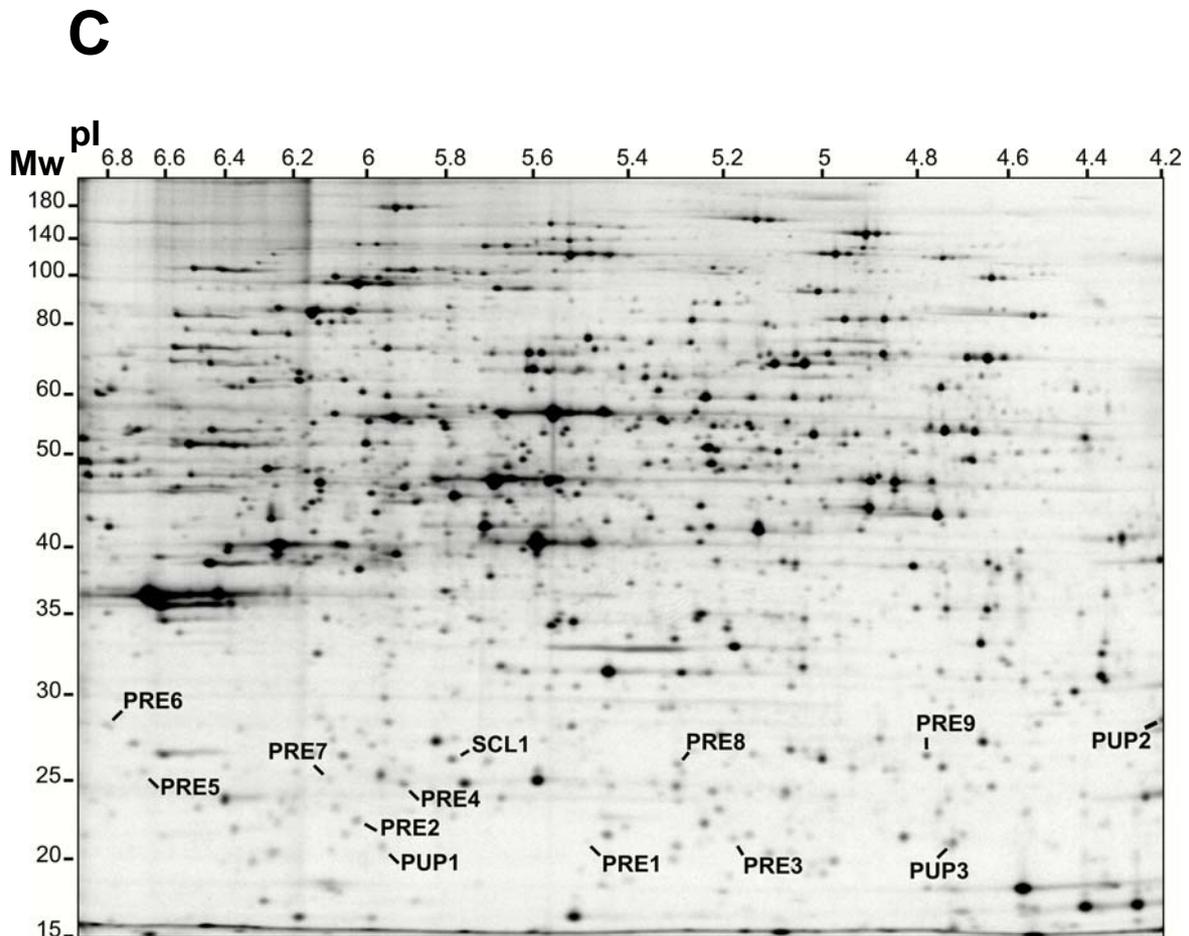
In order to localize and identify the proteasome subunits on the proteome map, proteasomes were purified from *S. cerevisiae* SCO973 (*SCL1*-tagged) and SCO832 (*RPN5*-tagged) cells using the TAP-tag approach (Rigaut *et al.*, 1999; Puig *et al.*, 2001), allowing the recovery of the 20S and 19S sub-complexes, respectively. This method consists of two consecutive affinity purifications using different tags, allowing the recovery of protein complexes with less contaminants. Purified proteasome subunits were then co-migrated on a 2D gel with <sup>35</sup>S-methionine labelled proteins extracted from CEN-PK2 control cells. Their exact location on the 2D map was achieved by superimposing the radioactive image of total extracted proteins with the coomassie-stained gel with the separated TAP-purified proteasome subunits. The identification of the subunits was achieved by mass spectrometry and this map was then compared with previously obtained gels.



**B**



**Figure 34 – Localization and identification of proteasome subunits purified using the TAP-tag method, on the *S. cerevisiae* 2D-map.** A – Silver-stained gel showing the proteins recovered after purification of TAP-tagged Rpn5p (lane 1) and Scl1p (lane 2), subunits of the 19S regulatory particle and 20S core particle, respectively. Molecular weight markers are represented in the left lane. B - Coomassie blue-stained gel presenting the TAP-purified 20S proteasome subunits. Spots were identified by mass spectrometry. C – *S. cerevisiae* proteome map showing the localization of the TAP-purified 20S proteasome subunits.



The results for the 20S proteasome subunits are shown in Figure 34. Unfortunately, due to technical problems the gel for the proteasome 19S regulatory sub-complex, purified from the *RPN5*-tagged strain, was not obtained in proper conditions, so it is not possible to take any conclusions for the subunits belonging to the 19S particle. On the other hand, the results for the 20S proteasome structural subunits were very close to what has been described during the course of this work (Iwafune *et al.*, 2002).

The localization of the 20S proteasome subunits on the 2D-map allowed the assignment of unidentified spots on previous experiments, corresponding to spots induced by CUG ambiguity, as proteasome subunits. In particular, on the proteome map of *S. cerevisiae* cells expressing the *C. albicans* Ser-tRNA<sub>CAG</sub>, when compared to the control cells, at the normal growth temperature of 25°C, the spot number 684 (Figure 25 on chapter III.2.2.) could correspond to Pre3. If so, in addition to the increased expression levels verified for the proteasome subunits Rpn10p, Rpn12p, Scl1p and Pup2p in response to genetic code ambiguity, also Pre3p shows a 4.2-fold induction ( $p = 0.0014$ ). Similarly, on the proteome map of *S. cerevisiae* cells expressing the *C. albicans* Ser-tRNA<sub>CAG</sub>, when compared to the control cells, grown at the sub-optimal temperature of 37°C, the spot number 659 (Figure 30 on chapter III.2.3.1.) could correspond to Pre4. If so, in addition to the increased expression levels verified for the proteasome subunits Rpn10p, Rpn12p, Rpt3p, Rpt4p, Pre8p and Pre9p in response to genetic code ambiguity under an additional permanent stress, also Pre4p shows a 3.4-fold induction ( $p = 0.0056$ ).

These protein identifications further support the hypothesis that the proteasome is a major target of genetic code ambiguity. However, the fact that it was not possible to determine whether all the proteasome subunits are induced by CUG mistranslation, or whether only some subunits increase expression, prompted the study of proteasome activity (see chapter III.3.3.).

## **2.5. Discussion**

2D-PAGE is a good approach to separate and identify proteins and also allows the detection of post-translational modifications, but its scope is limited since only abundant and soluble proteins are detected. Low abundance proteins, namely transcriptional factors and regulatory proteins, are under-represented on the 2D maps. Hydrophobic membrane proteins and proteins with extreme molecular weight or isoelectric point (either very high or very low) also fail to be detected by this technique. Therefore, 2D-PAGE can only identify a subset of the whole proteome, hence the need to analyse gene expression at a genomic scale using DNA microarrays.

Data from transcriptomics and proteomics provide complementary information, and consequently characterization of a biological system is best achieved by comparison of the two data sets and combination of the results from both approaches. For many genes, the mRNA levels cannot predict the corresponding protein abundance in yeast (Gygi *et al.*, 1999; Ideker *et al.*, 2001; Griffin *et al.*, 2002). Microarrays can provide data on the expression levels of potentially all genes from a single genome, although they may not correlate with protein expression levels due to translational regulation of gene expression or post-translational modifications of proteins.

The profiling of mRNAs associated to polysomes could give a closer picture of gene expression, as only the mRNAs that are actually being translated will be detected (Arava *et al.*, 2003). Despite this, post-translational modifications of proteins will be missed in the microarrays. Therefore, an integrated approach that takes into account both transcriptome and proteome profiling is likely to provide valuable information about gene expression.

The results described in this chapter showed that genetic code ambiguity alters the expression of several proteins, belonging to specific functional categories. The main changes occur at the level of the stress response, with the induction of molecular chaperones and other stress responsive proteins, as well as protein degradation, as deduced from the up-regulation of several proteasome subunits. Protein synthesis and metabolic pathways such as amino acid metabolism are also repressed in response to CUG mistranslation. The proteome analysis data agrees with the transcriptome profiling data, with one exception: cell wall and membrane proteins are not represented on the 2D-gels, although the expression of their genes is altered in CUG mistranslating cells. However, because the fold increase in the expression of particular genes does not correlate with the fold increase in the expression of the corresponding proteins, these results suggest that translational control of gene expression might occur in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, highlighting once more the importance of combining proteomics and transcriptomics data in order to achieve a global gene expression characterization under a given condition.

It would be interesting to find a positive control condition that could mimic the effect of the *C. albicans* ser-tRNA<sub>CAG</sub> on these cells. Although the proteome analysis of *S. cerevisiae* cells grown in the presence of geneticin or canavanin, to induce mistranslation and misfolding, respectively, was inconclusive (data not shown), treatment with AZC or tunicamycin could be good candidates.

The cell stress response is activated in response to the *C. albicans* ser-tRNA<sub>CAG</sub>, with increased expression levels of molecular chaperones. As mistranslation of the CUG codon results in the production of aberrant proteins, the mobilization of molecular chaperones to correct the misfolded proteins occurs to prevent the formation of toxic protein aggregates. However, such mobilization of the chaperones does not compromise the cellular response to an additional stress, as demonstrated for the heat shock. Indeed, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> are able to further induce the expression of heat shock proteins following a temperature upshift from 25°C to 37°C, during 30 minutes.

Therefore, cells expressing genetic code ambiguity are under a permanent and internal stress that causes major changes in protein expression, but are not impaired in their capacity to respond to different unfavourable conditions that also require gene expression reprogramming. Despite their normal response to a transient stress, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> grown under an additional permanent stress do not show further induction of the stress-protective proteins, as observed for the control cells. Due to pre-existent increased amounts of these proteins, in the long-term these cells are able to cope with adverse environments, which has implications for their adaptation and evolution. In fact, organisms that experience a genetic code change have to deal with constant proteome destabilization, however by reprogramming gene expression in response to ambiguity, cells also trigger mechanisms that enhance cell survival under distinct stress conditions, thus preventing cell collapse and viability loss.

### 3. Physiological characterization of *S. cerevisiae* cells expressing CUG ambiguity

#### 3.1. Introduction

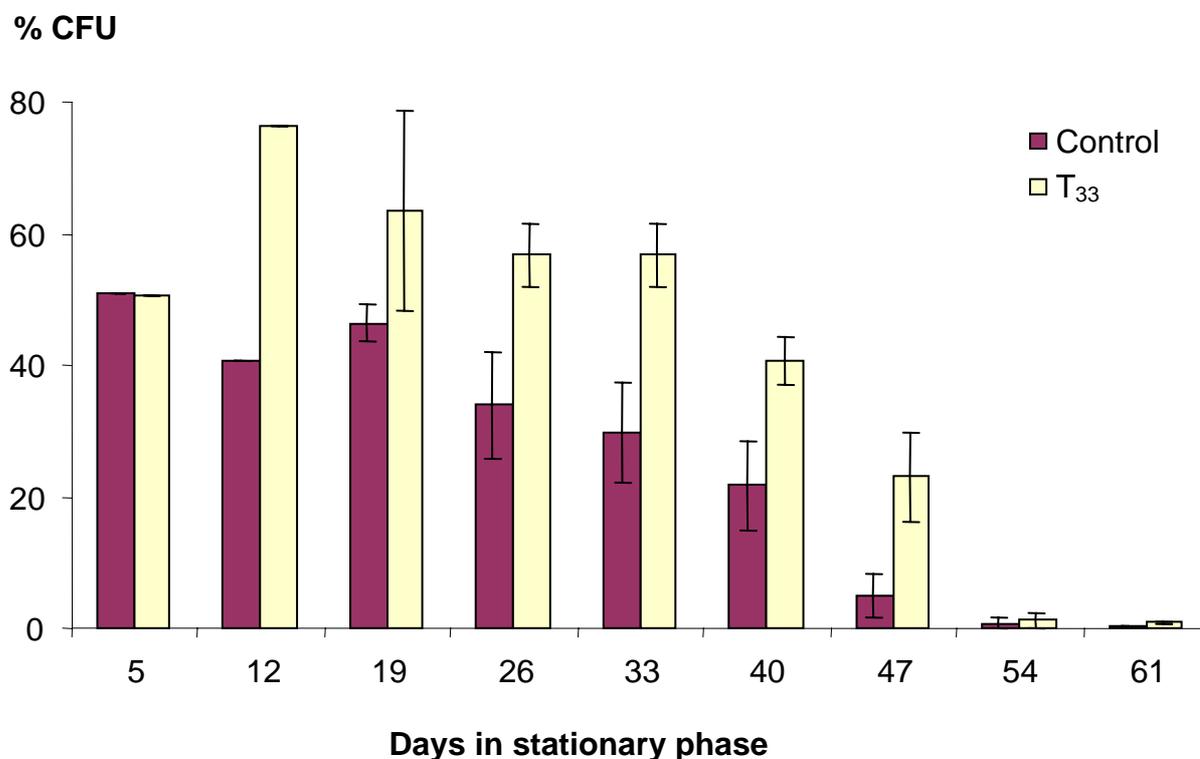
Transcriptome and proteome analysis of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> provided important insight on the impact of genetic code alteration on normal cellular functioning. The remodelling of gene and protein expression, triggered by genetic code ambiguity, with changes in metabolic pathways such as carbohydrate, amino acid or phosphate metabolism, cell wall structure and function, the stress response, protein synthesis and protein degradation, shows that these cells adapt and survive to a permanent stress condition that generates proteome and genome destabilization. Understanding in detail how cells cope with such disruption is, therefore, of paramount importance to elucidate the evolution of the genetic code.

Some of the questions that ought to be answered are “what is the role of Pnc1p, the protein that is most induced in ambiguous cells, on the protection from the damaging effect of an ambiguous genetic code?” and also, “is the proteasome activity increased in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>?”. If so, “are the aberrant proteins produced by CUG mistranslation being targeted for degradation by a mechanism related to increased ubiquitylation or oxidation levels?”. Other issues remained unclear, namely “is there accumulation of the compatible solutes trehalose and glycogen in response to genetic code ambiguity?” , or “are there karyotype alterations in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>?”. To clarify these questions, several gene and protein targets uncovered both by transcriptomics and proteomics were further tested, as described in the next chapters.

### 3.2. Effect of CUG ambiguity on cell viability

As highlighted before, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> are under permanent stress, resulting in gene expression reprogramming. These CUG-mistranslating cells grow slower than control cells, possibly because they have to deal with the constant destabilization of their genome and proteome, and have to mobilize large quantities of energy to protein degradation and repair systems. An increased genomic and proteomic instability has been associated with aging (McMurray and Gottschling, 2003; Reverter-Branchat *et al.*, 2004), which raises the important question of “are *S. cerevisiae* cells expressing CUG ambiguity able to cope with increasing disruption of their transcriptome and proteome, as errors accumulate during cell growth and aging?”. In *S. cerevisiae*, there are two types of aging, replicative and chronological (Reverter-Branchat *et al.*, 2004). Replicative life span refers to the number of times a cell divides, and yeast cells go through a finite number of divisions. On the other hand, chronological life span is related to the ability of stationary-phase cultures to maintain viability over time. In these non-dividing cells there is accumulation of damaged cellular components, and a system that could get rid of them or prevent such damage would increase chronological life span (Reverter-Branchat *et al.*, 2004). Otherwise, the gradual increase of damaged cellular components reduces chronological life span, and yeast cells lose viability. As *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> are likely to accumulate damaged proteins even in exponential growth, due to CUG mistranslation, the question that has to be answered is “do these cells maintain viability on the long term, or is their chronological life span affected by CUG ambiguity?”.

In order to determine the viability of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T33</sub>, cultures were grown at 30°C to stationary phase and were kept in culture for a period of about 70 days. At each time point, culture aliquots were taken and after adequate dilution cells were counted, plated and the colony forming units (CFU) determined as a percentage of the total number of cells plated (Figure 35).



**Figure 35 – CUG ambiguous cells do not lose viability when compared to control cells.** Viability of *S. cerevisiae* control cells and cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-T<sub>33</sub>, from 5 to 61 days in stationary phase. Results are an average of 3 independent cultures, and are presented in colony forming units (CFU) as a percentage of the cells plated.

Surprisingly, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> do not lose viability in stationary phase, and cells expressing CUG ambiguity are even more viable than the control cells at most time points. Indeed, the percentage of CFU from ambiguous cells is significantly higher than the percentage of CFU from control cells ( $p < 0.05$ ), after 26, 33, 40, 47 and 61 days in stationary phase (Figure 35). Although the normal cellular functioning is altered by genetic code ambiguity, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> showed increased viability in stationary phase, therefore, their chronological life span is increased.

As yeast cells age, metabolic and structural changes occur. For example, cells become stress resistant, tend to accumulate glycogen and trehalose and the cell wall thickens (Werner-Washburne *et al.*, 1996). Curiously, these aging phenotypes such as stress resistance (Santos *et al.*, 1996; Santos *et al.*, 1999), glycogen and trehalose accumulation (chapter III.3.4.) and cell wall alterations (chapter III.1.) were observed in exponentially-growing CUG ambiguous cells. Could these cells be pre-adapted to survive in stationary phase, as described for the growth at the sub-optimal temperature of 37°C (chapter III.2.3.1.)? Like a cross-protection mechanism, the stress response triggered by expression of the ser-tRNA<sub>CAG</sub> could enable CUG ambiguous cells to adapt easier to the additional stress caused by entry in stationary-phase. If so, what would be the proteins underlying such capacity?

One possibility was raised after recent work that showed that the *PNC1* gene plays an important role in aging (Anderson *et al.*, 2003). This gene encodes a pyrazinamidase and nicotinamidase, an enzyme that deaminates nicotinamide and activates Sir2p, which is a NAD-dependent histone deacetylase required for lifespan extension by calorie restriction (Lin and Guarente, 2003; Anderson *et al.*, 2003; Gallo *et al.*, 2004). Sir2p is involved in silencing chromatin at the telomeres, ribosomal DNA (rDNA) and mating type loci, and deletion of the *SIR2* gene promotes aging by increasing recombination at the rDNA locus (Sinclair and Guarente, 1997). Recent studies suggest that Sir2p is also a negative regulator of chromosomal DNA replication in yeast (Pappas, Jr. *et al.*, 2004). NAD

dependence of Sir2p activity links metabolism and aging (Lin and Guarente, 2003), and might provide the cell with a mechanism to slow down the aging process during times of starvation, allowing survival when the conditions improve.

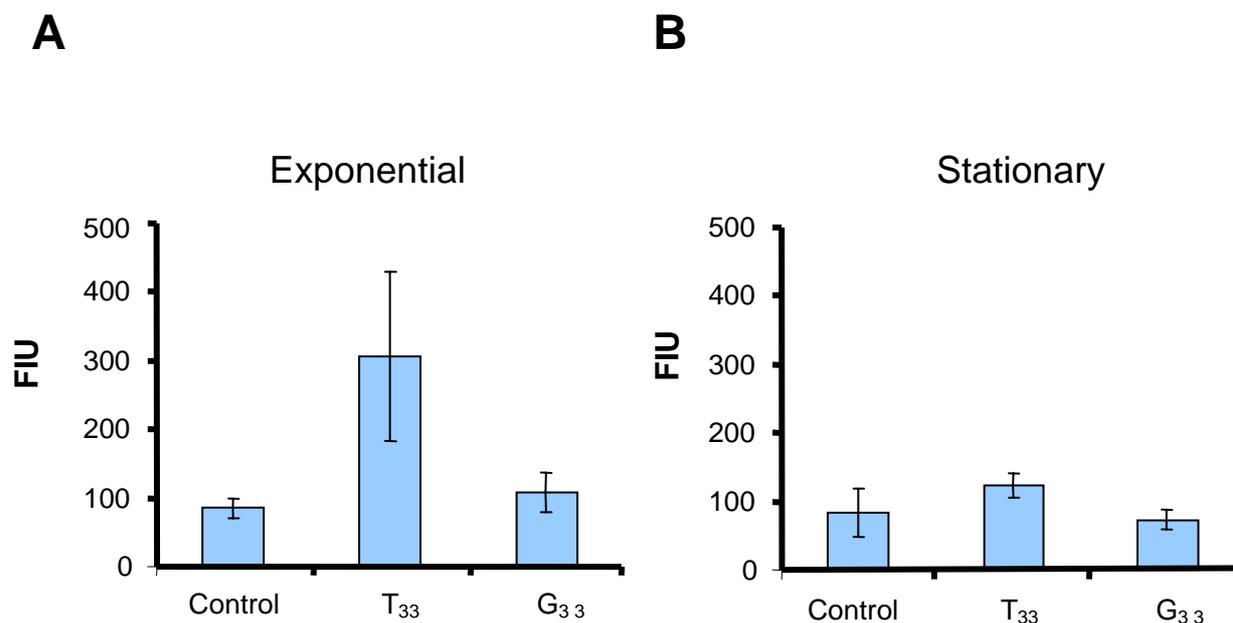
*PNC1* was identified as a longevity gene responsive to all stimuli that extend lifespan. In fact, it has been demonstrated that lifespan extension by a mild stress requires up-regulation of *PNC1*, and that a yeast strain with 5 copies of *PNC1* lived 70 % longer than the wild type (Anderson *et al.*, 2003). Furthermore, these authors discovered that Pnc1p localized to the peroxisome under stress conditions, suggesting that this enzyme might regulate proteins other than Sir2p.

Interestingly, *PNC1* is very sensitive to environmental conditions and was, therefore, identified as a gene with increased expression levels in presence of several stresses. For example, it increases after heat shock (Boy-Marcotte *et al.*, 1999), following oxidative stress (Godon *et al.*, 1998), cadmium exposure (Vido *et al.*, 2001) or treatment with the proline analog AZC (Trotter *et al.*, 2002). Although in other studies the observed fold increase for Pnc1 was 6.8 at the protein expression level (Godon *et al.*, 1998) and 5.2 at the mRNA expression level (Trotter *et al.*, 2002), Pnc1p showed a fold increase of 29.5 in *S. cerevisiae* cells expressing the *C. albicans* Ser-tRNA<sub>CAG</sub> (chapter III.2.2.1.). In fact, Pnc1p is the protein most responsive to genetic code ambiguity explaining why these ambiguous yeast cells do not lose viability. In addition, the molecular chaperones induced by genetic code ambiguity (see chapters III.1.2.1. and III.2.2.1.) have also been associated with longevity in *Caenorhabditis elegans* (Morley and Morimoto, 2004) and *Drosophila melanogaster* (Morrow *et al.*, 2004), adding strength to the observed lack of loss of viability of yeast cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>.

### 3.3. Effect of CUG ambiguity on proteasome activity

Proteomics studies described before unravelled the proteasome as a major potential target of genetic code ambiguity. Several proteasome subunits (Pup2p, Scl1p, Rpn10p and Rpn12p) were induced in response to CUG mistranslation, showing a 4 to 5 fold increase in expression under normal growth conditions, although such increase was not observed under heat stress (see chapter III.2. for further details). Since proteasome mapping by 2D-PAGE (chapter III.2.4.) failed to detect most proteasome subunits, and transcriptome analysis (chapter III.1.) was unable to uncover any alteration of expression of the genes that code for proteasomal subunits, several questions remained unanswered. For example, “is the altered expression levels found by 2D-PAGE limited to certain proteasome subunits?”. Or, “is the full set of proteasome subunits induced in *S. cerevisiae* cells expressing the *C. albicans* Ser-tRNA<sub>CAG</sub>?” and “will such alteration have an impact on the activity of the proteasome?”. In an attempt of elucidating these questions, the activity of the proteasome was quantified in *S. cerevisiae* cells expressing the *C. albicans* Ser-tRNA<sub>CAG</sub>. For this, cells were grown at 30°C to an OD<sub>600nm</sub> of 0.5 (exponential phase) or 3.5 (stationary phase). Protein extracts were obtained as described in methods and proteasome activity was assayed using the fluorogenic peptide s-LLVY-MCA as substrate (Grune *et al.*, 1998; Demasi *et al.*, 2003).

Exponentially growing *S. cerevisiae* cells expressing genetic code ambiguity show increased proteasome activity levels (Figure 36A). Indeed, proteasome activity of cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> was always 2 to 4 fold induced relative to the control strain (p<0.01), in 6 independent experiments performed. The G<sub>33</sub> strain also showed a slight increase in proteasome activity when compared to control cells. These results confirm the proteomics data presented previously and reinforce the possibility that the expression of proteasome subunits in the presence of CUG ambiguity is regulated at the translational level, since no difference was detected at the mRNA level (chapters III.1. and III.2.).



**Figure 36 – Cells expressing CUG ambiguity have increased proteasome activity levels.** Proteasome activity was assayed using the fluorogenic peptide s-LLVY-MCA as substrate. Protein extracts (100  $\mu$ g) were incubated at 37°C with 50  $\mu$ M s-LLVY-MCA for 60 minutes and fluorescence emission was read at 435 nm. The results are expressed as mean  $\pm$  S.D. of four to six independent experiments. Fluorescence intensity (FIU) is shown in arbitrary units. A – Proteasome activity from *S. cerevisiae* cells growing in exponential phase ( $OD_{600nm} = 0.5$ ) at 30°C. Control represents *S. cerevisiae* control cells; T<sub>33</sub> and G<sub>33</sub>, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub>, respectively. B – Proteasome activity from *S. cerevisiae* cells grown to stationary phase ( $OD_{600nm} = 3.5$ ). Control, T<sub>33</sub> and G<sub>33</sub> are as in A.

It has been demonstrated that cells with increased proteasome activity lose viability upon entry into stationary phase (Bajorek *et al.*, 2003). Although proteasome activity is about 3 times higher in the *S. cerevisiae* strain expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> than in the control strain, CUG ambiguous cells do not lose viability in stationary phase, as demonstrated in the previous chapter. These results led us to determine the proteasome activity of *S. cerevisiae* cells expressing CUG ambiguity in stationary phase (Figure 36B). Proteasome activity of control cells showed no significant decrease upon entry in stationary phase, which is somehow puzzling, as it has been shown that proteasome-dependent proteolysis decreases during stationary phase (Bajorek *et al.*, 2003). However, these experiments were performed in “early” stationary phase, with cells from a 2-day culture, whereas other authors report decreases of proteasome activity in stationary-phase cells after 1 week in culture (Bajorek *et al.*, 2003). Therefore, those differences may reflect differences in the experimental procedure. The *S. cerevisiae* strain expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> shows a decreased proteasome activity upon entry into stationary phase, when compared with the proteasome activity level of the same cells growing in exponential phase ( $p < 0.05$ ). Moreover, the proteasome activity of these cells in stationary phase is close to the proteasome activity of control cells. This decreased proteasome activity in stationary-phase *S. cerevisiae* cells expressing the *C. albicans* Ser-tRNA<sub>CAG-T<sub>33</sub></sub> correlates with the fact that these cells do not lose viability (previous chapter), suggesting the existence of a regulatory mechanism that allows a balance between degradation of aberrant proteins and viability maintenance.

There is undoubtedly increased protein degradation in *S. cerevisiae* cells expressing genetic code ambiguity. Mistranslation of the leucine-CUG codon as serine will result in the production of aberrant proteins that cannot be correctly folded, even with the increased levels of molecular chaperones. These proteins are potentially toxic to the cell as they may have altered functions or form aggregates, and a mechanism to prevent such toxicity is protein degradation. Indeed, it has been suggested that the induction of heat shock proteins during stress functions essentially to remove misfolded and aggregated proteins, implying

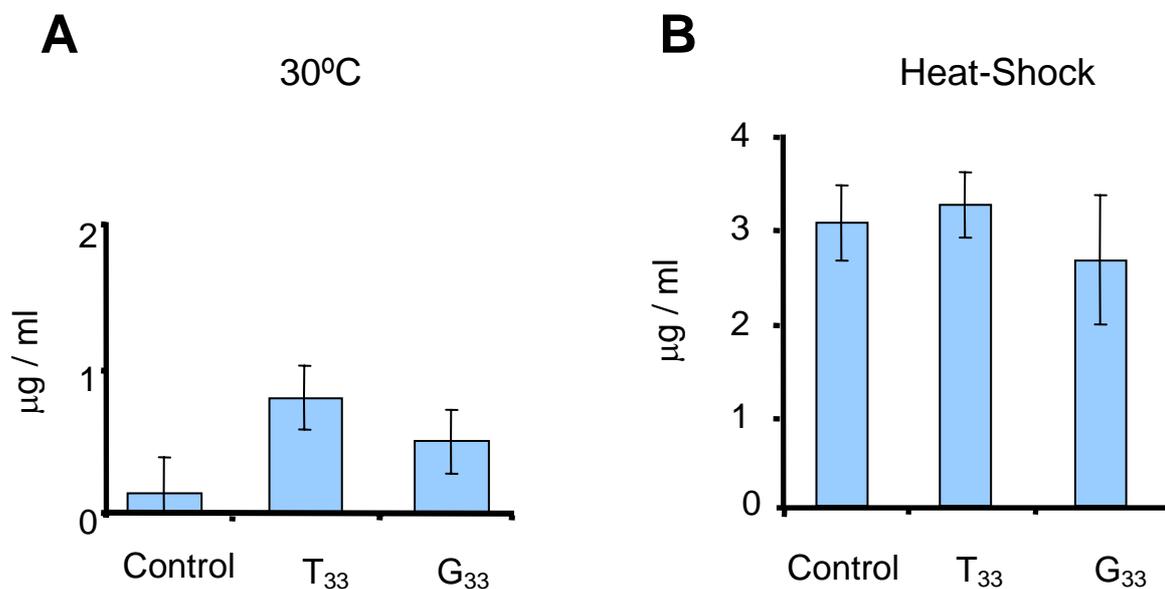
that cell death by heat shock occurs due to the toxicity of aggregated proteins instead of the loss of function of the denatured proteins (Friant *et al.*, 2003). Moreover, the same work demonstrated that an increase in the ubiquitin-dependent degradation could replace the role of the molecular chaperones in stressed cells. Therefore, the induction of the proteasome in *S. cerevisiae* cells expressing CUG ambiguity is crucial to protect these cells, by increasing the degradation of the abnormal proteins produced and impairing their aggregation. But the proteasome functions beyond just getting rid of abnormal proteins, since there is recent evidence that the ubiquitin-proteasome system is involved in the control of gene expression (reviewed in Conaway *et al.*, 2002; Muratani and Tansey, 2003). The role of ubiquitin in transcription is not new, as it has been found that the ubiquitylated forms of histones H2A and H2B are associated with actively transcribed genes and that the RNA polymerase II is a target for regulation by ubiquitylation. Additionally, the ubiquitin-proteasome system tightly controls the abundance and function of transcriptional factors (Conaway *et al.*, 2002; Muratani and Tansey, 2003). For some transcription factors, binding to the promoters of target genes triggers their ubiquitylation and, therefore, tags them for destruction by the proteasome. This allows the expression of target genes to occur, and then the degradation of the activated transcription factor would shut off the signal rapidly. Using this mechanism, the cell is ready to respond to the stimuli that activate gene expression, assuring that the same pathway will not remain active when it is no longer needed (Conaway *et al.*, 2002; Muratani and Tansey, 2003).

Increased proteasome activity in *S. cerevisiae* cells expressing CUG ambiguity might, therefore, result not only from increased levels of mistranslated proteins that have to be destroyed, but may also reflect increased degradation of proteins implicated in regulatory processes, namely the transcription factors that control and remodel gene expression in response to stress. Anyway, proteasome induction is a protection mechanism in cells expressing CUG ambiguity.

### **3.4. Effect of CUG ambiguity on trehalose and glycogen accumulation**

Glycogen and trehalose are two major reserve carbohydrates in *S. cerevisiae*, and are known to accumulate in yeast cells as part of the stress response (Parrou *et al.*, 1997). Under stress conditions, trehalose is more than just an energy reserve. Despite its function as compatible osmolyte, a major role of trehalose is protein stabilization, since this carbohydrate stabilizes the structure of enzymes at high temperatures (Kaushik and Bhat, 2003) and even decreases the aggregation of unfolded or heat-denatured proteins (Singer and Lindquist, 1998; Ueda *et al.*, 2001). Therefore, trehalose could be fundamental to the protection of *S. cerevisiae* cells by promoting stabilization of the misfolded proteins produced by CUG mistranslation. Interestingly, recent work indicates that there is no accumulation of trehalose or glycogen in *C. albicans* cells in response to heat and oxidative stresses, and hyperosmotic stress only slightly increased the production of these reserve sugars (Enjalbert *et al.*, 2003). These authors argue that these and other evidences point to a lack of the general stress response in *C. albicans* cells. However, these observations also prompt important questions. “Is there accumulation of trehalose or glycogen in *S. cerevisiae* cells in response to CUG ambiguity?” and, if so, “are these levels similar to wild-type *S. cerevisiae* cells exposed to standard stress conditions?”.

In order to better characterize the stress response of *S. cerevisiae* cells expressing the *C. albicans* Ser-tRNA<sub>CAG</sub>, the accumulation of trehalose and glycogen in these cells was investigated and compared with the response to a standard stress, namely heat shock. Cells were grown at 30°C to an OD<sub>600nm</sub> of 0.5. For heat shock, cultures were grown at 25°C to an OD<sub>600nm</sub> of 0.5 and then transferred to 37°C for 30 minutes. Samples were prepared as described in methods and incubated either with trehalase or amyloglucosidase, for the quantification of trehalose and glycogen, respectively. Glucose present on the supernatant, originated from trehalose or glycogen breakdown, was determined using the glucose oxidase/oxidase kit from Sigma (see methods for details).



**Figure 37 - Trehalose accumulation increases in cells expressing CUG ambiguity.** Trehalose accumulation was measured as glucose content (in  $\mu\text{g}$  per ml cells) after treatment of cell extracts with trehalase. The results are expressed as mean  $\pm$  S.D. of four independent experiments. A – Trehalose accumulation in *S. cerevisiae* cells growing in exponential phase ( $\text{OD}_{600\text{nm}} = 0.5$ ) at 30°C. Control represents *S. cerevisiae* control cells; T<sub>33</sub> and G<sub>33</sub>, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub>, respectively. B - Trehalose accumulation in *S. cerevisiae* cells after temperature upshift (25°C to 37°C, 30 minutes). Control shows *S. cerevisiae* control cells; T<sub>33</sub> and G<sub>33</sub>, *S. cerevisiae* cells expressing CUG ambiguity (*C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub>, respectively).

The data indicate that there is an increase in trehalose content in yeast cells expressing genetic code ambiguity when compared to the control cells ( $p < 0.05$ ), although this difference is higher for the ser-tRNA<sub>CAG-T<sub>33</sub></sub> than for the ser-tRNA<sub>CAG-G<sub>33</sub></sub> transformed strain (Figure 37A). However, this trehalose accumulation is much lower under normal growth conditions than under heat-shock (Figure 37B). When compared to the control cells, the response to the heat shock is similar in *S. cerevisiae* cells expressing the *C. albicans* Ser-tRNA<sub>CAG</sub>, in both T<sub>33</sub> and G<sub>33</sub> strains (Figure 37B).

Trehalose content has been shown to be close to zero in wild-type yeast cells grown in rich medium. However, trehalose levels increase transiently after a temperature upshift from 25 to 37°C, reaching a maximum at 30 minutes of stress and decreasing after this period (Parrou *et al.*, 1997). The same study showed that mutant strains defective in trehalase can accumulate trehalose at higher levels and in a sustained manner, and that trehalose content increases with an increase in the temperature at which the heat shock is performed (Parrou *et al.*, 1997). Accordingly, the trehalose content of *S. cerevisiae* control cells is close to zero and in cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> trehalose levels are close to the observed in heat stressed wild-type cells. On the other hand, trehalose values after a 30-minute temperature upshift from 25 to 37°C are much higher than expected, even in the control cells. This could be due to the differences in the strains and culture conditions, since *S. cerevisiae* cells were grown in minimal medium and nutrient starvation triggers trehalose and glycogen accumulation (Parrou *et al.*, 1997).

Proteomics results from *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> showed no increase in trehalose metabolism enzymes under normal growth conditions, although a 2-fold increase on Tps1p expression was observed but was not considered significant due to high variability of the data for this protein (chapter III.2.2.2.). However, after temperature upshift from 25 to 37°C both control and cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> had increased expression of proteins belonging to the trehalose biosynthesis pathway, and with a similar fold

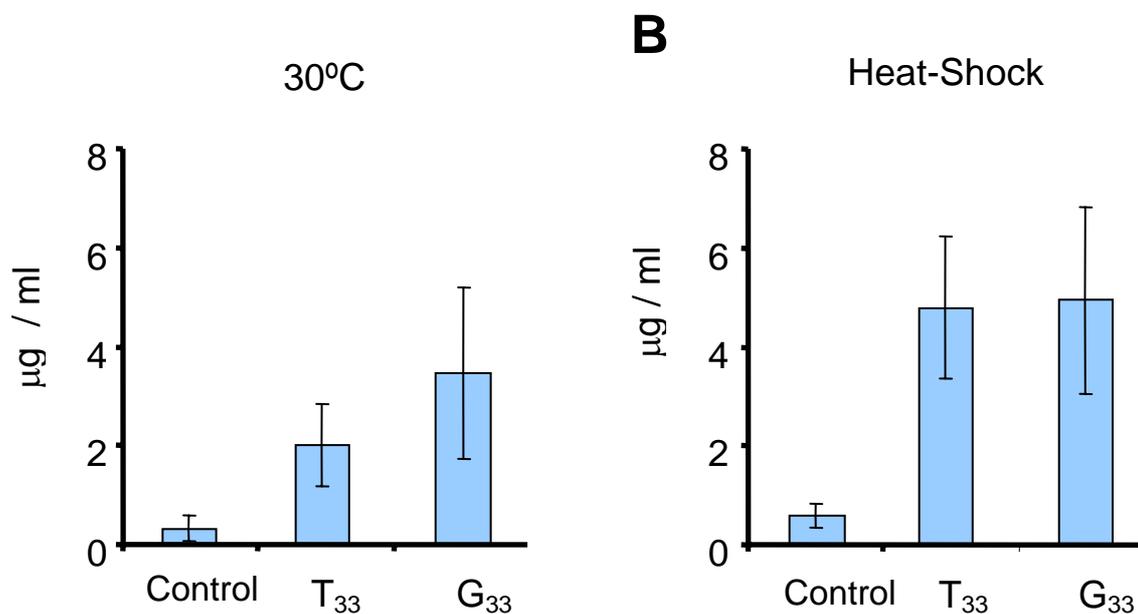
induction (chapter III.2.3.2.), consistent with the trehalose levels presented in Figure 36.

Transcriptomics data (chapter III.1.) revealed that the expression of the components of the trehalose synthase, *TPS1*, *TPS2* and *TSL1*, is increased 1.7, 1.5 and 2.9 fold in the T<sub>33</sub> strain and 2.6, 2.1 and 3.6 in the G<sub>33</sub> clone, respectively. Therefore, one expects that the G<sub>33</sub> strain would have accumulated more trehalose than the T<sub>33</sub>, but the opposite was observed. This paradox may be explained by the observation that the trehalase *NTH1* shows an induction of 1.2 in the T<sub>33</sub> strain and 1.6 in the G<sub>33</sub>, suggesting that more trehalose degradation also occurs on the G<sub>33</sub> strain. An increased recycling of trehalose and glycogen has already been reported, when both synthesis and degradation pathways are activated, and hence the lack of correlation between transcriptional induction and accumulation of these reserve carbohydrates (Parrou *et al.*, 1997). The simultaneous induction of genes that encode enzymes from the trehalose and glycogen synthesis and degradation pathways was also described in the ESR (Gasch *et al.*, 2000). As the activity of these enzymes might be regulated post-translationally, this mechanism enables the cell to quick and precisely respond to stress conditions by controlling the flux of carbohydrates in and out of its energy stores.

Taking into account the role of trehalose in maintaining the structure and function of proteins (Kaushik and Bhat, 2003), and impairing the formation of toxic protein aggregates (Singer and Lindquist, 1998; Ueda *et al.*, 2001) under stress conditions, the increased trehalose levels in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> might function besides carbohydrate storage. Indeed, trehalose accumulation would act as yet another mechanism to protect the cell from misfolded proteins produced by CUG mistranslation. Additionally, during heat shock, trehalose accumulation would be important both as energy reserve and as a stabilizer of proteins denatured by elevated temperatures.

There is an increase of glycogen levels in yeast cells expressing genetic code ambiguity when compared to the control cells ( $p < 0.05$ ), in both *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub> and G<sub>33</sub> (Figure 38A). In addition, glycogen accumulation is higher in heat-shocked cells (Figure 38B). Glycogen content in a wild-type strain grown in rich medium is close to 1  $\mu\text{g}$  of glucose per  $10^7$  cells. After a temperature upshift from 25 to 37°C, glycogen levels increase reaching a maximum of about 3  $\mu\text{g}$  of glucose per  $10^7$  wild-type cells after 60 minutes of stress, a value that is maintained at least until 120 minutes of stress (Parrou *et al.*, 1997). Glycogen accumulation for the control strain is in agreement with these previously published data, and the cells expressing CUG ambiguity show glycogen levels similar to a heat shocked wild-type culture. Control heat stressed cells should show higher levels than those observed, and despite an increase in glycogen content it was not significant compared to regular levels. Maybe a 30-minute stress was not enough to detect a significant glycogen accumulation in the control condition, as glycogen level peaks at 60-120 minutes after a temperature upshift (Parrou *et al.*, 1997).

There is no proteomics data for the enzymes involved on glycogen metabolism, as these proteins are not identified on the 2D maps used. However, transcriptomics data (chapter III.1.) show that the genes encoding enzymes belonging to glycogen biosynthesis pathways, *GSY1*, *GSY2* and *GLC3*, are increased 1.2, 1.4 and 1.5 fold in the T<sub>33</sub> strain and 1.5, 2.5 and 2.4 in the G<sub>33</sub> mutant, respectively. These results support the observed accumulation of glycogen at higher levels in the G<sub>33</sub> strain than in the T<sub>33</sub> mutant. Also *GPH1*, involved in glycogen degradation, shows a fold induction of 3.9 fold in the T<sub>33</sub> strain and 5.4 in the G<sub>33</sub> mutant, pointing towards an increased recycling of this reserve carbohydrate as observed before (Parrou *et al.*, 1997).



**Figure 38 - Glycogen accumulation increases in cells expressing CUG ambiguity.** Glycogen accumulation was measured as glucose content (in  $\mu\text{g}$  per ml cells) after treatment of cell extracts with amyloglucosidase. The results are expressed as mean  $\pm$  S.D. of four independent experiments. A – Glycogen content of *S. cerevisiae* cells growing in exponential phase ( $\text{OD}_{600\text{nm}} = 0.5$ ) at 30°C. Control represents *S. cerevisiae* control cells; T<sub>33</sub> and G<sub>33</sub>, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub>, respectively. B - Glycogen content of *S. cerevisiae* cells after temperature upshift (25°C to 37°C, 30 minutes). Control shows the results of *S. cerevisiae* control cells; T<sub>33</sub> and G<sub>33</sub>, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub>, respectively.

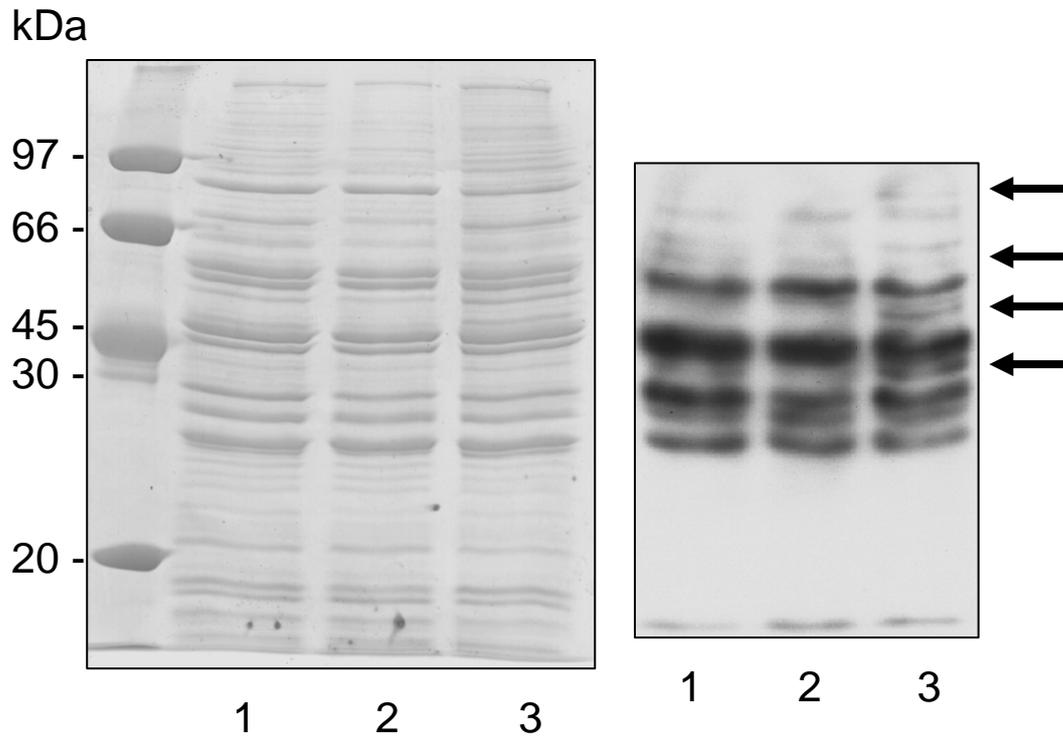
Recent work demonstrated that glycogen accumulation is dependent on glycogen synthase activation by glucose-6-P *in vivo* (Pederson *et al.*, 2004). Therefore, the induction of Glk1p in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> might increase the concentration of glucose-6-P available for glycogen synthesis. On the other hand, glycogen synthase activity is reduced after phosphorylation by the Pho85p cyclin-dependent protein kinase. The altered phosphate metabolism detected in cells expressing genetic code ambiguity (chapter III.1.2.3.) suggests phosphate starvation in these cells, which would also result in glycogen accumulation. As glycogen accumulation occurs when there is nutrient deprivation, this might provide the cell with a mechanism of survival in scarce resources.

Trehalose and glycogen levels are higher in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> than in control cells, which is a hallmark of nutrient-starved and stationary-phase yeast cells. However, the accumulation of these reserve carbohydrates in CUG ambiguous cells, in response to the production of misfolded proteins, serves both as an energy source for the ATP-dependent processes of folding and degradation of abnormal proteins, and as an additional mechanism for the stabilization of denatured proteins.

### 3.5. Effect of CUG ambiguity on protein oxidation and ubiquitylation

As cells age, there is an increase in the oxidative damage to cellular components, namely the accumulation of carbonylated proteins, which is thought to result from the production of reactive oxygen species during aerobic metabolism (Reverter-Branchat *et al.*, 2004). Additionally, it has been shown that error-prone ribosomes increase protein oxidation levels in *E. coli* (Ballesteros *et al.*, 2001) and similarly, treatment of cells with antibiotics that cause mistranslation also results in increased levels of protein carbonylation (Dukan *et al.*, 2000). These results indicate that a reduced translational fidelity, by producing misfolded and aberrant proteins that are prone to oxidative damage, consists of a distinct mechanism involved in protein oxidation (Nystrom, 2002). In order to prevent their cross-linking and aggregation, oxidized proteins are targeted for degradation by the proteasome independently of the ubiquitin signal (Davies, 2001). Indeed, it has been shown that the 20S proteasome recognizes and degrades oxidatively modified proteins, in contrast to the 26S proteasome, and the exposure of hydrophobic residues might function as the signal required for degradation (Davies, 2001).

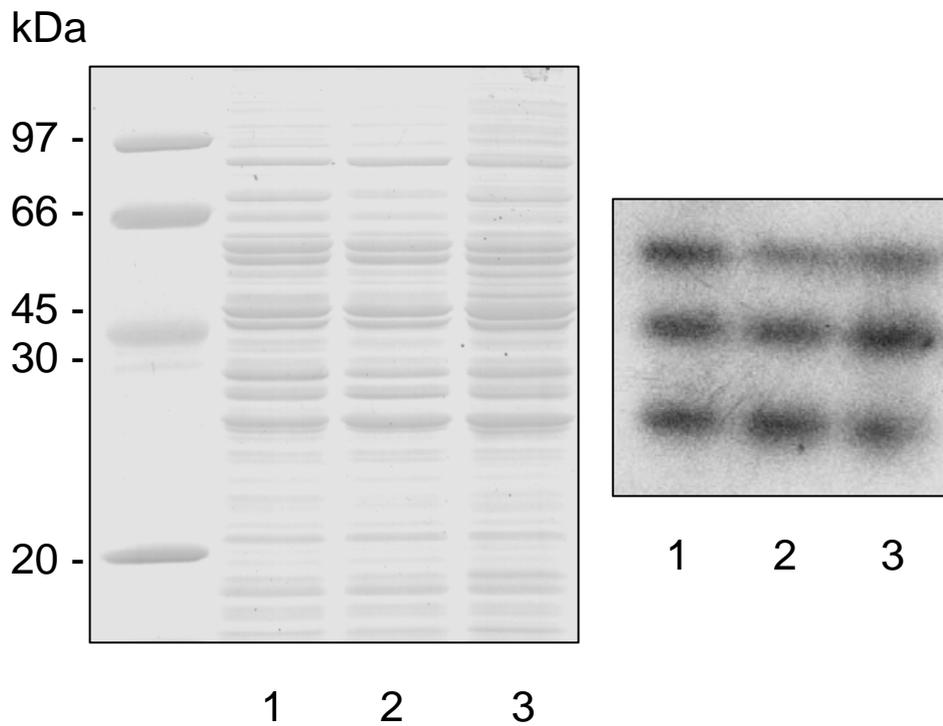
These observations raise the possibility that *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> have a higher content of oxidized proteins resulting from CUG mistranslation, and this increased level of oxidized proteins would result in increased proteasome activity as described (chapter III.3.3). To evaluate the extent of protein carbonylation in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, cultures were grown at 30°C to an OD<sub>600nm</sub> of 0.5. Protein extracts were prepared as described in methods and the carbonyl groups in the protein side chains were derivatized to 2,4-dinitrophenylhydrazone by 2,4-dinitrophenylhydrazine. Proteins were separated by SDS-PAGE and modified proteins were detected by Western blot with an anti-DNP antibody (Figure 39).



**Figure 39– Western blot analysis of carbonylated proteins.** 12  $\mu\text{g}$  of total protein extracts were separated by SDS-PAGE and stained with Coomassie Blue as a loading control (left panel), or incubated with an anti-DNP rabbit polyclonal antibody (Dako) and detected by ECL (Amersham) (right panel). Molecular weight markers are indicated in kDa. Lane 1, carbonylated proteins from *S. cerevisiae* control cells; lanes 2 and 3, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (T<sub>33</sub> and G<sub>33</sub>, respectively). Arrows indicate proteins with distinct carbonylation levels.

Surprisingly, there was no global increase on the content of protein carbonyls in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (Figure 39). However, there are some subtle differences on the pattern of carbonylated proteins, as shown by the arrows in Figure 38. This indicates that there might be distinct oxidation-prone proteins between the control cells and the strains that mistranslate the CUG codon. In order to identify such proteins, a possible approach would be to detect protein carbonyls on 2D gels instead of 1D, to compare both different patterns and amounts of protein oxidation. An additional experiment that should be performed would be to detect protein carbonyls in stationary-phase cells, instead of exponentially growing cells, as oxidatively damaged proteins accumulate with aging in yeast cells (Aguilaniu *et al.*, 2003). However, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> have increased chronological age when compared to the control cells (chapter III.3.2.), which might be correlated with decreased oxidative damage in cellular components. Therefore, more than detecting differences in the levels of protein carbonylation, it would be important to identify the targets of protein oxidation in CUG mistranslating cells.

*S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T33</sub> show increased expression of proteasome subunits and higher proteasome activity than control cells (chapters III.2.2.3. and III.3.3.). In general, substrates are targeted for degradation by the 26S proteasome through covalent attachment of multi-ubiquitin chains (Thrower *et al.*, 2000). Therefore, it is possible that the increase in proteasome expression and activity would result from an increase in ubiquitylated proteins. To address this question, *S. cerevisiae* control cells and cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> were grown at 30°C to an OD<sub>600nm</sub> of 0.5. Protein extracts were prepared as described in methods and fractionated by SDS-PAGE. Ubiquitylated proteins were detected by Western blot analysis using an anti-ubiquitin antibody (Figure 40).



**Figure 40 – Western blot analysis of ubiquitylated proteins.** 12  $\mu$ g of total protein extracts were separated by SDS-PAGE and stained with Coomassie Blue as a loading control (left panel), or incubated with an anti-ubiquitin rabbit polyclonal antibody and detected by ECL (Amersham) (right panel). Molecular weight markers are indicated in kDa. Lane 1, ubiquitylated proteins from *S. cerevisiae* control cells; lanes 2 and 3, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (T<sub>33</sub> and G<sub>33</sub>, respectively).

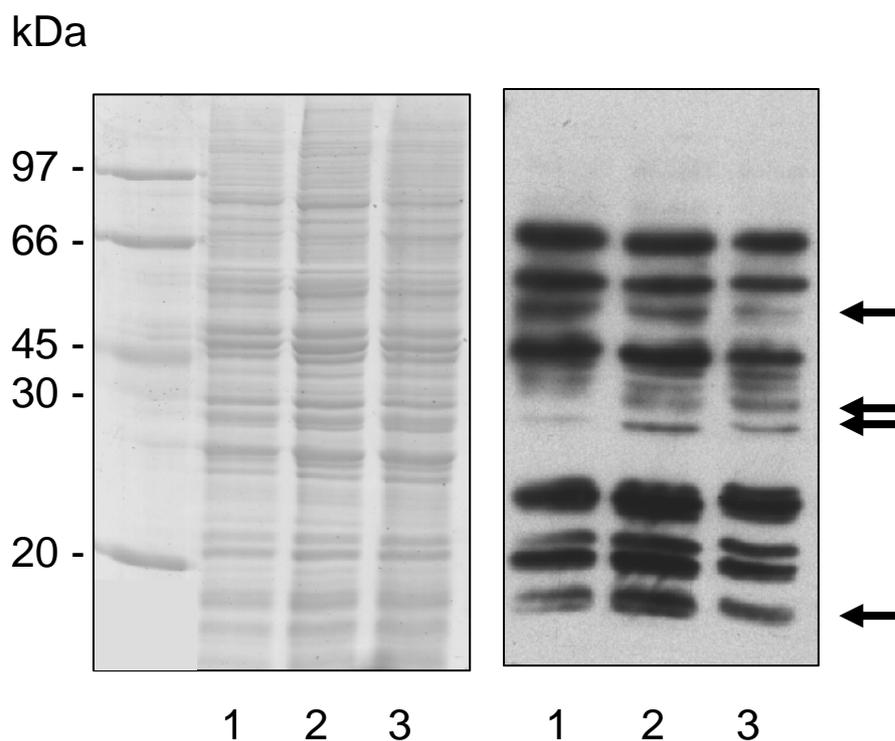
There was no increase in the content of ubiquitylated proteins in cells expressing genetic code ambiguity, and no differences were observed in the pattern of ubiquitylated proteins from control and *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (Figure 40). This unexpected result may be explained by the observation that ubiquitylation is reversible, and consequently a net increase in ubiquitylated proteins may be difficult to observe. However, detection of ubiquitin-protein conjugates can be accomplished by using proteasome inhibitors or aggregation-prone reporter proteins (Bence *et al.*, 2001), or alternatively the ubiquitin signal can be detected by Mass Spectrometry. Detection and identification of ubiquitylated proteins in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> is crucial to understand the mechanisms of protein degradation in these cells. Moreover, ubiquitylation is an important and complex post-translational modification of proteins. Multi-ubiquitin chains are the signal for substrate degradation by the 26S proteasome, but monoubiquitin has multiple roles as a regulator of protein localization and activity (reviewed in Hicke, 2001; Weissman, 2001; Schnell and Hicke, 2003). It is, therefore, important to detect both mono- and poli-ubiquitylated proteins in CUG ambiguous cells, in future studies.

Mistranslation of the leucine-CUG codon as serine produces aberrant proteins that have to be degraded to avoid their aggregation. The induction of proteasome activity (chapter III.3.3.), without increased levels of oxidized or ubiquitylated proteins (see above), suggests that protein degradation in these cells might proceed through a distinct mechanism. It has been described that some denatured proteins can be degraded without the ubiquitin signal (Benaroudj *et al.*, 2001) and, thus, unfolded proteins resulting from CUG mistranslation can also be degraded in an ubiquitin-independent manner, due to the exposure of motifs in their sequences or conformations that might be recognized by the 20S proteasome.

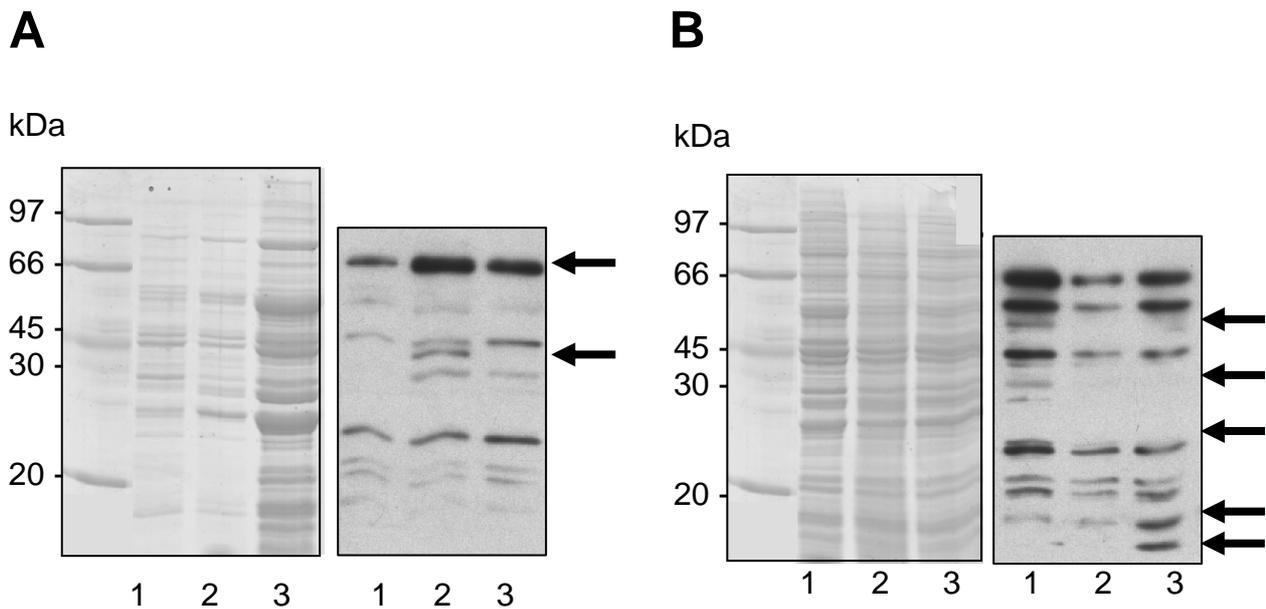
### 3.6. Effect of CUG ambiguity on phosphoserine proteins and HSF-1

Comparison between transcriptomics and proteomics data (chapters III.1. and III.2.) suggests that gene expression in *S. cerevisiae* cells expressing CUG ambiguity could be regulated at the translational level. Phosphorylation of serine 51 of the A subunit of the eukaryotic translation initiation factor 2 (eIF2) occurs under diverse stress conditions, such as nutrient starvation, ER stress and oxidative stress (Cox and Walter, 1996; Hinnebusch, 1996). This adaptation results in inhibition of protein synthesis and is thought to protect the cell against unfolded or modified proteins that accumulate under stress (Novoa *et al.*, 2003). Phosphorylation is one of the most important post-translational modifications that modulates protein activity, and occurs in tyrosine, threonine and serine residues. Besides the stress imposed on *S. cerevisiae* cells by the expression of the *C. albicans* ser-tRNA<sub>CAG</sub>, CUG mistranslation replaces leucines with serines. Although the level of altered proteins is a small subset of the normal cellular proteome, it is enough to cause significant alterations in the transcriptome and proteome of ambiguous cells. Therefore, it is possible that some of the altered proteins will give rise to new phosphorylation sites, or indirectly alter the level of phosphorylation of proteins as a consequence of regulatory processes. This raises the following questions: i) could this overall phosphorylation level be measured? ii) do *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> have different phosphoproteins? To tackle these questions, cells were grown at 30°C to an OD<sub>600nm</sub> of 0.5. Protein extracts were prepared as described in methods and fractionated by SDS-PAGE. Serine-phosphorylated proteins were detected by Western blot analysis using an anti-phosphoserine antibody.

Only minor differences in the phosphoprotein patterns were detected (Figure 41). However, these differences were highlighted by exposing cells to additional stress, such as heat shock and growth at the sub-optimal temperature of 37°C (Figure 42).



**Figure 41 – Western blot analysis of phosphorylated proteins from exponentially growing cells, at 30°C.** 5  $\mu$ g of total protein extracts were fractionated by SDS-PAGE and stained with Coomassie Blue as a loading control (left panel), or incubated with an anti-PhosphoSerine mouse monoclonal antibody (Qiagen) and detected by ECL (Amersham) (right panel). Molecular weight markers are indicated in kDa. Lane 1, PhosphoSerine proteins from *S. cerevisiae* control cells; lanes 2 and 3, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub>, respectively. Arrows indicate proteins with different phosphorylation levels.

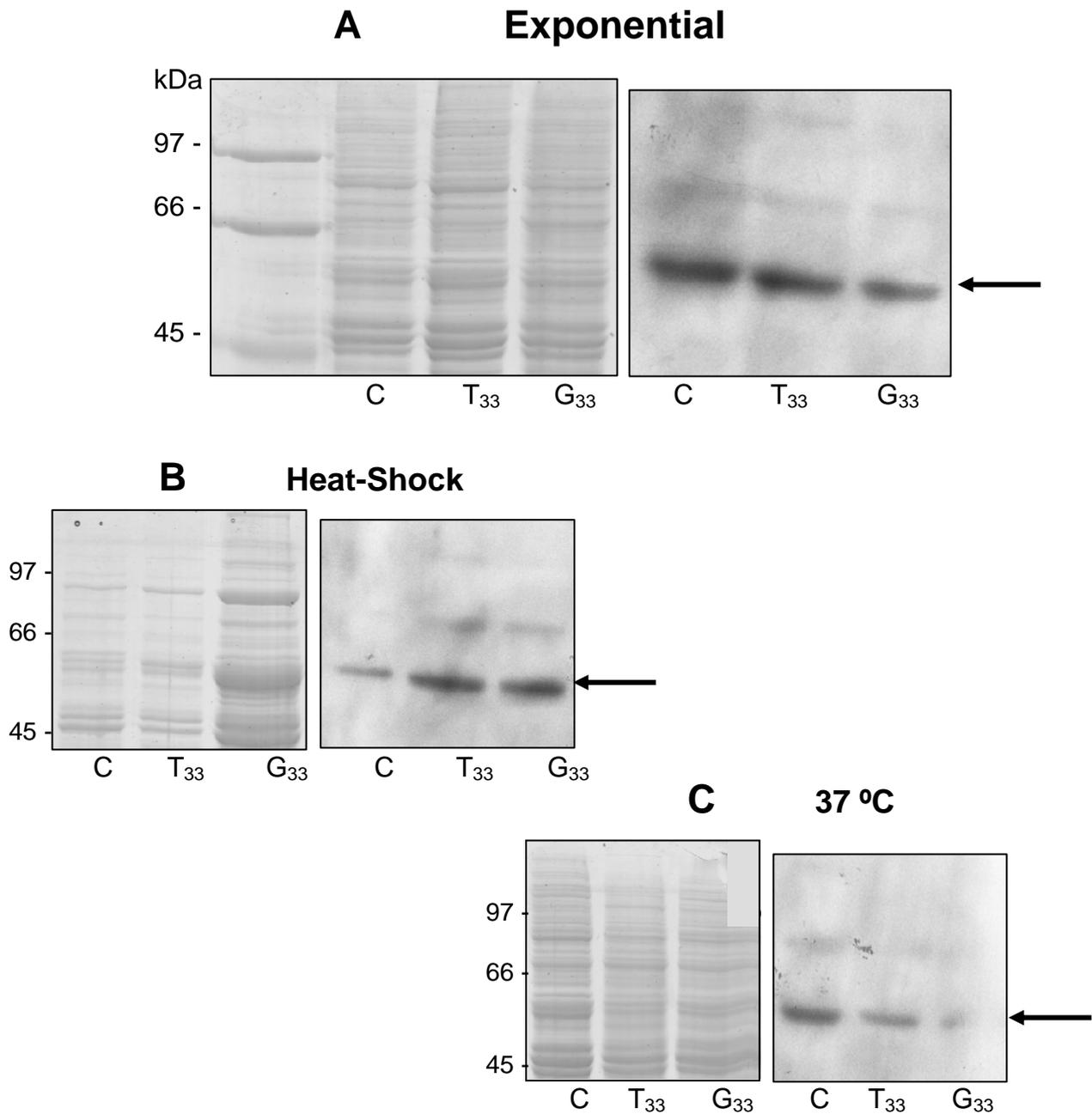


**Figure 42 – Western blot analysis of phosphorylated proteins from heat shocked cultures (A) and cultures grown at 37°C (B).** 5  $\mu$ g of total protein extracts were separated by SDS-PAGE and stained with Coomassie Blue as a loading control (left panel), or incubated with an anti-PhosphoSerine mouse monoclonal antibody (Qiagen) and detected by ECL (Amersham) (right panel). Molecular weight markers are indicated in kDa. Lanes 1, PhosphoSerine proteins from *S. cerevisiae* control cells; lanes 2 and 3, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub>, respectively. Arrows indicate proteins with different phosphorylation levels.

As before, 2D-gels may reveal differences in protein phosphorylation that could not be detected by 1D-PAGE. Additionally, the specific phosphorylation of translation factors, such as eIF2, in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, should be studied to elucidate how gene expression is regulated at the translational level in these cells.

HSF-1 is a master regulator of the stress response, controlling the transcription of many stress genes including the genes that code for the heat shock proteins (Boy-Marcotte *et al.*, 1999). HSF-dependent genes have increased mRNA (chapter III.1.) and protein (chapter III.2.) levels in response to genetic code ambiguity, thus raising the following questions: i) do *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> show alterations in HSF-1 expression? ii) does the expression of HSF-1 change in response to additional stress? To address these questions, cells were grown to an OD<sub>600nm</sub> of 0.5 at 30°C, 37°C and under heat shock. Protein extracts were prepared as described in methods and fractionated by SDS-PAGE. HSF-1 was detected by Western blot analysis using an anti-HSF-1 antibody (Figure 43).

The expression level of HSF-1 does not increase in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> (Figure 43), which might correlate with the complex regulation of this transcription factor. HSF-1 activity is regulated by post-translational modification, namely its phosphorylation is related with a decrease in activity (Bonner *et al.*, 2000). Therefore, one experiment that should be performed is the detection of phosphorylation changes on HSF-1 in CUG mistranslating cells, both under normal and stress conditions. Moreover, the interaction with SSA and SSB proteins modulates HSF activity, possibly through conformational changes (Bonner *et al.*, 2000). Since these proteins show altered expression levels in response to CUG mistranslation (chapter III.3.2.), it would be interesting to determine if these chaperones or additional proteins that might also interact with HSF-1 contribute to the activation of HSF activity in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>.

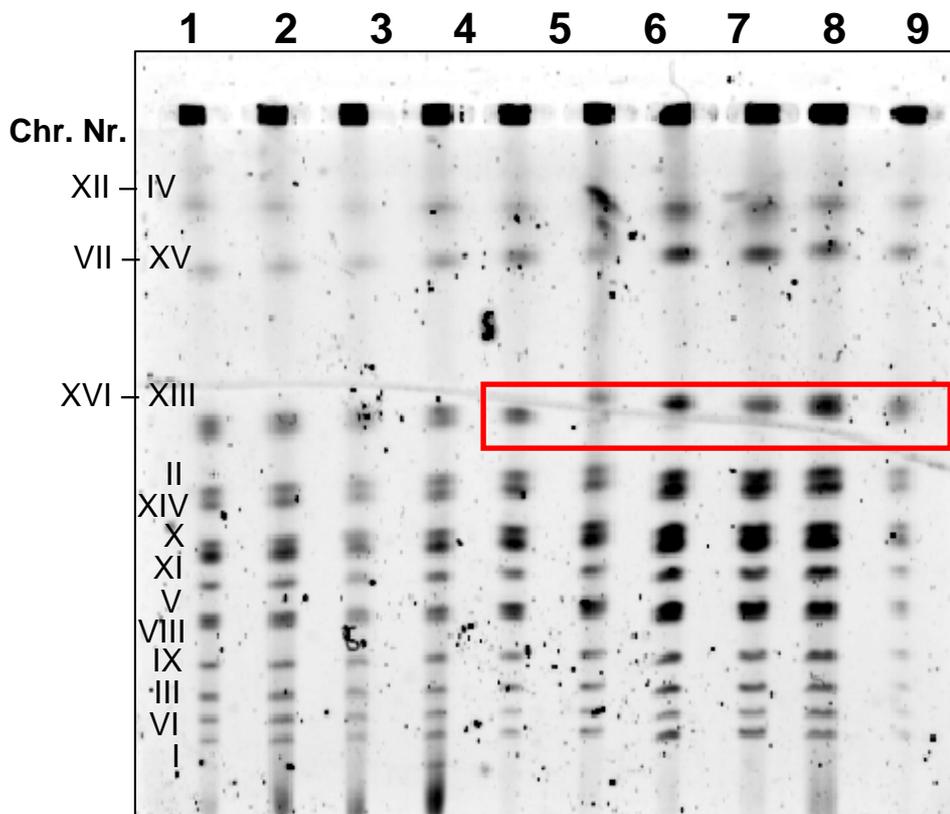


**Figure 43 – Western blot analysis of HSF-1 from CUG ambiguous cells.** 5  $\mu$ g of total protein extracts were fractionated by SDS-PAGE, incubated with an anti-HSF-1 rabbit polyclonal antibody (Stressgen) and detected by ECL (Amersham). Molecular weight markers are indicated in kDa. A, extracts from exponentially growing cells; B, extracts from heat shocked cells; C, extracts from cultures grown at 37°C. Control represents *S. cerevisiae* control cells; T<sub>33</sub> and G<sub>33</sub> represent *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub>, respectively. Arrows indicate HSF-1.

### 3.7. Karyotype of *S. cerevisiae* cells expressing CUG ambiguity

*S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> have an unstable genome, as pointed out by chromosome-wide expression biases (chapters III.1.3. and III.1.4.). Indeed, when the genes whose expression levels are altered in CUG ambiguous strains were distributed according to their chromosomal localization (Figures 22 and 24), there were entire chromosomes that appeared to be globally induced or repressed. The possibility that some chromosomes have been duplicated or deleted in these cells, raises the questions of “does CUG mistranslation generate genome instability?”, and “are mistranslating cells polyploid or aneuploid?”.

Genomic instability is a hallmark of cancer, and aging increases such instability (Sinclair, 2003; McMurray and Gottschling, 2003). Unstable genomes most often display chromosome number alterations and chromosome rearrangements. Cells with an increase in the DNA content that is equal for all chromosomes are polyploid, however deviations from the normal 2N set of chromosomes, along with structural rearrangements, creates aneuploid cells (Storchova and Pellman, 2004). Ploidy alterations can be found during normal development of organisms, namely in *Drosophila* salivary glands or mammalian megakaryocytes, and the generation of polyploids increases during cell senescence or under stress (Storchova and Pellman, 2004). Also during evolution, the formation of polyploids was not infrequent and consisted of a mechanism to generate variability. Indeed, genome duplication is evolutionary advantageous, as the divergence of duplicated genes allows the development of novel phenotypes and the potential to adapt to different environments. Interestingly, recent evidence shows that several yeast species have evolved through genome duplication followed by gene loss (Dujon *et al.*, 2004), suggesting that *S. cerevisiae* is a degenerate tetraploid (Kellis *et al.*, 2004).



**Figure 44 – Karyotype analysis of *S. cerevisiae* cells expressing genetic code ambiguity.** Lanes 1 to 4, *S. cerevisiae* control cells; lanes 5 to 7, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-G<sub>33</sub>; and lanes 8 to 10, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub>. Chromosome number is indicated on the left. Box indicates chromosome alterations in CUG ambiguous cells, possibly the loss of either chromosome XIII or XVI.</sub></sub>

In order to detect and localize possible chromosome duplications and deletions occurring in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, a karyotype analysis was performed (Figure 44). Strikingly, there are not many alterations of the chromosome number and pattern in CUG ambiguous cells, with the exception of the region highlighted in red in Figure 44 that corresponds to the chromosomes XIII and XVI. These chromosomes are altered in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-G<sub>33</sub> (lanes 5 to 7) and T<sub>33</sub> (lanes 8 to 10), and the presence of only one band in some samples suggests that one of the chromosomes has been lost. When compared with the results from gene expression analysis (chapter III.1.), strains expressing the ser-tRNA<sub>CAG</sub>-G<sub>33</sub> apparently had lost one copy of the chromosome XVI and gained an extra chromosome XIII, which may correspond to the pattern displayed in Figure 44. However, these strains had additional chromosomal alterations that were not detected in the karyotype analysis, for example the duplication of chromosomes IX and XII and deletion of chromosomes X, XI, XIV and XV (chapter III.1.4.). Moreover, for the strains expressing the ser-tRNA<sub>CAG</sub>-T<sub>33</sub> the only alteration observed referred to the loss of chromosome XIV (chapter III.1.3.), again not represented on the karyotype.

Because the karyotype analysis did not reflect the bias observed in gene expression, a flow cytometry analysis was carried out for detection of ploidy alterations in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>. The increased DNA content of CUG mistranslating cells suggests that these cells are polyploid and tend to become aneuploid (Rita Rocha, unpublished results). Additionally, the analysis of the *C. albicans* karyotype (Isabel Miranda, unpublished results) showed major alterations on the chromosome number that included both deletions and duplications, and also possible rearrangements, which was confirmed by flow cytometry (Rita Rocha, unpublished results). Therefore, *C. albicans* and *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> have unstable genomes.

Genome instability can be used to generate genetic variability. Even at the cost of decreasing fitness of the organism, an unstable genome might be important for the adaptation and evolution of cells, namely under stress conditions. Indeed, it has been shown that chromosome rearrangements occur in yeast cells responding to a strong and permanent selective pressure (Dunham *et al.*, 2002). Thus, to understand the response of *S. cerevisiae* cells to CUG mistranslation, induced by the expression of the *C. albicans* ser-tRNA<sub>CAG</sub>, it is crucial to characterize the chromosome number alterations and rearrangements that occur in these cells. For this, microarray-based comparative genomic hybridization (CGH) that allows the detection of DNA copy number at the single-gene level and assesses if there is functional meaning for specific chromosome rearrangements in the context of mistranslation, may be the ideal experimental approach. As ploidy can regulate the expression of some genes, namely *FLO11* that is related with invasive growth (Galitski *et al.*, 1999), it would be interesting to address if ploidy changes correlate with stress resistance or virulence phenotypes. Therefore, the mechanisms generating genome instability in mistranslating cells will be further studied.

## **IV. General Discussion**

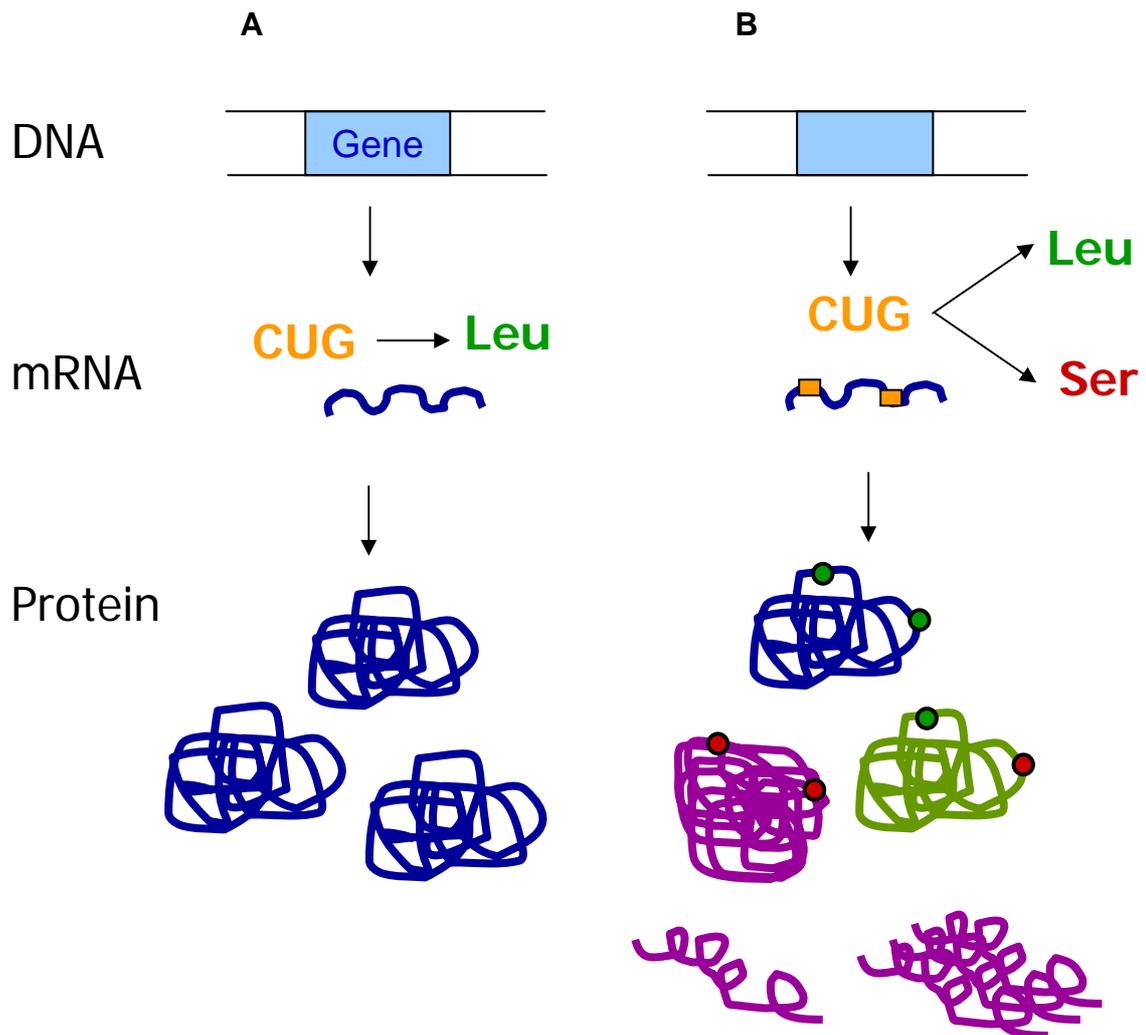
1. The cellular response to CUG ambiguity
2. Conclusions
3. Future work



## 1. The cellular response to CUG ambiguity

The aims of this work were to re-construct the early steps of the *C. albicans* genetic code alteration in *S. cerevisiae*, in order to shed new light on the impact that the reassignment of the leucine CUG codon to serine had on *C. albicans* evolution. To achieve this, the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub> were expressed in *S. cerevisiae*, and the effect of CUG ambiguity on cell physiology, gene expression and genome stability was characterized.

Expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* results in the mistranslation of the leucine CUG codon as serine. The consequences of mistranslation are represented in Figure 45. In cells expressing CUG ambiguity, whenever the ribosome reads a CUG codon during translation it will incorporate either a leucine or a serine into the nascent polypeptide chain (Figure 45). If a leucine is incorporated, the protein will be normal. However, if a serine is incorporated in response to a CUG codon, there is a strong possibility that the protein will not be able to achieve the correct structure and will not fold. The accumulation of unfolded or misfolded proteins decreases activity levels, leading to an haploinsufficiency phenotype. Also, misfolded proteins may form toxic aggregates. To counteract these deleterious effects, the cell triggers a constitutive stress response to increase the levels of molecular chaperones that assist protein folding and remove aberrant proteins. A similar response is observed under a stress like heat shock that also results in protein unfolding. However, unlike heat shock, which consists of a transient and external variation of the environmental conditions, CUG mistranslation imposes a permanent and internal stress on the cell. Therefore, the cellular response to mistranslation was studied at the molecular level through the analysis of the transcriptome and proteome of *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub>.



**Figure 45 – Scheme representing the consequences of genetic code ambiguity.** A – In a normal cell, translation of the mRNA of one gene results in synthesis of a functional corresponding protein. This situation is valid in cells expressing CUG ambiguity for the mRNAs that do not contain CUG codons. B – In CUG ambiguous cells, translation of a CUG codon might result in the production of the normal protein, if leucine is incorporated into the polypeptide chain. However, if serine is incorporated at the place of leucine, the protein will not achieve the native conformation and will unfold. The accumulation of misfolded proteins may be toxic to the cell.

### 1.1. CUG ambiguity alters gene expression

Transcriptome profiling of *S. cerevisiae* strains expressing the *C. albicans* ser-tRNA<sub>CAG</sub> T<sub>33</sub> and G<sub>33</sub> showed that the stress response is induced in these cells. The increased expression levels of genes that code for molecular chaperones indicates that there is a requirement to correct misfolded proteins resulting from the decoding of the CUG codon as serine instead of leucine. The simultaneous up regulation of stress-responsive genes suggests that CUG ambiguous cells trigger a stress response, which is correlated with the stress resistance phenotypes previously described (Santos *et al.*, 1996; Santos *et al.*, 1999). The observed higher expression levels of genes involved in carbohydrate metabolism, namely in the synthesis and degradation of the reserves trehalose and glycogen, supports the hypothesis of a tightly regulated energy metabolism. The accumulation of trehalose and glycogen points to the role that these storage carbohydrates play in the regulation of glucose availability to the cell and also in protein stabilization (Kaushik and Bhat, 2003). Genes related to the cell wall and transporters are also induced in response to CUG mistranslation, which could be linked to resistance to non-favourable environments. Taking into account that *C. albicans* is an opportunistic pathogen, and the cell wall is the structure that first interacts with the host, the remodelling of the cell wall triggered by CUG ambiguity could be an important mechanism during infection. Indeed, it has been shown that *S. cerevisiae* mutant strains that have altered cell wall composition and architecture are more virulent (Wheeler *et al.*, 2003). Protein synthesis is repressed and protein degradation is induced, as reported for yeast cells growing under adverse conditions. Genes belonging to the phosphate metabolism have increased expression levels as well, and together with the altered amino acid metabolism indicate that the expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* is a stressful situation that affects the regular functioning of the cell.

Proteome characterization of *S. cerevisiae* expressing the *C. albicans* ser-tRNA<sub>CAG</sub>-T<sub>33</sub> indicates that, as obtained for the transcriptome, the stress response is induced with increased expression levels of proteins associated with protein

folding, carbohydrate metabolism, and proteins responsive to general stress conditions. Also, protein synthesis and amino acid metabolism enzymes are repressed and protein degradation is up regulated. The higher expression of proteasome subunits and an enhanced proteasome activity in cells expressing CUG ambiguity are related to the production of altered proteins arising from mistranslation that have to be destroyed. Indeed, the increased expression levels of chaperones might be insufficient to correct the misfolded proteins and impair their aggregation. Therefore, an additional response is required to allow cell survival, such as the induction of protein degradation to eliminate abnormal proteins and prevent their accumulation on the stressed cell. The simultaneous induction of chaperones and the proteasome is thus fundamental for protein quality control. It is interesting that *C. albicans* has a high protease activity, which could reflect an increased amount of abnormal proteins normally produced in this yeast as a natural consequence of CUG ambiguity (Catarina Gomes, unpublished observations).

The proteome alterations observed in the strain expressing the ser-tRNA<sub>CAG-T<sub>33</sub></sub> are in agreement with previous studies described for the G<sub>33</sub> strain (Santos *et al.*, 1999), as both ser-tRNA<sub>CAG-T<sub>33</sub></sub> and ser-tRNA<sub>CAG-G<sub>33</sub></sub> increase expression of Hsp104p and members of the Hsp70 family of molecular chaperones. The proteins induced are different from the ones up regulated by heat shock, reinforcing that the response to genetic code ambiguity is unique and underlying the distinct nature of the stress imposed by the expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae*. However, the consequences of genetic code ambiguity are not restricted to proteome alterations and gene expression changes. CUG ambiguity will first impact protein synthesis with the production of aberrant proteins. If these proteins belong to DNA or RNA synthesis and repair pathways, the fidelity of such information processes might be compromised and the result is an increased global error rate. In this way, the consequences of genetic code ambiguity might also be reflected at the genome level. A similar mechanism of stress-induced mutagenesis has been described in *E. coli* (Bjedov *et al.*, 2003). These authors showed that stress conditions triggered the occurrence of mutations during DNA synthesis,

mainly due to impaired repair systems, and that selection favoured strains with increased mutation rates. The produced genetic diversity allowed cells to survive under stress and even explore novel environments. Moreover, increased mistranslation in *E. coli* also results in increased mutagenesis (Al Mamun *et al.*, 2002). In this work, an error-prone polymerase was shown to be the responsible for the mutations arisen during DNA replication. By analogy, the unstable genome of CUG mistranslating cells might be considered a generator of variability, promoting adaptation to novel ecological niches and the evolution of new phenotypes.

Cells expressing CUG ambiguity show extensive gene expression reprogramming, indicating that a genetic code change is a dramatic event that results in proteome destabilization, genome instability and global disruption of the normal functioning of the cell. However, the possibility that these mechanisms create phenotypic diversity suggests that cells expressing genetic code ambiguity have the potential to adapt and evolve under unfavourable growth conditions.

The comparative analysis of transcriptomics and proteomics data highlighted differences between the mRNA and protein content for some genes, suggesting that gene expression might be regulated at the translational level in CUG ambiguous cells. The evidence that there is translational control of gene expression under stress and the recent finding that eukaryotic gene expression can be directly regulated by metabolic enzymes rather than transcriptional factors (Hall *et al.*, 2004) make it feasible that additional levels of control can be found in mistranslating cells. Therefore, in order to elucidate the mechanisms of translational control of gene expression in mistranslating cells, the phosphorylation status of eIF2A and additional translation factors should be determined. The phosphorylation of eIF2A can justify the translational repression observed in CUG mistranslating cells, as a mechanism to protect the cell from its own aberrant proteins produced under stress. Nevertheless, stress-induced gene expression occurs and is also subject of control (Novoa *et al.*, 2003). In CUG ambiguous cells, specific mRNAs are selected for translation from the mRNA pool by yet unknown

pathways that could be controlled by the tRNA itself or might require phosphorylation and dephosphorylation of translational factors. As the phosphoproteome is different between CUG ambiguous and control cells and under stress conditions, it could be a good approach to detect candidate intervenients of the translational control of gene expression in response to CUG ambiguity.

## **1.2. The stress response triggered by CUG ambiguity**

The characterization of the cellular response of *S. cerevisiae* to the *C. albicans* ser-tRNA<sub>CAG</sub> has identified the genes and proteins responsible for tolerance of yeast cells to CUG ambiguity. It is also of paramount importance to elucidate how *C. albicans* responds to CUG ambiguity if one is to fully understand the evolution of CUG reassignment. *C. albicans* apparently lacks a general stress response similar to that observed in *S. cerevisiae* (Enjalbert *et al.*, 2003). In *C. albicans* there is a transient transcriptional induction of specific genes in response to various environmental challenges, however there is no common set of genes with increased expression levels to all stresses tested. Additionally, the *C. albicans* homologues of the *S. cerevisiae* genes involved in the general stress response show only a slight up regulation under the same stress conditions (Enjalbert *et al.*, 2003). Interestingly, this situation is mimicked in *S. cerevisiae* expressing the *C. albicans* ser-tRNA<sub>CAG</sub>. After heat shock, the transient induction of the stress responsive targets is observed, however, when these cells are grown under the sub-optimal temperature of 37°C the induction of most of the stress response genes is not detected. The main reason for this behaviour is that *S. cerevisiae* ambiguous cells already express higher levels of the stress proteins necessary to face the unfavourable growth temperature. Indeed, the presence of the ser-tRNA<sub>CAG</sub> in *S. cerevisiae* increases the expression of stress proteins needed to cope with adverse environments, thus pre-adapting cells to tolerate such conditions. Similarly, in *C. albicans* the basal expression levels of stress-protective proteins might be high in response to CUG mistranslation, and hence the apparent

absence of a general stress response in *C. albicans*. The presence of the ser-tRNA<sub>CAG</sub> in *C. albicans* for millions of years provided enough evolutionary time for the stress response to specialize and diverge from that observed in *S. cerevisiae*. Indeed, the reduced number of STREs in the promoters of *C. albicans* stress responsive genes points to distinct control mechanisms of the stress response in both yeast species. There are candidate homologues of the Msn2p and Msn4p transcription factors, which regulate STRE-driven gene expression, in *C. albicans* (Enjalbert *et al.*, 2003). However, their role as general stress transcription factors is yet to be determined, and it is possible that their function could be related to a different form of stress, namely the internal and permanent pressure on gene expression elicited by the ambiguity of the CUG codon that was preserved to the present day.

### **1.3. The role of molecular chaperones and the proteasome**

*S. cerevisiae* cells expressing genetic code ambiguity showed induction of the molecular chaperones and increased proteasome activity, suggesting that these are the main responses of the cell against the misfolded and aggregated proteins resulting from CUG mistranslation. Indeed, the roles of molecular chaperones and the proteasome are tightly related, as both recognize non-native proteins and prevent their accumulation, thus contributing to protein homeostasis in the cell. Interestingly, besides protein folding and the stress response, the molecular chaperone Hsp90p is fundamental for the assembly and maintenance of the 26S proteasome (Imai *et al.*, 2003). Furthermore, the induction of the heat shock proteins during heat shock, essential for cell survival under stress, could be replaced by an increase of the ubiquitin-dependent protein degradation (Friant *et al.*, 2003), reinforcing the connection between the two pathways. Therefore, it is fundamental to assess the distinct contributions of the two mechanisms to the cellular response to CUG mistranslation. For this, the response to CUG mistranslation should be studied in *S. cerevisiae* strains deleted of crucial genes, namely Hsp104p, which is required for rescuing misfolded and aggregated

proteins, and non-essential proteasome subunits, such as Rpn10p, involved in protein degradation. Additionally, if an excess of aberrant proteins is being produced that needs to be removed from mistranslating cells, what would be the impact of blocking such “waste disposal” mechanisms with proteasome inhibitors? The ubistatins might be quite useful to understand the processes of protein degradation operating in mistranslating cells, as they inhibit ubiquitin-dependent proteolysis (Verma *et al.*, 2004). Thus, they could be important to uncover alternative ubiquitin-independent pathways that might be used to target mistranslated proteins for degradation by the proteasome.

The molecular chaperones Hsp90p and Hsp70p also play a role in protein degradation, besides protein folding (Hohfeld *et al.*, 2001). The mechanism that enables the choice for either one of these processes is not known, but it is determined by the interaction of additional proteins with the complex formed by Hsp90p and Hsp70p. Sti1p is a co-chaperone that binds both Hsp90p and Hsp70p as part of the folding complex, and this protein is induced in CUG mistranslating cells. This result indicates that Sti1p might be promoting an increase in protein folding by activating Hsp70p (Wegele *et al.*, 2003). However, since protein degradation is also increased in these cells, there must be another co-chaperone that is able to switch the activity of the Hsp90p-Hsp70p complex from folding to degradation, for example by redirecting the unfolded substrates to the proteasome. Thus, protein quality control in CUG ambiguous cells is coordinated by factors that ought to be revealed. In order to determine which co-chaperones associate with the complexes formed by Hsp90p and Hsp70p, a possible experiment would be co-immunoprecipitation using antibodies against these heat shock proteins, followed by fractionation and identification of the interacting proteins by mass spectrometry.

#### 1.4. Is the *C. albicans* ser-tRNA<sub>CAG</sub> a generator of phenotypic diversity?

Mutations that arise under stress conditions are a source of genetic diversity that can be used for evolution (Bjedov *et al.*, 2003), and epigenetic mechanisms can also confer growth advantage in fluctuating environments (True *et al.*, 2004). By revealing hidden genetic variation, such epigenetic mechanisms enable phenotypic plasticity, with consequences for both survival of the organism and evolution of new traits. Work on Hsp90p (Rutherford and Lindquist, 1998; Queitsch *et al.*, 2002) and [PSI<sup>+</sup>] (True and Lindquist, 2000; True *et al.*, 2004) show that, although acting by different mechanisms, they both induce alterations in protein folding that influence the protein content of the cell.

Hsp90p interacts with unstable signalling molecules, keeping these proteins ready for activation by conformational changes that stabilize them and that are related with signal transduction. Blocking of Hsp90p function in *Drosophila* (Rutherford and Lindquist, 1998) and *Arabidopsis thaliana* (Queitsch *et al.*, 2002) exposes phenotypic diversity, indicating that Hsp90p buffers pre-existent genetic variation. This suggests that under stress conditions Hsp90p is redirected from its normal targets towards damaged proteins and variation is then revealed, allowing selection to act on such expressed traits to evolve distinct phenotypes that will be eventually produced even when Hsp90p is fully functional.

[PSI<sup>+</sup>] is a prion that arises from a conformational change of the translation termination factor Sup35p (Paushkin *et al.*, 1996). The  $\psi$  prion exists in an aggregated form of Sup35p causing reduced translation termination efficiency (Eaglestone *et al.*, 1999). The consequent read through of stop codons gives rise to the production of longer proteins (C-terminally extended), which in turn produce genetic variation (True *et al.*, 2004; True and Lindquist, 2000). The phenotypes related to [PSI<sup>+</sup>] acquisition involve stress-resistance and morphological variation, and although dependent on the genetic background of the strain used, these traits are inherited by the progeny and can even be fixed in the population (True and Lindquist, 2000; True *et al.*, 2004). Therefore, exposure of hidden genetic variation

by epigenetic mechanisms has an important role in producing phenotypes from genotypes and has significant impact on evolution.

The *C. albicans* ser-tRNA<sub>CAG</sub> could provide such an epigenetic mechanism to produce genetic variation. In this situation, cells that share the same genome will contain different statistical proteomes, since CUG ambiguity can result in a myriad of distinct proteins in a random manner, depending on the insertion of either serine or leucine. The observations that *C. albicans* can manipulate the levels of the ser-tRNA<sub>CAG</sub> leucylation and serylation *in vivo* according to the physiological state of the cell and environmental conditions (Catarina Gomes, unpublished results) support this hypothesis. Additionally, *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> are more stress tolerant than wild-type cells, thus, the ser-tRNA<sub>CAG</sub> provides growth advantage in unfavourable environments (Santos *et al.*, 1996; Santos *et al.*, 1999). Therefore, the *C. albicans* ser-tRNA<sub>CAG</sub> could act as a trigger to unveil pre-existent variation by sequestering the molecular chaperones that assist the folding of aberrant proteins, and at the same time provide a mechanism to originate some proteins with novel properties. One interesting experiment to address this possibility would be to reduce Hsp82p (the Hsp90p homologue) activity in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>, either by mutation or growth in the presence of inhibitors such as geldanamycin and radicicol (Queitsch *et al.*, 2002). This would increase phenotypic diversity, as the buffering capacity of Hsp90p would be impaired. The less functional Hsp90p would still be challenged with the requirement to fold damaged proteins arising from CUG mistranslation, but since there would be less chaperone available the hidden phenotypic variability is likely to be revealed. An additional approach is to grow *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> under additional stress conditions for several generations, and then allow them to lose the ser-tRNA<sub>CAG</sub>. After, cells that have lost the ser-tRNA<sub>CAG</sub> would be grown in the stress conditions previously determined, in order to check if the stress-resistance phenotypes can be fixed in the population. If the variation generated by CUG ambiguity is selected upon, the traits will be kept regardless of the presence of the tRNA.

### 1.5. Is there a role for CUG ambiguity during *C. albicans* infection?

Quorum sensing is mainly used in bacteria to monitor the cell density of a population, in order to regulate the production of bioluminescence, virulence factors or biofilm formation, among other physiological processes. This signalling mechanism is based on diffusible molecules that influence the cells' behaviour (Dong and Zhang, 2005). Similarly, it has been shown that volatile ammonia can control yeast colony growth (Palkova *et al.*, 1997; Palkova and Forstova, 2000; Palkova *et al.*, 2002). It would be interesting to determine if *C. albicans* infection depends on a quorum sensing-like mechanism that controls CUG ambiguity, and identify signals could be transmitted between cells under ambiguity conditions. One possible candidate of such cell-to-cell communication could be ammonia, as it is produced during the nicotinamidase activity of Pnc1p that is induced about 30 fold in *S. cerevisiae* expressing the *C. albicans* ser-tRNA<sub>CAG</sub>. Additionally, because CUG ambiguity increases in *C. albicans* under growth at 37 °C and acidic pH (Catarina Gomes, unpublished results), it would be particularly relevant to establish the role of genetic code ambiguity in *C. albicans* during infection, which could allow the development of novel anti-fungal targets.

## 2. Conclusions

The expression of the *C. albicans* ser-tRNA<sub>CAG</sub> in *S. cerevisiae* cells showed that CUG ambiguity constitutively alters the expression of mRNAs and proteins. The ambiguity introduced in *S. cerevisiae* is low, due to competition between the *C. albicans* ser-tRNA<sub>CAG</sub>, which translates the CUG codon as serine, and the endogenous tRNAs that decode it as leucine. Moreover, the CUG codon is one of the least used leucine codons in *S. cerevisiae* proteins, resulting in the alteration of only part of proteome. Even though, CUG ambiguity drives extensive gene expression reprogramming and regulation of gene expression at the translational level. These results indicate that genetic code ambiguity consists of an internal and permanent stress to the cell, causing proteome disruption and genomic instability.

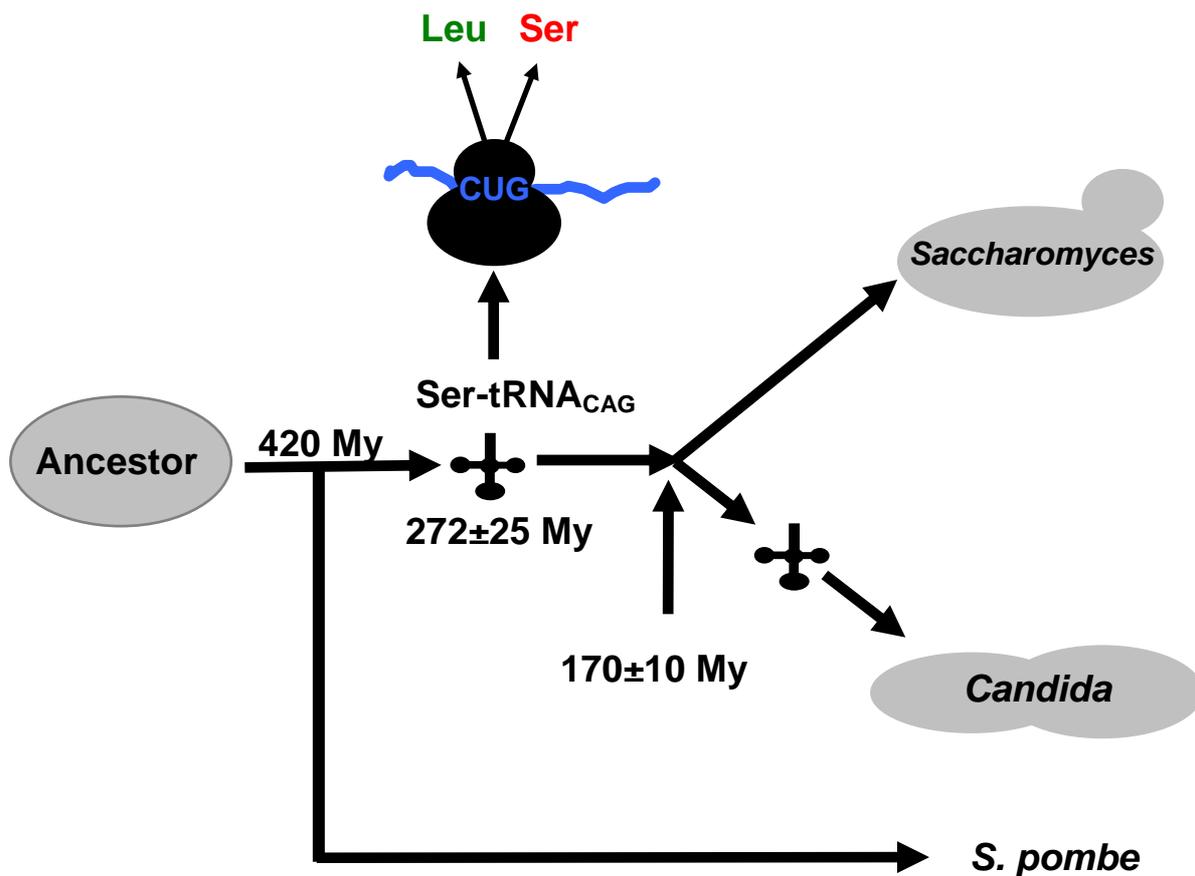
However, the stress imposed by CUG mistranslation does not impair cells to respond to additional stress. Indeed, CUG ambiguous cells were able to induce stress responsive proteins in response to heat-shock, namely heat-shock proteins and carbohydrate metabolism enzymes. Most remarkably, the stress response triggered by mRNA decoding ambiguity, by increasing the expression levels of the molecular chaperones and other stress protective proteins, pre-adapts cells to tolerate adverse growth environments, as shown by the increased stress resistance of CUG ambiguous cells (Santos *et al.*, 1996; Santos *et al.*, 1999; this work). Therefore, although genetic code ambiguity is toxic to the cell, due to the accumulation of aberrant proteins that are produced by mRNA mistranslation, these cells have selective advantages under stress. Thus, organisms that have altered the genetic code through ambiguous codon decoding have to adapt to global proteome and genome destabilization, but simultaneously the response to this decreased decoding fidelity provides variability between the cells of the same population. Such phenotypic diversity allows not only cell survival, but also the exploration of novel ecological niches.

Heterogeneity of the cellular stress response among individuals of the same population has been described before in *S. cerevisiae* (Attfield *et al.*, 2001), and such variability between cells provides a fertile ground for selection to act upon. Indeed, altering gene expression levels can drive evolution through acquisition of novel phenotypes. Instead of accumulating mutations on the genes themselves, which would alter the properties of the proteins, mutations that occur rather on the regulatory sequences that control gene expression will result in altered levels of the expressed genes. Increasing or decreasing the corresponding gene products alters the protein content of different individuals of the same population and allow new characters to arise (Oleksiak *et al.*, 2002).

This study shows that yeast cells are able to tolerate genetic code ambiguity. Realizing that eukaryotic cells can cope with mistranslation opens the possibility to alter or expand the genetic code, both to create novel proteins and to understand the origins and evolution of life. Recent work indicates that it is possible to expand the genetic code in *E. coli* (Döring *et al.*, 2001; Wang *et al.*, 2001; Mehl *et al.*, 2003), bacteriophages (Bacher *et al.*, 2003) and yeast (Chin *et al.*, 2003) to accept nonnatural amino acids and even to introduce the 22<sup>nd</sup> amino acid, pyrrolysine, into *E. coli* proteins (Blight *et al.*, 2004), supporting the notion that organisms are amenable to genetic code manipulations (discussed in chapter 1.2.3.).

Despite having unstable proteomes and genomes, *S. cerevisiae* cells that misread the leucine CUG codon as serine do not lose viability in stationary phase. As many tumours show genomic instability and genomes tend to be unstable with age (McMurray and Gottschling, 2003) the study of mistranslation in *S. cerevisiae* could have implications for the understanding of aging mechanisms and cancer. Additionally, genome duplication has been proposed as an important mechanism for the evolution of *S. cerevisiae* and other yeast species (Dujon *et al.*, 2004), whether occurring as whole genome duplication (Kellis *et al.*, 2004) or independent local duplication events (Kozsul *et al.*, 2004). The recent evidence for whole genome duplication in *S. cerevisiae* suggests this yeast is a degenerate tetraploid (Kellis *et al.*, 2004). Indeed, genome duplication followed by gene loss

and specialization offers the potential for adaptation to new environments as it allows new functions to arise. Flow cytometry analysis of *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub> showed that CUG ambiguous cells have higher DNA content, with the formation of polyploid and aneuploid cells (Rita Rocha, unpublished results). Therefore, one has to ask whether the introduction of an ambiguous tRNA triggered the genome duplication on the ancestor of *S. cerevisiae* and *C. albicans*? The polyploidy originated with this event would have brought genome instability, but also the potential to create new functions and new phenotypes. The possibility to adapt and colonize different ecological niches would result in the divergence between cells, and novel species could evolve. Thus, this work has implications to the understanding of evolution of yeast species, as the ser-tRNA<sub>CAG</sub> was a major driving force of the evolution of the genus *Candida*. The appearance of the ser-tRNA<sub>CAG</sub> about 272 million years ago generated ambiguity at the CUG codon (Figure 46). This genetic code ambiguity had a major impact on the ancestor's genome and proteome, as highlighted by the alterations described in this work in *S. cerevisiae* expressing the *C. albicans* ser-tRNA<sub>CAG</sub>. In this experimental model, the attempt to rewind evolutionary time, to the point where a cell is challenged with a tRNA mutation that introduces codon ambiguity, showed that this was a dramatic but crucial event for the evolution of *Candida* spp. Indeed, *S. cerevisiae* ambiguous cells show induction of a constitutive stress response, resulting in increased stress resistance, remodelling of the cell wall and general metabolism, and also an increase in ploidy. Possibly, higher levels of ambiguity and enough evolutionary distance triggered the expression of distinct phenotypes that drove the divergence of the genera *Candida* and *Saccharomyces* in a 272 million year time-scale. CUG mistranslation could have provided an alternative mechanism to generate variability among the population and allow cell survival under stress, and ultimately driven evolution of new species. Genetic code ambiguity can have, therefore, an active evolutionary role instead of representing a mere by-product of biased genome GC content.



**Figure 46 – The *ser-tRNA<sub>CAG</sub>* was a major driving force of the evolution of the genus *Candida*.** The standard leucine CUG codon is translated as serine by a novel tRNA, which appeared approximately 272 million years (My) ago (Massey *et al.*, 2003). The divergence of the genera *Candida* and *Saccharomyces* occurred later, at about 170 My, and the *ser-tRNA<sub>CAG</sub>* was kept in the lineage that originated the genus *Candida* but was lost in the lineage leading to the genus *Saccharomyces*.

### 3. Future work

Characterization of the cellular response to a genetic code alteration showed that CUG mistranslating cells have increased expression of molecular chaperones and proteasome activity, indicating their key role in the tolerance to genetic code ambiguity. Therefore, what happens to CUG ambiguous cells if either the activity of the molecular chaperones or the proteasome is impaired? To answer this question, selected genes encoding molecular chaperones and proteasome subunits will be disrupted using homologous recombination, and the resulting knockout strains will be transformed with the *C. albicans* ser-tRNA<sub>CAG</sub>. The fitness of these cells will be studied by comparing their growth rate, viability and stress tolerance with control cells. Additionally, gene expression and ploidy changes in the knockout strains will be characterized using DNA microarrays.

The possibility that CUG ambiguous cells form protein and / or RNA aggregates will also be addressed, using immunofluorescence methodologies. This will be important to elucidate the pathways of protein accumulation and degradation, as well as the mechanisms of translational control of gene expression in CUG mistranslating cells.

CUG ambiguous cells have increased life span, despite having unstable genome and proteome. This raises the questions of “how does permanent proteome instability promote genome rearrangements?”, and “what are the long term consequences of extended longevity with unstable genomes and proteomes?”. To answer these questions, the role of the viability-promoting genes will be studied in CUG mistranslating cells, using a similar strategy as described above. *PNC1* and *SIR2* are not required for viability in yeast (Giaever *et al.*, 2002), but what would be the impact of their deletion on *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG</sub>? Would these cells lose viability? Would the deletion of *PNC1* or *SIR2* be lethal in these cells? The elucidation of the mechanisms that generate genome instability in long-lived cells could also shed new light on the pathways that link aging and cancer.

## **V. References**



## References

- Agris, P.F. (2004). Decoding the genome: a modified view. *Nucleic Acids Res.* **32**: 223-238.
- Aguilaniu, H., Gustafsson, L., Rigoulet, M., and Nystrom, T. (2003). Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. *Science* **299**: 1751-1753.
- Al Mamun, A.A., Marians, K.J., and Humayun, M.Z. (2002). DNA polymerase III from *Escherichia coli* cells expressing mutA mistranslator tRNA is error-prone. *J. Biol. Chem.* **277**: 46319-46327.
- Almaguer, C., Cheng, W., Nolder, C., and Patton-Vogt, J. (2004). Glycerophosphoinositol, a novel phosphate source whose transport is regulated by multiple factors in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**: 31937-31942.
- Anderson, R.M., Bitterman, K.J., Wood, J.G., Medvedik, O., and Sinclair, D.A. (2003). Nicotinamide and PNC1 govern lifespan extension by calorie restriction in *Saccharomyces cerevisiae*. *Nature* **423**: 181-185.
- Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A* **100**: 3889-3894.
- Arnez, J.G. and Moras, D. (1997). Structural and functional considerations of the aminoacylation reaction. *Trends Biochem. Sci.* **22**: 211-216.
- Attfield, P.V., Choi, H.Y., Veal, D.A., and Bell, P.J. (2001). Heterogeneity of stress gene expression and stress resistance among individual cells of *Saccharomyces cerevisiae*. *Mol. Microbiol.* **40**: 1000-1008.
- Auesukaree, C., Homma, T., Tochio, H., Shirakawa, M., Kaneko, Y., and Harashima, S. (2004). Intracellular phosphate serves as a signal for the

- regulation of the PHO pathway in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**: 17289-17294.
- Bacher, J.M., Bull, J.J., and Ellington, A.D. (2003). Evolution of phage with chemically ambiguous proteomes. *BMC. Evol. Biol.* **3**: 24-
- Bacher, J.M., Crecy-Lagard, V., and Schimmel, P.R. (2005). Inhibited cell growth and protein functional changes from an editing-defective tRNA synthetase. *Proc. Natl. Acad. Sci. U. S. A* **102**: 1697-1701.
- Bacher, J.M., Hughes, R.A., Wong, J.T.F., and Ellington, A.D. (2004). Evolving new genetic codes. *Trends in Ecology and Evolution* **19**: 69-75.
- Bajorek, M., Finley, D., and Glickman, M.H. (2003). Proteasome disassembly and downregulation is correlated with viability during stationary phase. *Curr. Biol.* **13**: 1140-1144.
- Ballesteros, M., Fredriksson, A., Henriksson, J., and Nystrom, T. (2001). Bacterial senescence: protein oxidation in non-proliferating cells is dictated by the accuracy of the ribosomes. *EMBO J.* **20**: 5280-5289.
- Baron, C. and Bock, A. (1995). The Selenocysteine-Inserting tRNA Species: Structure and Function. in *tRNA structure, Biosynthesis and function*. (eds Soll, D. and RajBhandary, U. L.): 529-544. American Society for Microbiology, Washington, D.C.
- Barrell, B.G., Bankier, A.T., and Drouin, J. (1979). A different genetic code in human mitochondria. *Nature* **282**: 189-194.
- Beier, H. and Grimm, M. (2001). Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res.* **29**: 4767-4782.
- Bein, M., Schaller, M., and Korting, H.C. (2002). The secreted aspartic proteinases as a new target in the therapy of candidiasis. *Curr. Drug Targets.* **3**: 351-357.

- Benaroudj, N., Tarcsa, E., Cascio, P., and Goldberg, A.L. (2001). The unfolding of substrates and ubiquitin-independent protein degradation by proteasomes. *Biochimie* **83**: 311-318.
- Bence, N.F., Sampat, R.M., and Kopito, R.R. (2001). Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292**: 1552-1555.
- Bini, L., Liberatori, S., Magi, B., Marzocchi, B., Raggiaschi, R., and Pallini, V. (2000). Protein Blotting and Immunoblotting. in *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods* (eds Rabilloud, T.): 127-141. Springer Verlag,
- Bjedov, I., Tenailon, O., Gerard, B., Souza, V., Denamur, E., Radman, M., Taddei, F., and Matic, I. (2003). Stress-induced mutagenesis in bacteria. *Science* **300**: 1404-1409.
- Björk, G.R. (1995a). Biosynthesis and Function of Modified Nucleosides. in *tRNA Structure, Biosynthesis and Function*. (eds Söll, D. and RajBhandary, U.): 165-205. American Society for Microbiology, Washington, D.C.
- Björk, G.R. (1995b). Genetic dissection of synthesis and function of modified nucleosides in bacterial transfer RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **50**: 263-338.
- Blight, S.K., Larue, R.C., Mahapatra, A., Longstaff, D.G., Chang, E., Zhao, G., Kang, P.T., Green-Church, K.B., Chan, M.K., and Krzycki, J.A. (2004). Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* **431**: 333-335.
- Bonner, J.J., Carlson, T., Fackenthal, D.L., Paddock, D., Storey, K., and Lea, K. (2000). Complex regulation of the yeast heat shock transcription factor. *Mol. Biol. Cell* **11**: 1739-1751.
- Boucherie, H., Dujardin, G., Kermorgant, M., Monribot, C., Slonimski, P., and Perrot, M. (1995). Two-dimensional protein map of *Saccharomyces cerevisiae*: construction of a gene-protein index. *Yeast* **11**: 601-613.

- Boy-Marcotte, E., Lagniel, G., Perrot, M., Bussereau, F., Boudsocq, A., Jacquet, M., and Labarre, J. (1999). The heat shock response in yeast: differential regulations and contributions of the Msn2p/Msn4p and Hsf1p regulons. *Mol. Microbiol.* **33**: 274-283.
- Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248-254.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29**: 365-371.
- Breitschopf, K. and Gross, H.J. (1994). The exchange of the discriminator base A73 for G is alone sufficient to convert human tRNA-Leu into a serine-acceptor in vitro. *EMBO J.* **13**: 3166-3169.
- Brown, P.O. and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21**: 33-37.
- Castresana, J., Feldmaier-Fuchs, G., and Paabo, S. (1998). Codon reassignment and amino acid composition in hemichordate mitochondria. *Proc. Natl. Acad. Sci. U. S. A* **95**: 3703-3707.
- Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S., and Young, R.A. (2001). Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* **12**: 323-337.
- Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W., and Harrison, P.R. (1986). The structure of the mouse glutathione peroxidase gene: the

- selenocysteine in the active site is encoded by the 'termination' codon, TGA. *EMBO J.* **5**: 1221-1227.
- Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z., and Schultz, P.G. (2003). An expanded eukaryotic genetic code. *Science* **301**: 964-967.
- Chu, W.S., Magee, B.B., and Magee, P.T. (1993). Construction of an Sfil macrorestriction map of the *Candida albicans* genome. *J. Bacteriol.* **175**: 6637-6651.
- Clemens, M.J. (1996). Protein kinases that phosphorylate eIF2 and eIF2B, and their role in eukaryotic cell translational control. in *Translational Control* (eds Hershey, J. W. B., Mathews, M. B, and Sonenberg, N.): 139-172. Cold Spring Harbor Laboratory Press,
- Coleman, S.T., Fang, T.K., Rovinsky, S.A., Turano, F.J., and Moye-Rowley, W.S. (2001). Expression of a glutamate decarboxylase homologue is required for normal oxidative stress tolerance in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**: 244-250.
- Commans, S. and Bock, A. (1999). Selenocysteine inserting tRNAs: an overview. *FEMS Microbiol. Rev.* **23**: 335-351.
- Conaway, R.C., Brower, C.S., and Conaway, J.W. (2002). Emerging roles of ubiquitin in transcription regulation. *Science* **296**: 1254-1258.
- Costa, V.M., Amorim, M.A., Quintanilha, A., and Moradas-Ferreira, P. (2002). Hydrogen peroxide-induced carbonylation of key metabolic enzymes in *Saccharomyces cerevisiae*: the involvement of the oxidative stress response regulators Yap1 and Skn7. *Free Radic. Biol. Med.* **33**: 1507-1515.
- Cox, J.S. and Walter, P. (1996). A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. *Cell* **87**: 391-404.
- Crick, F.H. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* **12**: 138-163.

- Crick, F.H. (1968). The origin of the genetic code. *J. Mol. Biol.* **38**: 367-379.
- Cropp, T.A. and Schultz, P.G. (2004). An expanding genetic code. *Trends Genet.* **20**: 625-630.
- Crothers, D.M., Seno, T., and Soll, G. (1972). Is there a discriminator site in transfer RNA? *Proc. Natl. Acad. Sci. U. S. A* **69**: 3063-3067.
- Curnow, A.W., Hong, K., Yuan, R., Kim, S., Martins, O., Winkler, W., Henkin, T.M., and Soll, D. (1997). Glu-tRNA<sup>Gln</sup> amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl. Acad. Sci. U. S. A* **94**: 11819-11826.
- Curnow, A.W., Ibba, M., and Soll, D. (1996). tRNA-dependent asparagine formation. *Nature* **382**: 589-590.
- Davies, K.J. (2001). Degradation of oxidized proteins by the 20S proteasome. *Biochimie* **83**: 301-310.
- Day, D.A. and Tuite, M.F. (1998). Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *J. Endocrinol.* **157**: 361-371.
- Delarue, M. and Moras, D. (1993). The aminoacyl-tRNA synthetase family: modules at work. *Bioessays* **15**: 675-687.
- Demasi, M., Silva, G.M., and Netto, L.E. (2003). 20 S proteasome from *Saccharomyces cerevisiae* is responsive to redox modifications and is S-glutathionylated. *J. Biol. Chem.* **278**: 679-685.
- DiDomenico, B.J., Bugaisky, G.E., and Lindquist, S. (1982a). Heat shock and recovery are mediated by different translational mechanisms. *Proc. Natl. Acad. Sci. U. S. A* **79**: 6181-6185.
- DiDomenico, B.J., Bugaisky, G.E., and Lindquist, S. (1982b). The heat shock response is self-regulated at both the transcriptional and posttranscriptional levels. *Cell* **31**: 593-603.

- Dietrich, A., Romby, P., Marechal-Drouard, L., Guillemaut, P., and Giege, R. (1990). Solution conformation of several free tRNA<sup>Leu</sup> species from bean, yeast and *Escherichia coli* and interaction of these tRNAs with bean cytoplasmic Leucyl-tRNA synthetase. A phosphate alkylation study with ethylnitrosourea. *Nucleic Acids Res.* **18**: 2589-2597.
- Dirheimer, G., Keith, G., Dumas, P., and Westhof, E. (1995). Primary, Secondary, and Tertiary Structures of tRNAs. in *tRNA structure, Biosynthesis and function*. (eds Soll, D. and RajBhandary, U.): 93-126. American Society for Microbiology, Washington, D.C.
- Dock-Bregeon, A.C., Westhof, E., Giege, R., and Moras, D. (1989). Solution structure of a tRNA with a large variable region: yeast tRNA<sup>Ser</sup>. *J. Mol. Biol.* **206**: 707-722.
- Dong, Y.H. and Zhang, L.H. (2005). Quorum sensing and quorum-quenching enzymes. *J. Microbiol.* **43 Spec No**: 101-109.
- Döring, V., Mootz, H.D., Nangle, L.A., Hendrickson, T.L., Crécy-Lagard, V., Schimmel, P., and Marlière, P. (2001). Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway. *Science* **292**: 501-504.
- Driscoll, D.M. and Copeland, P.R. (2003). Mechanism and regulation of selenoprotein synthesis. *Annu. Rev. Nutr.* **23**: 17-40.
- Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., de Montigny, J., Marck, C., Neuveglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Bleykasten, C., Boisrame, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.F., Straub, M.L.,

- Suleau, A., Swennen, D., Tekaiia, F., Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., and Souciet, J.L. (2004). Genome evolution in yeasts. *Nature* **430**: 35-44.
- Dukan, S., Farewell, A., Ballesteros, M., Taddei, F., Radman, M., and Nystrom, T. (2000). Protein oxidation in response to increased transcriptional or translational errors. *Proc. Natl. Acad. Sci. U. S. A* **97**: 5746-5749.
- Duncan, R.F. (1996). Translational control during heat shock. in *Translational control* (eds Hershey, J. W. B., Mathews, M. B, and Sonenberg, N.): 271-294. Cold Spring Harbor Laboratory Press,
- Dunham, M.J., Badrane, H., Ferea, T., Adams, J., Brown, P.O., Rosenzweig, F., and Botstein, D. (2002). Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A* **99**: 16144-16149.
- Eaglestone, S.S., Cox, B.S., and Tuite, M.F. (1999). Translation termination efficiency can be regulated in *Saccharomyces cerevisiae* by environmental stress through a prion-mediated mechanism. *EMBO J.* **18**: 1974-1981.
- Edelmann, P. and Gallant, J. (1977). Mistranslation in *E. coli*. *Cell* **10**: 131-137.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A* **95**: 14863-14868.
- Ellis, C.D., Wang, F., MacDiarmid, C.W., Clark, S., Lyons, T., and Eide, D.J. (2004). Zinc and the Msc2 zinc transporter protein are required for endoplasmic reticulum function. *J. Cell Biol.* **166**: 325-335.
- Enjalbert, B., Nantel, A., and Whiteway, M. (2003). Stress-induced gene expression in *Candida albicans*: absence of a general stress response. *Mol. Biol. Cell* **14**: 1460-1467.

- Feng, Y.X., Copeland, T.D., Oroszlan, S., Rein, A., and Levin, J.G. (1990). Identification of amino acids inserted during suppression of UAA and UGA termination codons at the gag-pol junction of Moloney murine leukemia virus. *Proc. Natl. Acad. Sci. U. S. A* **87**: 8860-8863.
- Fenske, C., Palm, G.J., and Hinrichs, W. (2003). How unique is the genetic code? *Angew. Chem. Int. Ed Engl.* **42**: 606-610.
- Forchhammer, K. and Bock, A. (1991). Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. *J. Biol. Chem.* **266**: 6324-6328.
- Friant, S., Meier, K.D., and Riezman, H. (2003). Increased ubiquitin-dependent degradation can replace the essential requirement for heat shock protein induction. *EMBO J.* **22**: 3783-3791.
- Galitski, T., Saldanha, A.J., Styles, C.A., Lander, E.S., and Fink, G.R. (1999). Ploidy regulation of gene expression. *Science* **285**: 251-254.
- Gallo, C.M., Smith, D.L., Jr., and Smith, J.S. (2004). Nicotinamide clearance by Pnc1 directly regulates Sir2-mediated silencing and longevity. *Mol. Cell Biol.* **24**: 1301-1312.
- Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**: 4241-4257.
- Gasch, A.P. and Werner-Washburne, M. (2002). The genomics of yeast responses to environmental stress and starvation. *Funct. Integr. Genomics* **2**: 181-192.
- Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A.P., Astromoff, A., El Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J.H.,

- Hempel, S., Herman, Z., Jaramillo, D.F., Kelly, D.E., Kelly, S.L., Kotter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S.L., Revuelta, J.L., Roberts, C.J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K., Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C.Y., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W., and Johnston, M. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387-391.
- Gietz, R.D. and Woods, R.A. (1994). High efficiency transformation with lithium acetate. in *Molecular Genetics of Yeast: A Practical Approach* (eds Johnston, J. R.): 121-134. IRL Press, Oxford
- Giots, F., Donaton, M.C., and Thevelein, J.M. (2003). Inorganic phosphate is sensed by specific phosphate carriers and acts in concert with glucose as a nutrient signal for activation of the protein kinase A pathway in the yeast *Saccharomyces cerevisiae*. *Mol. Microbiol.* **47**: 1163-1181.
- Gladyshev, V.N. and Kryukov, G.V. (2001). Evolution of selenocysteine-containing proteins: significance of identification and functional characterization of selenoproteins. *Biofactors* **14**: 87-92.
- Glover, J.R. and Lindquist, S. (1998). Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. *Cell* **94**: 73-82.
- Godon, C., Lagniel, G., Lee, J., Buhler, J.M., Kieffer, S., Perrot, M., Boucherie, H., Toledano, M.B., and Labarre, J. (1998). The H<sub>2</sub>O<sub>2</sub> stimulon in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**: 22480-22489.
- Goffeau, A., Park, J., Paulsen, I.T., Jonniaux, J.L., Dinh, T., Mordant, P., and Saier, M.H., Jr. (1997). Multidrug-resistant transport proteins in yeast:

- complete inventory and phylogenetic characterization of yeast open reading frames with the major facilitator superfamily. *Yeast* **13**: 43-54.
- Gounalaki, N. and Thireos, G. (1994). Yap1p, a yeast transcriptional activator that mediates multidrug resistance, regulates the metabolic stress response. *EMBO J.* **13**: 4036-4041.
- Griffin, T.J., Gygi, S.P., Ideker, T., Rist, B., Eng, J., Hood, L., and Aebersold, R. (2002). Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol. Cell Proteomics.* **1**: 323-333.
- Grune, T., Blasig, I.E., Sitte, N., Roloff, B., Haseloff, R., and Davies, K.J. (1998). Peroxynitrite increases the degradation of aconitase and other cellular proteins by proteasome. *J. Biol. Chem.* **273**: 10857-10862.
- Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **19**: 1720-1730.
- Hall, D.A., Zhu, H., Zhu, X., Royce, T., Gerstein, M., and Snyder, M. (2004). Regulation of gene expression by a metabolic enzyme. *Science* **306**: 482-484.
- Hao, B., Gong, W., Ferguson, T.K., James, C.M., Krzycki, J.A., and Chan, M.K. (2002). A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296**: 1462-1466.
- Hatfield, D.L. and Gladyshev, V.N. (2002). How selenium has altered our understanding of the genetic code. *Mol. Cell. Biol.* **22**: 3565-3576.
- Hellinga, H.W., Wynn, R., and Richards, F.M. (1992). The hydrophobic core of *Escherichia coli* thioredoxin shows a high tolerance to nonconservative single amino acid substitutions. *Biochemistry* **31**: 11203-11209.

- Hendrickson, T.L., Crecy-Lagard, V., and Schimmel, P. (2004). Incorporation of nonnatural amino acids into proteins. *Annu. Rev. Biochem.* **73**: 147-176.
- Hicke, L. (2001). Protein regulation by monoubiquitin. *Nat. Rev. Mol. Cell Biol.* **2**: 195-201.
- Himeno, H., Yoshida, K., Soma, A., and Nishikawa, K. (1997). Only one nucleotide insertion to the long variable arm confers an efficient serine acceptor activity upon *Saccharomyces cerevisiae* tRNA-Leu in vivo. *J. Mol. Biol.* **268**: 704-711.
- Hinnebusch, A.G. (1996). Translational control of *GCN4*: gene-specific regulation by phosphorylation of eIF2. in *Translational Control* (eds Hershey, J. W. B., Mathews, M. B, and Sonenberg, N.): 199-244. Cold Spring Harbor Laboratory Press,
- Hinnebusch, A.G. and Natarajan, K. (2002). Gcn4p, a master regulator of gene expression, is controlled at multiple levels by diverse signals of starvation and stress. *Eukaryot. Cell* **1**: 22-32.
- Hohfeld, J., Cyr, D.M., and Patterson, C. (2001). From the cradle to the grave: molecular chaperones that may choose between folding and degradation. *EMBO Rep.* **2**: 885-890.
- Hohmann, S. and Mager, W.H. (1997). *Yeast stress response*. Chapman & Hall, New York
- Hruby, V.J. (2002). Designing peptide receptor agonists and antagonists. *Nat. Rev. Drug Discov.* **1**: 847-858.
- Huang, A.M., Lee, J.I., King, S.C., and Wilson, T.H. (1992). Amino acid substitution in the lactose carrier protein with the use of amber suppressors. *J. Bacteriol.* **174**: 5436-5441.
- Hughes, R.A. and Ellington, A.D. (2005). Mistakes in translation don't translate into termination. *Proc. Natl. Acad. Sci. U. S. A* **102**: 1273-1274.

- Hughes, T.R., Roberts, C.J., Dai, H., Jones, A.R., Meyer, M.R., Slade, D., Burchard, J., Dow, S., Ward, T.R., Kidd, M.J., Friend, S.H., and Marton, M.J. (2000). Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.* **25**: 333-337.
- Ibba, M. and Soll, D. (2000). Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69**: 617-650.
- Ibba, M. and Soll, D. (2002). Genetic code: introducing pyrrolysine. *Curr. Biol.* **12**: R464-R466.
- Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., and Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**: 929-934.
- Imai, J., Maruya, M., Yashiroda, H., Yahara, I., and Tanaka, K. (2003). The molecular chaperone Hsp90 plays a role in the assembly and maintenance of the 26S proteasome. *EMBO J.* **22**: 3557-3567.
- Iwafune, Y., Kawasaki, H., and Hirano, H. (2002). Electrophoretic analysis of phosphorylation of the yeast 20S proteasome. *Electrophoresis* **23**: 329-338.
- Jelinsky, S.A., Estep, P., Church, G.M., and Samson, L.D. (2000). Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes. *Mol. Cell Biol.* **20**: 8157-8167.
- Jung, U.S. and Levin, D.E. (1999). Genome-wide analysis of gene expression regulated by the yeast cell wall integrity signalling pathway. *Mol. Microbiol.* **34**: 1049-1057.
- Kano, A., Andachi, Y., Ohama, T., and Osawa, S. (1991). Novel anticodon composition of transfer RNAs in *Micrococcus luteus*, a bacterium with a high genomic G + C content. Correlation with codon usage. *J. Mol. Biol.* **221**: 387-401.

- Kapp, L.D. and Lorsch, J.R. (2004). The molecular mechanics of eukaryotic translation. *Annu. Rev. Biochem.* **73**: 657-704.
- Katze, M.G. (1996). Translational control in cells infected with influenza virus and reovirus. in *Translational Control* (eds Hershey, J. W. B., Mathews, M. B, and Sonenberg, N.): 607-630. Cold Spring Harbor Laboratory Press,
- Kaushik, J.K. and Bhat, R. (2003). Why is trehalose an exceptional protein stabilizer? An analysis of the thermal stability of proteins in the presence of the compatible osmolyte trehalose. *J. Biol. Chem.* **278**: 26458-26465.
- Kellis, M., Birren, B.W., and Lander, E.S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624.
- Kiesewetter, S., Ott, G., and Sprinzl, M. (1990). The role of modified purine 64 in initiator/elongator discrimination of tRNA(iMet) from yeast and wheat germ. *Nucleic Acids Res.* **18**: 4677-4682.
- Kiso, Y., Matsumoto, H., Mizumoto, S., Kimura, T., Fujiwara, Y., and Akaji, K. (1999). Small dipeptide-based HIV protease inhibitors containing the hydroxymethylcarbonyl isostere as an ideal transition-state mimic. *Biopolymers* **51**: 59-68.
- Kloetzel, P.M. (2001). Antigen processing by the proteasome. *Nat. Rev. Mol. Cell Biol.* **2**: 179-187.
- Knight, R.D., Freeland, S.J., and Landweber, L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. *Nat. Rev Genet.* **2**: 49-58.
- Kostova, Z. and Wolf, D.H. (2003). For whom the bell tolls: protein quality control of the endoplasmic reticulum and the ubiquitin-proteasome connection. *EMBO J.* **22**: 2309-2317.

- Koszul, R., Caburet, S., Dujon, B., and Fischer, G. (2004). Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* **23**: 234-243.
- Ladner, J.E., Jack, A., Robertus, J.D., Brown, R.S., Rhodes, D., Clark, B.F., and Klug, A. (1975). Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution. *Proc. Natl. Acad. Sci. U. S. A* **72**: 4414-4418.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.
- Lee, J., Godon, C., Lagniel, G., Spector, D., Garin, J., Labarre, J., and Toledano, M.B. (1999). Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast. *J. Biol. Chem.* **274**: 16040-16046.
- Lee, N., Bessho, Y., Wei, K., Szostak, J.W., and Suga, H. (2000). Ribozyme-catalyzed tRNA aminoacylation. *Nat. Struct. Biol.* **7**: 28-33.
- Lehman, N. (2001). Molecular evolution: Please release me, genetic code. *Curr. Biol.* **11**: R63-R66.
- Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M.A., and Bock, A. (1988). Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* **331**: 723-725.
- Levine, R.L., Williams, J.A., Stadtman, E.R., and Shacter, E. (1994). Carbonyl assays for determination of oxidatively modified proteins. *Methods Enzymol.* **233**: 346-357.
- Lin, S.J. and Guarente, L. (2003). Nicotinamide adenine dinucleotide, a metabolic regulator of transcription, longevity and disease. *Curr. Opin. Cell Biol.* **15**: 241-246.
- Lloyd, A.T. and Sharp, P.M. (1992). Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **20**: 5289-5295.

- Lovett, P.S., Ambulos, N.P., Jr., Mulbry, W., Noguchi, N., and Rogers, E.J. (1991). UGA can be decoded as tryptophan at low efficiency in *Bacillus subtilis*. *J. Bacteriol.* **173**: 1810-1812.
- Lozupone, C.A., Knight, R.D., and Landweber, L.F. (2001). The molecular basis of nuclear genetic code change in ciliates. *Curr. Biol.* **11**: 65-74.
- Lum, R., Tkach, J.M., Vierling, E., and Glover, J.R. (2004). Evidence for an unfolding/threading mechanism for protein disaggregation by *Saccharomyces cerevisiae* Hsp104. *J. Biol. Chem.* **279**: 29139-29146.
- Lyons, T.J., Gasch, A.P., Gaither, L.A., Botstein, D., Brown, P.O., and Eide, D.J. (2000). Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast. *Proc. Natl. Acad. Sci. U. S. A* **97**: 7957-7962.
- Ma, Y. and Hendershot, L.M. (2001). The unfolding tale of the unfolded protein response. *Cell* **107**: 827-830.
- MacDiarmid, C.W., Gaither, L.A., and Eide, D. (2000). Zinc transporters that regulate vacuolar zinc storage in *Saccharomyces cerevisiae*. *EMBO J.* **19**: 2845-2855.
- Mak, C.C., Brik, A., Lerner, D.L., Elder, J.H., Morris, G.M., Olson, A.J., and Wong, C.H. (2003). Design and synthesis of broad-based mono- and bi- cyclic inhibitors of FIV and HIV proteases. *Bioorg. Med. Chem.* **11**: 2025-2040.
- Mannhaupt, G., Schnall, R., Karpov, V., Vetter, I., and Feldmann, H. (1999). Rpn4p acts as a transcription factor by binding to PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast. *FEBS Lett.* **450**: 27-34.
- Martinez-Lopez, R., Monteoliva, L., Diez-Orejas, R., Nombela, C., and Gil, C. (2004). The GPI-anchored protein CaEcm33p is required for cell wall integrity, morphogenesis and virulence in *Candida albicans*. *Microbiology* **150**: 3341-3354.

- Massey, S.E., Moura, G., Beltrao, P., Almeida, R., Garey, J.R., Tuite, M.F., and Santos, M.A. (2003). Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida* spp. *Genome Res.* **13**: 544-557.
- Matthews, M.B. (1996). Interactions between viruses and the cellular machinery for protein synthesis. in *Translational Control* (eds Hershey, J. W. B., Mathews, M. B, and Sonenberg, N.): 505-548. Cold Spring Harbor Laboratory Press,
- McClain, W.H. (1993). Rules that govern tRNA identity in protein synthesis. *J. Mol. Biol.* **234**: 257-280.
- McMurray, M.A. and Gottschling, D.E. (2003). An age-induced switch to a hyper-recombinational state. *Science* **301**: 1908-1911.
- Mehl, R.A., Anderson, J.C., Santoro, S.W., Wang, L., Martin, A.B., King, D.S., Horn, D.M., and Schultz, P.G. (2003). Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**: 935-939.
- Meinzel, T., Mechulam, Y., and Blanquet, S. (1995). Aminoacyl-tRNA Synthetases: Occurrence, Structure, and Function. in *tRNA structure, Biosynthesis and function*. (eds Soll, D. and RajBhandary, U.): 251-292. American Society for Microbiology, Washington, D.C.
- Mizzen, L.A. and Welch, W.J. (1988). Characterization of the thermotolerant cell. I. Effects on protein synthesis activity and the regulation of heat-shock protein 70 expression. *J. Cell Biol.* **106**: 1105-1116.
- Monribot, C. and Boucherie, H. (2000). Two-Dimensional Electrophoresis with Carrier Ampholytes. in *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods* (eds Rabilloud, T.): 31-52. Springer Verlag,
- Mori, K., Kawahara, T., Yoshida, H., Yanagi, H., and Yura, T. (1996). Signalling from endoplasmic reticulum to nucleus: transcription factor with a basic-

- leucine zipper motif is required for the unfolded protein-response pathway. *Genes Cells* **1**: 803-817.
- Morley, J.F. and Morimoto, R.I. (2004). Regulation of longevity in *Caenorhabditis elegans* by heat shock factor and molecular chaperones. *Mol. Biol. Cell* **15**: 657-664.
- Morrow, G., Battistini, S., Zhang, P., and Tanguay, R.M. (2004). Decreased lifespan in the absence of expression of the mitochondrial small heat shock protein hsp22 in *Drosophila*. *J. Biol. Chem.* **279**: 43382-43385.
- Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, T., and Yokoyama, S. (1988). Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* **336**: 179-181.
- Muratani, M. and Tansey, W.P. (2003). How the ubiquitin-proteasome system controls transcription. *Nat. Rev. Mol. Cell Biol.* **4**: 192-201.
- Nangle, L.A., de, C.L., V, Doring, V., and Schimmel, P. (2002). Genetic Code Ambiguity. Cell viability related to severity of editing defects in mutant tRNA synthetases. *J. Biol. Chem.* **277**: 45729-45733.
- Ng, D.T., Spear, E.D., and Walter, P. (2000). The unfolded protein response regulates multiple aspects of secretory and membrane protein biogenesis and endoplasmic reticulum quality control. *J. Cell Biol.* **150**: 77-88.
- Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**: 920-930.
- Nitta, I., Kamada, Y., Noda, H., Ueda, T., and Watanabe, K. (1998). Reconstitution of peptide bond formation with *Escherichia coli* 23S ribosomal RNA domains. *Science* **281**: 666-669.

- Norbeck, J. and Blomberg, A. (1997). Metabolic and regulatory changes associated with growth of *Saccharomyces cerevisiae* in 1.4 M NaCl. Evidence for osmotic induction of glycerol dissimilation via the dihydroxyacetone pathway. *J. Biol. Chem.* **272**: 5544-5554.
- Novoa, I., Zhang, Y., Zeng, H., Jungreis, R., Harding, H.P., and Ron, D. (2003). Stress-induced gene expression requires programmed recovery from translational repression. *EMBO J.* **22**: 1180-1187.
- Nystrom, T. (2002). Translational fidelity, protein oxidation, and senescence: lessons from bacteria. *Ageing Res. Rev.* **1**: 693-703.
- O'Sullivan, J.M., Mihr, M.J., Santos, M.A., and Tuite, M.F. (2001a). Seryl-tRNA synthetase is not responsible for the evolution of CUG codon reassignment in *Candida albicans*. *Yeast* **18**: 313-322.
- O'Sullivan, J.M., Mihr, M.J., Santos, M.A., and Tuite, M.F. (2001b). The *Candida albicans* gene encoding the cytoplasmic leucyl-tRNA synthetase: implications for the evolution of CUG codon reassignment. *Gene* **275**: 133-140.
- Oba, T., Andachi, Y., Muto, A., and Osawa, S. (1991). CGG: an unassigned or nonsense codon in *Mycoplasma capricolum*. *Proc. Natl. Acad. Sci. U. S. A* **88**: 921-925.
- Ogawa, N., DeRisi, J., and Brown, P.O. (2000). New components of a system for phosphate accumulation and polyphosphate metabolism in *Saccharomyces cerevisiae* revealed by genomic expression analysis. *Mol. Biol. Cell* **11**: 4309-4321.
- Ogle, J.M., Brodersen, D.E., Clemons, W.M., Jr., Tarry, M.J., Carter, A.P., and Ramakrishnan, V. (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**: 897-902.

- Ogle, J.M., Carter, A.P., and Ramakrishnan, V. (2003). Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem. Sci.* **28**: 259-266.
- Ogle, J.M., Murphy, F.V., Tarry, M.J., and Ramakrishnan, V. (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111**: 721-732.
- Ohama, T., Muto, A., and Osawa, S. (1990). Role of GC-biased mutation pressure on synonymous codon choice in *Micrococcus luteus*, a bacterium with a high genomic GC-content. *Nucleic Acids Res.* **18**: 1565-1569.
- Ohsumi, Y. (2001). Molecular dissection of autophagy: two ubiquitin-like systems. *Nat. Rev. Mol. Cell Biol.* **2**: 211-216.
- Oleksiak, M.F., Churchill, G.A., and Crawford, D.L. (2002). Variation in gene expression within and among natural populations. *Nat. Genet.* **32**: 261-266.
- Osawa, S. and Jukes, T.H. (1989). Codon reassignment (codon capture) in evolution. *J. Mol. Evol.* **28**: 271-278.
- Osawa, S., Jukes, T.H., Watanabe, K., and Muto, A. (1992). Recent evidence for evolution of the genetic code. *Microbiol. Rev.* **56**: 229-264.
- Owsianik, G., Balzi I L, and Ghislain, M. (2002). Control of 26S proteasome expression by transcription factors regulating multidrug resistance in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **43**: 1295-1308.
- Palkova, Z., Devaux, F., Ilicova, M., Minarikova, L., Le Crom, S., and Jacq, C. (2002). Ammonia pulses and metabolic oscillations guide yeast colony development. *Mol. Biol. Cell* **13**: 3901-3914.
- Palkova, Z. and Forstova, J. (2000). Yeast colonies synchronise their growth and development. *J. Cell Sci.* **113 ( Pt 11)**: 1923-1928.

- Palkova, Z., Janderova, B., Gabriel, J., Zikanova, B., Pospisek, M., and Forstova, J. (1997). Ammonia mediates communication between yeast colonies. *Nature* **390**: 532-536.
- Pallanck, L., Pak, M., and Schulman, L.H. (1995). tRNA Discrimination in Aminoacylation. in *tRNA structure, Biosynthesis and function*. (eds Soll, D. and RajBhandary, U.): 371-394. American Society for Microbiology, Washington, D.C.
- Pan, X., Harashima, T., and Heitman, J. (2000). Signal transduction cascades regulating pseudohyphal differentiation of *Saccharomyces cerevisiae*. *Curr. Opin. Microbiol.* **3**: 567-572.
- Pape, L.K., Koerner, T.J., and Tzagoloff, A. (1985). Characterization of a yeast nuclear gene (MST1) coding for the mitochondrial threonyl-tRNA1 synthetase. *J. Biol. Chem.* **260**: 15362-15370.
- Pape, T., Wintermeyer, W., and Rodnina, M. (1999). Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J.* **18**: 3800-3807.
- Pappas, D.L., Jr., Frisch, R., and Weinreich, M. (2004). The NAD(+)-dependent Sir2p histone deacetylase is a negative regulator of chromosomal DNA replication. *Genes Dev.* **18**: 769-781.
- Park, H.D., Scott, S., Rai, R., Dorrington, R., and Cooper, T.G. (1999). Synergistic operation of the CAR2 (Ornithine transaminase) promoter elements in *Saccharomyces cerevisiae*. *J. Bacteriol.* **181**: 7052-7064.
- Parker, J. and Precup, J. (1986). Mistranslation during phenylalanine starvation. *Mol. Gen. Genet.* **204**: 70-74.
- Parrou, J.L. and Francois, J. (1997). A simplified procedure for a rapid and reliable assay of both glycogen and trehalose in whole yeast cells. *Anal. Biochem.* **248**: 186-188.

- Parrou, J.L., Teste, M.A., and Francois, J. (1997). Effects of various types of stress on the metabolism of reserve carbohydrates in *Saccharomyces cerevisiae*: genetic evidence for a stress-induced recycling of glycogen and trehalose. *Microbiology* **143 ( Pt 6)**: 1891-1900.
- Patil, C. and Walter, P. (2001). Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. *Curr. Opin. Cell Biol.* **13**: 349-355.
- Paushkin, S.V., Kushnirov, V.V., Smirnov, V.N., and Ter Avanesyan, M.D. (1996). Propagation of the yeast prion-like [psi+] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor. *EMBO J.* **15**: 3127-3134.
- Pederson, B.A., Wilson, W.A., and Roach, P.J. (2004). Glycogen synthase sensitivity to glucose-6-P is important for controlling glycogen accumulation in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**: 13764-13768.
- Perrot, M., Sagliocco, F., Mini, T., Monribot, C., Schneider, U., Shevchenko, A., Mann, M., Jenö, P., and Boucherie, H. (1999). Two-dimensional gel protein database of *Saccharomyces cerevisiae* (update 1999). *Electrophoresis* **20**: 2280-2298.
- Pezo, V., Metzgar, D., Hendrickson, T.L., Waas, W.F., Hazebrouck, S., Doring, V., Marliere, P., Schimmel, P., and Crecy-Lagard, V. (2004). Artificially ambiguous genetic code confers growth yield advantage. *Proc. Natl. Acad. Sci. U. S. A*
- Pickart, C.M. and Cohen, R.E. (2004). Proteasomes and their kin: proteases in the machine age. *Nat. Rev. Mol. Cell Biol.* **5**: 177-187.
- Pinson, B., Merle, M., Franconi, J.M., and Daignan-Fornier, B. (2004). Low affinity orthophosphate carriers regulate PHO gene expression independently of internal orthophosphate concentration in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**: 35273-35280.

- Poole, E. and Tate, W. (2000). Release factors and their role as decoding proteins: specificity and fidelity for termination of protein synthesis. *Biochim. Biophys. Acta* **1493**: 1-11.
- Praetorius-Ibba, M. and Ibba, M. (2003). Aminoacyl-tRNA synthesis in archaea: different but not unique. *Mol. Microbiol.* **48**: 631-637.
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**: 218-229.
- Queitsch, C., Sangster, T.A., and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. *Nature* **417**: 618-624.
- Rabilloud, T. and Charmont, S. (2000). Detection of Proteins on Two-Dimensional Electrophoresis Gels. in *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods* (eds Rabilloud, T.): 107-126. Springer Verlag,
- Rabilloud, T., Strub, J.M., Carte, N., Luche, S., Van Dorsselaer, A., Lunardi, J., Giege, R., and Florentz, C. (2002). Comparative proteomics as a new tool for exploring human mitochondrial tRNA disorders. *Biochemistry* **41**: 144-150.
- RajBhandary, U.L. and Ming Chow, C. (1995). Initiator tRNAs and Initiation of Protein Synthesis. in *tRNA structure, Biosynthesis and function*. (eds Soll, D. and RajBhandary, U. L.): 511-528. American Society for Microbiology, Washington, D.C.
- Reverter-Branchat, Cabiscol, E., Tamarit, J., and Ros, J. (2004). Oxidative Damage to Specific Proteins in Replicative and Chronological-aged *Saccharomyces cerevisiae*: COMMON TARGETS AND PREVENTION BY CALORIE RESTRICTION. *J. Biol. Chem.* **279**: 31983-31989.

- Ribas de Pouplana, L. and Schimmel, P. (2001a). Aminoacyl-tRNA synthetases: potential markers of genetic code development. *Trends Biochem. Sci.* **26**: 591-596.
- Ribas de Pouplana, L. and Schimmel, P. (2001b). Operational RNA code for amino acids in relation to genetic code in evolution. *J. Biol. Chem.* **276**: 6881-6884.
- Rich, A. and Kim, S.H. (1978). The three-dimensional structure of transfer RNA. *Sci. Am.* **238**: 52-62.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**: 1030-1032.
- Rocak, S., Landeka, I., and Weygand-Durasevic, I. (2002). Identifying Pex21p as a protein that specifically interacts with yeast seryl-tRNA synthetase. *FEMS Microbiol. Lett.* **214**: 101-106.
- Rodnina, M.V. and Wintermeyer, W. (2001a). Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annu. Rev. Biochem.* **70**: 415-435.
- Rodnina, M.V. and Wintermeyer, W. (2001b). Ribosome fidelity: tRNA discrimination, proofreading and induced fit. *Trends Biochem. Sci.* **26**: 124-130.
- Ruan, B., Ahel, I., Ambrogelly, A., Becker, H.D., Bunjun, S., Feng, L., Tumbula-Hansen, D., Ibba, M., Korencic, D., Kobayashi, H., Jacquin-Becker, C., Mejlhede, N., Min, B., Raczniak, G., Rinehart, J., Stathopoulos, C., Li, T., and Soll, D. (2001). Genomics and the evolution of aminoacyl-tRNA synthesis. *Acta Biochim. Pol.* **48**: 313-321.
- Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., and Moras, D. (1991). Class II aminoacyl

- transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science* **252**: 1682-1689.
- Rutherford, S.L. and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* **396**: 336-342.
- Ruusala, T., Ehrenberg, M., and Kurland, C.G. (1982). Is there proofreading during polypeptide synthesis? *EMBO J.* **1**: 741-745.
- Sagliocco, F., Guillemot, J.C., Monribot, C., Capdevielle, J., Perrot, M., Ferran, E., Ferrara, P., and Boucherie, H. (1996). Identification of proteins of the yeast protein map using genetically manipulated strains and peptide-mass fingerprinting. *Yeast* **12**: 1519-1533.
- Sanchez-Silva, R., Villalobo, E., Morin, L., and Torres, A. (2003). A new noncanonical nuclear genetic code: translation of UAA into glutamate. *Curr. Biol.* **13**: 442-447.
- Santos, M.A.S. (1992). *Molecular characterization of a novel leucyl-tRNA<sub>CAG</sub> from Candida albicans*. PhD Thesis. University of Kent,
- Santos, M.A.S., Cheesman, C., Costa, V., Moradas-Ferreira, P., and Tuite, M.F. (1999). Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in *Candida* spp. *Mol. Microbiol.* **31**: 937-947.
- Santos, M.A.S., Perreau, V.M., and Tuite, M.F. (1996). Transfer RNA structural change is a key element in the reassignment of the CUG codon in *Candida albicans*. *EMBO J.* **15**: 5060-5068.
- Santos, M.A.S. and Tuite, M.F. (1995). The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. *Nucleic Acids Res.* **23**: 1481-1486.

- Santos, M.A.S., Ueda, T., Watanabe, K., and Tuite, M.F. (1997). The non-standard genetic code of *Candida* spp.: an evolving genetic code or a novel mechanism for adaptation? *Mol. Microbiol.* **26**: 423-431.
- Schaaff-Gerstenschlager, I. and Zimmermann, F.K. (1993). Pentose-phosphate pathway in *Saccharomyces cerevisiae*: analysis of deletion mutants for transketolase, transaldolase, and glucose 6-phosphate dehydrogenase. *Curr. Genet.* **24**: 373-376.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467-470.
- Schimmel, P. and Beebe, K. (2004). Molecular biology: genetic code seizes pyrrolysine. *Nature* **431**: 257-258.
- Schimmel, P., Giege, R., Moras, D., and Yokoyama, S. (1993). An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci. U. S. A* **90**: 8763-8768.
- Schimmel, P. and Ribas de Pouplana, L. (1995). Transfer RNA: from minihelix to genetic code. *Cell* **81**: 983-986.
- Schmitt, E., Panvert, M., Blanquet, S., and Mechulam, Y. (1998). Crystal structure of methionyl-tRNA<sup>fMet</sup> transformylase complexed with the initiator formyl-methionyl-tRNA<sup>fMet</sup>. *EMBO J.* **17**: 6819-6826.
- Schmitt, M.E., Brown, T.A., and Trumpower, B.L. (1990). A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **18**: 3091-3092.
- Schneider, R.J. (1996). Adenovirus and vaccinia virus translational control. in *Translational Control* (eds Hershey, J. W. B., Mathews, M. B, and Sonenberg, N.): 575-606. Cold Spring Harbor Laboratory Press,

- Schnell, J.D. and Hicke, L. (2003). Non-traditional functions of ubiquitin and ubiquitin-binding proteins. *J. Biol. Chem.* **278**: 35857-35860.
- Schulman, L.H. (1991). Recognition of tRNAs by aminoacyl-tRNA synthetases. *Prog. Nucleic Acid Res. Mol. Biol.* **41**: 23-87.
- Schulman, L.H. and Her, M.O. (1973). Recognition of altered E. coli formylmethionine transfer RNA by bacterial T factor. *Biochem. Biophys. Res. Commun.* **51**: 275-282.
- Schulman, L.H. and Pelka, H. (1988). Anticodon switching changes the identity of methionine and valine transfer RNAs. *Science* **242**: 765-768.
- Schultz, D.W. and Yarus, M. (1994). Transfer RNA mutation and the malleability of the genetic code. *J. Mol. Biol.* **235**: 1377-1380.
- Seong, B.L. and RajBhandary, U.L. (1987). Mutants of Escherichia coli formylmethionine tRNA: a single base change enables initiator tRNA to act as an elongator in vitro. *Proc. Natl. Acad. Sci. U. S. A* **84**: 8859-8863.
- Serrano, R., Bernal, D., Simon, E., and Arino, J. (2004). Copper and iron are the limiting factors for growth of the yeast *Saccharomyces cerevisiae* in an alkaline environment. *J. Biol. Chem.* **279**: 19698-19704.
- Sherman, J.M., Rogers, M.J., and Soll, D. (1995). Recognition in the Glutamine tRNA System: from Structure to Function. in *tRNA structure, Biosynthesis and function*. (eds Soll, D. and RajBhandary, U.): 395-409. American Society for Microbiology, Washington, D.C.
- Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., and Mann, M. (1996). Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. U. S. A* **93**: 14440-14445.

- Siderius, M., Van Wuytswinkel, O., Reijenga, K.A., Kelders, M., and Mager, W.H. (2000). The control of intracellular glycerol in *Saccharomyces cerevisiae* influences osmotic stress response and resistance to increased temperature. *Mol. Microbiol.* **36**: 1381-1390.
- Silva, R.M., Miranda, I., Moura, G., and Santos, M.A. (2004). Yeast as a model organism for studying the evolution of non-standard genetic codes. *Brief. Funct. Genomic. Proteomic.* **3**: 35-46.
- Sinclair, D.A. (2003). Cell biology. An age of instability. *Science* **301**: 1859-1860.
- Sinclair, D.A. and Guarente, L. (1997). Extrachromosomal rDNA circles--a cause of aging in yeast. *Cell* **91**: 1033-1042.
- Singer, M.A. and Lindquist, S. (1998). Multiple effects of trehalose on protein folding in vitro and in vivo. *Mol. Cell* **1**: 639-648.
- Soma, A., Kumagai, R., Nishikawa, K., and Himeno, H. (1996). The anticodon loop is a major identity determinant of *Saccharomyces cerevisiae* tRNA-Leu. *J. Mol. Biol.* **263**: 707-714.
- Sprinzi, M., Horn, C., Brown, M., loudovitch, A., and Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**: 148-153.
- Srinivasan, G., James, C.M., and Krzycki, J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**: 1459-1462.
- Stansfield, I., Jones, K.M., Herbert, P., Lewendon, A., Shaw, W.V., and Tuite, M.F. (1998). Missense translation errors in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **282**: 13-24.
- Steinberg, S., Misch, A., and Sprinzi, M. (1993). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **21**: 3011-3015.

- Stolz, J., Hoja, U., Meier, S., Sauer, N., and Schweizer, E. (1999). Identification of the plasma membrane H<sup>+</sup>-biotin symporter of *Saccharomyces cerevisiae* by rescue of a fatty acid-auxotrophic mutant. *J. Biol. Chem.* **274**: 18741-18746.
- Storchova, Z. and Pellman, D. (2004). From polyploidy to aneuploidy, genome instability and cancer. *Nat. Rev. Mol. Cell Biol.* **5**: 45-54.
- Suzuki, T., Ueda, T., and Watanabe, K. (1997). The 'polysemous' codon--a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. *EMBO J.* **16**: 1122-1134.
- Suzuki, Y., Nishimaki, K., Taniyama, M., and Muramatsu, T. (2005). Mitochondrial tRNA (Leu(UUR)) mutation at position 3243 detected in patients with type 1 diabetes. *Diabetes Res. Clin. Pract.* **67**: 92-94.
- Szathmáry, E. (1999). The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends Genet.* **15**: 223-229.
- Takehige, K., Baba, M., Tsuboi, S., Noda, T., and Ohsumi, Y. (1992). Autophagy in yeast demonstrated with proteinase-deficient mutants and conditions for its induction. *J. Cell Biol.* **119**: 301-311.
- Thevelein, J.M., Cauwenberg, L., Colombo, S., de Winde, J.H., Donation, M., Dumortier, F., Kraakman, L., Lemaire, K., Ma, P., Nauwelaers, D., Rolland, F., Teunissen, A., Van Dijck, P., Versele, M., Wera, S., and Winderickx, J. (2000). Nutrient-induced signal transduction through the protein kinase A pathway and its role in the control of metabolism, stress resistance, and growth in yeast. *Enzyme Microb. Technol.* **26**: 819-825.
- Thevelein, J.M. and de Winde, J.H. (1999). Novel sensing mechanisms and targets for the cAMP-protein kinase A pathway in the yeast *Saccharomyces cerevisiae*. *Mol. Microbiol.* **33**: 904-918.
- Thompson, R.C. and Stone, P.J. (1977). Proofreading of the codon-anticodon interaction on ribosomes. *Proc. Natl. Acad. Sci. U. S. A* **74**: 198-202.

- Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **19**: 94-102.
- Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A* **76**: 4350-4354.
- Trachsel, H. (1996). Binding of Initiator Methionyl-tRNA to Ribosomes. in *Translational Control* (eds Hershey, J. W. B., Mathews, M. B, and Sonenberg, N.): 113-138. Cold Spring Harbor Laboratory Press,
- Travers, K.J., Patil, C.K., Wodicka, L., Lockhart, D.J., Weissman, J.S., and Walter, P. (2000). Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell* **101**: 249-258.
- Trotter, E.W., Kao, C.M., Berenfeld, L., Botstein, D., Petsko, G.A., and Gray, J.V. (2002). Misfolded proteins are competent to mediate a subset of the responses to heat shock in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **277**: 44817-44825.
- True, H.L., Berlin, I., and Lindquist, S.L. (2004). Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. *Nature* **431**: 184-187.
- True, H.L. and Lindquist, S.L. (2000). A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**: 477-483.
- Tujebajeva, R.M., Copeland, P.R., Xu, X.M., Carlson, B.A., Harney, J.W., Driscoll, D.M., Hatfield, D.L., and Berry, M.J. (2000). Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Rep.* **1**: 158-163.
- Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A* **98**: 5116-5121.

- Ueda, T., Nagata, M., and Imoto, T. (2001). Aggregation and chemical reaction in hen lysozyme caused by heating at pH 6 are depressed by osmolytes, sucrose and trehalose. *J. Biochem. (Tokyo)* **130**: 491-496.
- Ueda, T., Suzuki, T., Yokogawa, T., Nishikawa, K., and Watanabe, K. (1994). Unique structure of new serine tRNAs responsible for decoding leucine codon CUG in various *Candida* species and their putative ancestral tRNA genes. *Biochimie* **76**: 1217-1222.
- Valente, L. and Kinzy, T.G. (2003). Yeast as a sensor of factors affecting the accuracy of protein synthesis. *Cell Mol. Life Sci.* **60**: 2115-2130.
- van Leeuwen, F.W., Gerez, L., Benne, R., and Hol, E.M. (2002). +1 Proteins and aging. *Int. J. Biochem. Cell Biol.* **34**: 1502-1505.
- Verma, R., Peters, N.R., D'Onofrio, M., Tochtrop, G.P., Sakamoto, K.M., Varadan, R., Zhang, M., Coffino, P., Fushman, D., Deshaies, R.J., and King, R.W. (2004). Ubistatins inhibit proteasome-dependent degradation by binding the ubiquitin chain. *Science* **306**: 117-120.
- Vido, K., Spector, D., Lagniel, G., Lopez, S., Toledano, M.B., and Labarre, J. (2001). A proteome analysis of the cadmium response in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**: 8469-8474.
- Wang, L., Brock, A., Herberich, B., and Schultz, P.G. (2001). Expanding the genetic code of *Escherichia coli*. *Science* **292**: 498-500.
- Wegele, H., Haslbeck, M., Reinstein, J., and Buchner, J. (2003). Sti1 is a novel activator of the Ssa proteins. *J. Biol. Chem.* **278**: 25970-25976.
- Weiner, A.M. and Maizels, N. (1987). tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc. Natl. Acad. Sci. U. S. A* **84**: 7383-7387.
- Weissman, A.M. (2001). Themes and variations on ubiquitylation. *Nat. Rev. Mol. Cell Biol.* **2**: 169-178.

- Wemmie, J.A., Wu, A.L., Harshman, K.D., Parker, C.S., and Moye-Rowley, W.S. (1994). Transcriptional activation mediated by the yeast AP-1 protein is required for normal cadmium tolerance. *J. Biol. Chem.* **269**: 14690-14697.
- Werner-Washburne, M., Becker, J., Kasic-Smithers, J., and Craig, E.A. (1989). Yeast Hsp70 RNA levels vary in response to the physiological status of the cell. *J. Bacteriol.* **171**: 2680-2688.
- Werner-Washburne, M., Braun, E.L., Crawford, M.E., and Peck, V.M. (1996). Stationary phase in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **19**: 1159-1166.
- Werner-Washburne, M., Stone, D.E., and Craig, E.A. (1987). Complex interactions among members of an essential subfamily of hsp70 genes in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **7**: 2568-2577.
- Wheeler, R.T., Kupiec, M., Magnelli, P., Abeijon, C., and Fink, G.R. (2003). A *Saccharomyces cerevisiae* mutant with increased virulence. *Proc. Natl. Acad. Sci. U. S. A* **100**: 2766-2770.
- Wiegand, T.W., Janssen, R.C., and Eaton, B.E. (1997). Selection of RNA amide synthases. *Chem. Biol.* **4**: 675-683.
- Wilson, F.H., Hariri, A., Farhi, A., Zhao, H., Petersen, K.F., Toka, H.R., Nelson-Williams, C., Raja, K.M., Kashgarian, M., Shulman, G.I., Scheinman, S.J., and Lifton, R.P. (2004). A Cluster of Metabolic Defects Caused by Mutation in a Mitochondrial tRNA. *Science*
- Woese, C.R., Olsen, G.J., Ibba, M., and Soll, D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**: 202-236.
- Woo, N.H., Roe, B.A., and Rich, A. (1980). Three-dimensional structure of *Escherichia coli* initiator tRNA<sup>fMet</sup>. *Nature* **286**: 346-351.

- Wu, A., Wemmie, J.A., Edgington, N.P., Goebel, M., Guevara, J.L., and Moye-Rowley, W.S. (1993). Yeast bZip proteins mediate pleiotropic drug and metal resistance. *J. Biol. Chem.* **268**: 18850-18858.
- Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**: e15-
- Yang, Y.H. and Speed, T. (2002). Design issues for cDNA microarray experiments. *Nat. Rev. Genet.* **3**: 579-588.
- Yarus, M. and Smith, D. (1995). tRNA on the Ribosome: a Waggle Theory. in *tRNA structure, Biosynthesis and function*. (eds Soll, D. and RajBhandary, U.): 443-469. American Society for Microbiology, Washington, D.C.
- Yasukawa, T., Suzuki, T., Ueda, T., Ohta, S., and Watanabe, K. (2000). Modification defect at anticodon wobble nucleotide of mitochondrial tRNAs(Leu)(UUR) with pathogenic mutations of mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes. *J. Biol. Chem.* **275**: 4251-4257.
- Yokobori, S., Suzuki, T., and Watanabe, K. (2001). Genetic code variations in mitochondria: tRNA as a major determinant of genetic code plasticity. *J. Mol. Evol.* **53**: 314-326.
- Yokogawa, T., Suzuki, T., Ueda, T., Mori, M., Ohama, T., Kuchino, Y., Yoshinari, S., Motoki, I., Nishikawa, K., Osawa, S., and . (1992). Serine tRNA complementary to the nonuniversal serine codon CUG in *Candida cylindracea*: evolutionary implications. *Proc. Natl. Acad. Sci. U. S. A* **89**: 7408-7411.
- Yokoyama, S. and Nishimura, S. (1995). Modified Nucleosides and Codon Recognition. in *tRNA structure, Biosynthesis and function*. (eds Soll, D. and

RajBhandary, U.): 207-223. American Society for Microbiology, Washington, D.C.

Zabin, H.B., Horvath, M.P., and Terwilliger, T.C. (1991). Approaches to predicting effects of single amino acid substitutions on the function of a protein. *Biochemistry* **30**: 6230-6240.

Zinoni, F., Birkmann, A., Stadtman, T.C., and Bock, A. (1986). Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A* **83**: 4650-4654.

## **VI. Annexes**



**Table A – Genes whose expression is induced by genetic code ambiguity.**

ORF	GENE	Fold increase	ORF	GENE	Fold increase
YOL154W	ZPS1	19,77	YBL049W	YBL049W	3,19
YML123C	PHO84	17,62	YAR071W	PHO11	3,12
YFL014W	HSP12	12,35	YBR117C	TKL2	2,97
YOL052C-A	DDR2	9,02	YOR031W	CRS5	2,96
YNL160W	YGP1	8,75	YKL178C	STE3	2,94
YKL163W	PIR3	7,48	YHR053C	CUP1-1	2,91
YHR136C	SPL2	6,90	YBR093C	PHO5	2,89
YBR067C	TIP1	6,20	YDR077W	SED1	2,83
YFR053C	HXK1	6,17	YHR055C	CUP1-2	2,80
YBR072W	HSP26	5,83	YER150W	SPI1	2,75
YLR346C	YLR346C	5,82	YBR169C	SSE2	2,69
YHR087W	YHR087W	5,55	YGR248W	SOL4	2,69
YBR285W	YBR285W	5,34	YKL221W	MCH2	2,68
YBR296C	PHO89	4,67	YMR250W	GAD1	2,58
YPR160W	GPH1	4,65	YLR312C	YLR312C	2,54
YBR054W	YRO2	4,50	YGR142W	BTN2	2,43
YKL096W-A	CWP2	4,00	YMR105C	PGM2	2,39
YER103W	SSA4	3,95	YLL026W	HSP104	2,38
YDR171W	HSP42	3,89	YMR169C	ALD3	2,36
YGL121C	GPG1	3,76	YBL078C	AUT7	2,30
YCR021C	HSP30	3,71	YCL040W	GLK1	2,25
YLR142W	PUT1	3,70	YOR153W	PDR5	2,22
YML128C	MSC1	3,57	YDL145C	COP1	2,18
YGR008C	STF2	3,50	YBR126C	TPS1	2,14
YHR215W	PHO12	3,37	YIL117C	PRM5	2,13
YLR042C	YLR042C	3,34	YGL037C	PNC1	2,10
YML100W	TSL1	3,24	YOL151W	GRE2	2,08
YDL168W	SFA1	3,23	YKR093W	PTR2	2,05
YLR178C	TFS1	3,21	YER011W	TIR1	2,00
			YDL020C	RPN4	1,84

**Table B – Genes whose expression is repressed by genetic code ambiguity.**

ORF	GENE	Fold decrease
YKL218C	SRY1	2,53
YNR050C	LYS9	2,36
YMR321C	YMR321C	2,23
YFL034C-A	RPL22B	2,14
YDR341C	YDR341C	2,14
YML116W	ATR1	2,13
YNL067W	RPL9B	2,04
YGR065C	VHT1	2,03
YNR056C	BIO5	2,02
YNR009W	YNR009W	2,01
YPL273W	SAM4	2,01
YNR058W	BIO3	1,94
YGL202W	ARO8	1,91
YLL053C	YLL053C	1,85
YNR069C	BSC5	1,85
YPR058W	YMC1	1,85
YBR248C	HIS7	1,75
YCL009C	ILV6	1,73
YOR130C	ORT1	1,71
YHR029C	YHR029C	1,70
YGR239C	PEX21	1,62

**Table I – Proteins altered by genetic code ambiguity (tRNA strain vs. control strain, at 25°C), indicating the respective fold variation.** Proteins that are not expressed in the control condition are considered new (n) spots and proteins that are not expressed in the stress condition are considered disappeared (d) spots (note that their real fold variation, therefore, might not be accurately determined).

Proteins Up-regulated		Proteins Down-regulated		New spots	
Name	Fold variation	Name	Fold variation	Name	Fold variation
PNC1	29,5	YDJ1	d	SSA4	n
HSP104	13,1	TDH2c	d	HXK1a	n
RPN12	5,2	TDH2a	d	HXK1b	n
GLK1c	5,0	MET17	5,3		
SSA1	4,7	ARO8a	5,1	+ 5 unidentified spots	
YIL041W	4,0	ARG1	5,0		
RPN10	3,6	ARG4	4,4		
PUP2	3,6	ZWF1	3,7		
SCL1	3,6	ARG1b	3,3		
APT1	3,6	TDH2b	3,3		
SFA1	3,5	MET3	2,9		
YLR109Wa	3,4	LYS9	2,8		
STI1b	3,2	KRS1a	2,7		
SSC1c	3,2	KRS1b	2,7		
ARA1	3,1	GND1b	2,3		
SEC14	3,1	LEU2	2,3		
HSP78	2,7	SSB1a	2,1		
STI1a	2,7	SSB2	2,1		
YRB1	2,7	ARG1a	2,0		
SSA2a	2,5	THR4	2,0		
+ 15 unidentified spots		+ 14 unidentified spots			

**Table II – Proteins altered by growth at 37°C in the control strain (control at 37°C vs. control at 25°C), indicating the respective fold variation.** Proteins that are not expressed in the control condition are considered new (n) spots and proteins that are not expressed in the stress condition are considered disappeared (d) spots (note that their real fold variation, therefore, might not be accurately determined).

Proteins Up-regulated		Proteins Down-regulated		New spots	
Name	Fold variation	Name	Fold variation	Name	Fold variation
PNC1	15,0	EGD2	d	HXK1a	8,4
HSP104	9,5	SNZ1	3,5	UGP1	2,9
GLK1c	5,9	ASN1	3,1		
ADE3a	5,6	ARG3	2,8	+ 2 unidentified spots	
OYE2	3,9	ARG1b	2,3		
TUB1	3,9	HXK2b	2,2		
SSC1c	3,5	LYS9	2,2		
ALD4	2,5	TDH2a	2,2		
SOD1	2,5	GUA1b	2,1		
ENO1	2,5	ARG8	2,1		
MET3	2,3	ASN2	2,0		
TIF51Ac	2,3	SAM1b	2,0		
HSP60a	2,1				
ADK1	2,0	+ 9 unidentified spots			
STI1b	2,0				
ADH1a	2,0				
HSP78	2,0				
SSA1	1,9				
LPD1	1,0				
MET11	1,9				
+ 10 unidentified spots					

**Table III – Proteins altered by growth at 37°C in the tRNA strain (tRNA at 37°C vs. tRNA at 25°C), indicating the respective fold variation.** Proteins that are not expressed in the control condition are considered new (n) spots and proteins that are not expressed in the stress condition are considered disappeared (d) spots (note that their real fold variation, therefore, might not be accurately determined).

Proteins Up-regulated		Proteins Down-regulated		New spots	
Name	Fold variation	Name	Fold variation	Name	Fold variation
ADH1a	3,9	SNZ1	6,2	UGP1	3,3
OYE2	2,8	APT1	5,3		
MET3	2,5	YCL028W	4,4		
		YLR192C	4,0		
+ 4 unidentified spots		TSA1a	3,7		
		YIL041W	2,7		
		SFA1	2,6		
		LYS20b	2,6		
		SSA2a	2,4		
		SSA2b	2,4		
		TSA1b	2,2		
		SAM1b	2,2		
		CIM5a	2,1		
		YHB1	2,0		
		+ 7 unidentified spots			

**Table IV – Proteins altered by growth at 37°C (tRNA strain vs. control strain), indicating the respective fold variation.** Proteins that are not expressed in the control condition are considered new (n) spots and proteins that are not expressed in the stress condition are considered disappeared (d) spots (note that their real fold variation, therefore, might not be accurately determined).

Proteins Up-regulated		Proteins Down-regulated		New spots	
Name	Fold variation	Name	Fold variation	Name	Fold variation
PRE8	3,5	YHB1	2,4	SSA4	2,3
HSP104	2,6	MET17a	2,4		
RPN12	2,5	ADH1a	2,2	+ 1 unidentified spot	
CCT1b	2,3	MET3	2,1		
RPN10	2,2	GND1b	2,0		
PNC1	2,2	CYS4a	2,0		
SEC14	2,2				
YTA2	2,1	+ 4 unidentified spots			
CCT2	2,1				
CAR1	2,1				
PRE9	2,0				
SUG2	1,9				
+ 6 unidentified spots					

**Table V – Proteins altered by heat shock in the control strain (control heat shock vs. control at 25°C), indicating the respective fold variation.** Proteins that are not expressed in the control condition are considered new (n) spots and proteins that are not expressed in the stress conditions are considered disappeared (d) spots (note that their real fold variation, therefore, might not be accurately determined).

Proteins Up-regulated		Proteins Down-regulated		New spots	
Name	Fold variation	Name	Fold variation	Name	Fold variation
PNC1	135,4	SSB1a	d	SSA4	31,9
HSP104	112,3	TDH2a	12,0	HXK1a	7,9
GLK1c	45,1	ADE13	10,4	HXK1b	10,9
TPS1a	23,5	SSB1b	8,6	TKL2a	2,7
ALD4	18,5	TIF1c	7,2	TKL2b	17,2
ARO9	11,0	ADE17a	5,8	SSA3	2,6
TPS1b	10,4	SSB2	5,5	SSA4b	5,7
GPD1a	10,3	GUA1a	4,9	PGM2	14,4
YDR380W	9,1	GUA1b	4,9	CTT1	15,7
YDR032C	8,0	SAH1a	4,4	WTM1	8,9
HSP78	7,3	URA7	4,1	UGP1	21,0
GPD1b	7,1	TIF1b	4,0	ALD3a	57,5
STI1b	6,8	ARO4b	3,9	ALD3b	18,6
HSP82b	6,5	LYS9	3,8	PRB1	3,4
HSP82a	6,1	TDH2b	3,5	YER067W	0,7
GRS1a	6,1	ILV5	3,4		
ARA1	5,5	ASC1	3,3	+ 16 unidentified spots	
SSA1	5,1	HXK2b	3,1		
STI1a	4,9	SAM1b	3,0		
TSA1a	4,7	SAH1b	2,9		
GPS1c	4,4	YHB1	2,9		
YLR109Wa	4,1	YDR341C	2,8		
LAT1	4,0	SAM1a	2,7		
ALD6a	3,8	DYS1	2,7		
SOD1	3,7	SEC53a	2,6		
CIT2	3,6	RPA0	2,4		
SEC53b	3,6	SES1	2,3		
LPD1	3,5	HTS1	2,3		
CDC37	3,2	ERG6	2,0		
HAD1	3,1	TIF34	2,0		
DAK1b	2,9	ADH1b	2,0		
BMH2	2,7	GLN1b	1,8		
MET11	2,7	ADE1	d		
SBP1	2,7	GLN1c	d		

APT1	2,6	ADE17b	d		
ARP2	2,6				
PDA1	2,4	+ 10 unidentified spots			
BMH1	2,3				
ATP2	2,2				
CAR1	2,2				
MET3	2,2				
ALD6b	2,1				
TRR1	1,9				
SFA1	1,9				
+ 42 unidentified spots					

**Table VI - Proteins altered by heat shock in the tRNA strain (tRNA heat shock vs. tRNA 25°C), indicating the respective fold variation.** Proteins that are not expressed in the control condition are considered new (n) spots and proteins that are not expressed in the stress conditions are considered disappeared (d) spots (note that their real fold variation, therefore, might not be accurately determined).

Proteins Up-regulated		Proteins Down-regulated		New spots	
Name	Fold variation	Name	Fold variation	Name	Fold variation
ALD3	185,3	SSB1a	d	TKL2	5,1
TKL2	83,4	ARO8a	d	SSA3	6,3
HXK1a	54,1	SAM1c	d	SSA4	12,4
SSA4	26,4	SAM2a	d	PGM2	33,5
GLK1c	11,5	LYS20b	6,5	CTT1	25,7
HSP104	10,1	URA7	5,8	WTM1	7,2
TPS1a	9,6	ADE13	5,7	UGP1	19,0
HXK1b	9,0	SNZ1	5,6	ALD3	26,4
HSP82b	8,4	YHB1	5,4	PRB1	4,9
PNC1	8,1	SSB1b	5,0	YER067W	12,0
TPS1b	7,6	LYS12	4,3	GPD1	5,8
HSP82a	6,8	ARG1a	4,2		
ARO9	6,1	ARG1b	4,0	+ 7 unidentified spots	
GPD1a	5,9	SAH1a	4,0		
PFK2a	5,0	MET6b	4,0		
HSP78	4,0	ARO4b	3,9		

PYC2b	3,6	ASN1	3,6		
GPD1b	3,5	DYS1	3,5		
OYE2	3,3	LYS20a	3,5		
ALD7	3,3	ILV5	3,3		
STI1b	2,4	ASN2	3,3		
SSE1b	2,3	TIF1b	3,3		
YLR109Wa	2,1	ARG4	3,3		
SSC1c	2,0	PRO3	3,3		
ARA1	2,0	GUA1a	3,1		
SSA1	1,9	ERG6	3,0		
LAT1	1,9	LYS9	3,0		
SOD1	1,9	ACO2a	3,0		
TIF51Ac	1,8	MET6a	3,0		
STI1a	1,6	SSB2	2,8		
		YJR105Wb	2,7		
+33 unidentified spots		GRS1b	2,6		
		YIL041W	2,5		
		TIF1c	2,5		
		GPP1	2,5		
		PSA1	2,4		
		SAM1a	2,4		
		SAM2b	2,2		
		ERG19	2,2		
		SAM1b	2,2		
		SER1	2,2		
		SAH1b	2,1		
		YJR105Wa	2,0		
		SEC53a	2,0		
		KRS1a	2,0		
		APT1	1,9		
		ASC1	1,7		
		MET3	d		
		ADE1	d		
		+ 12 unidentified spots			

**Map 1 – 2D-map of proteins from *S. cerevisiae* control cells, grown at 37°C.**

Proteins were labelled with [<sup>35</sup>S]-methionine in cultures grown to mid-log phase (OD<sub>600nm</sub> = 0.5). The map shows in black the spots corresponding to the control cells grown at 25°C and indicates in colour the proteins whose expression level is altered by at least 2 fold in *S. cerevisiae* control cells grown at 37°C: in red, proteins that are induced; in green, proteins that are repressed; and in blue, proteins that are uniquely expressed in *S. cerevisiae* control cells when grown at 37°C (page 287).

**Map 2 – 2D-map of proteins from *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>, grown at 37°C.**

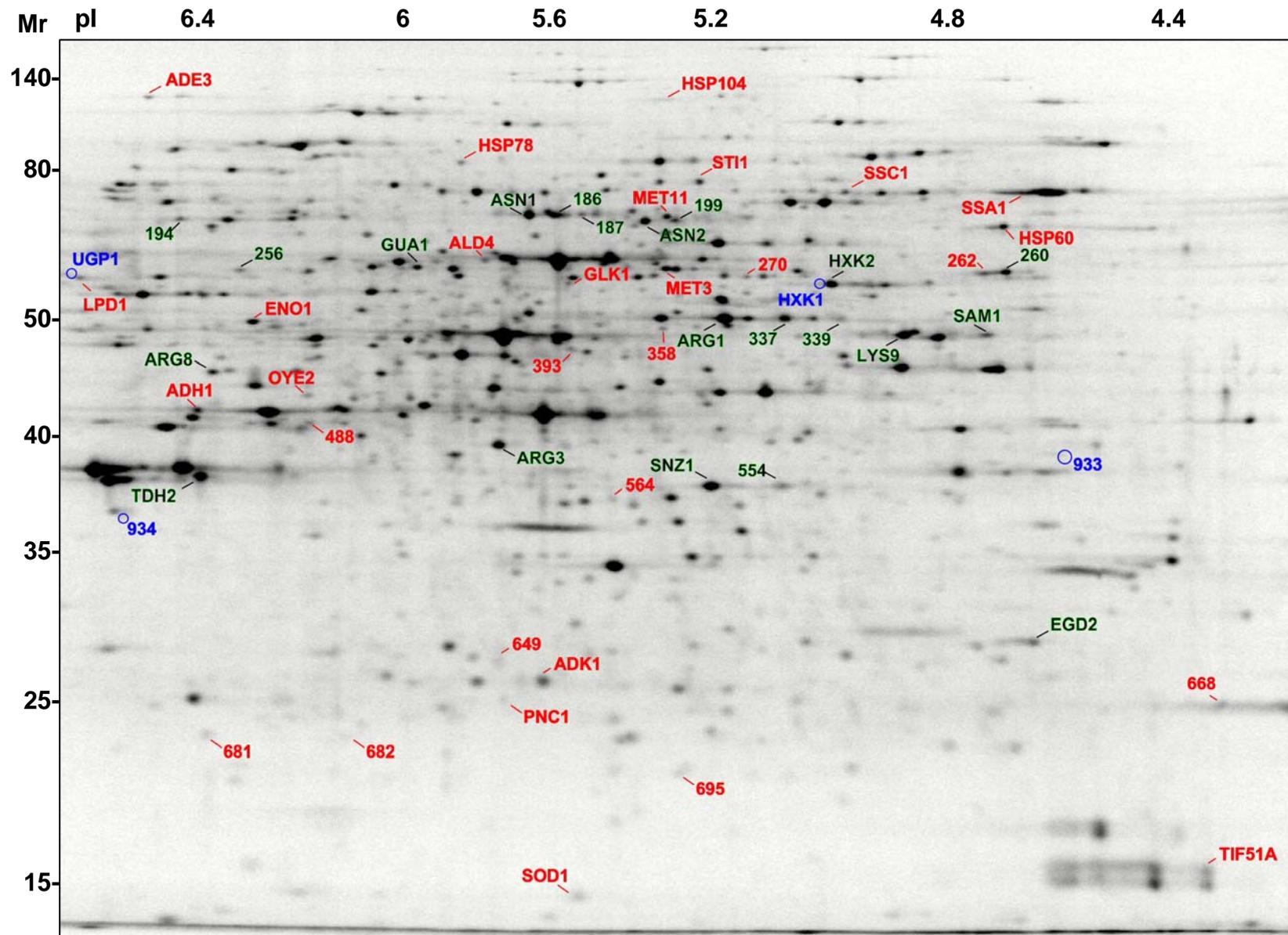
Proteins were labelled with [<sup>35</sup>S]-methionine in cultures grown to mid-log phase (OD<sub>600nm</sub> = 0.5). The map shows in black the spots corresponding to the control cells grown at 25°C and indicates in colour the proteins whose expression level is altered by at least 2 fold in *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG T<sub>33</sub></sub> grown at 37°C: in red, proteins that are induced; in green, proteins that are repressed; and in blue, proteins that are uniquely expressed in cells ambiguously decoding the CUG codon, when grown at 37°C (page 288).

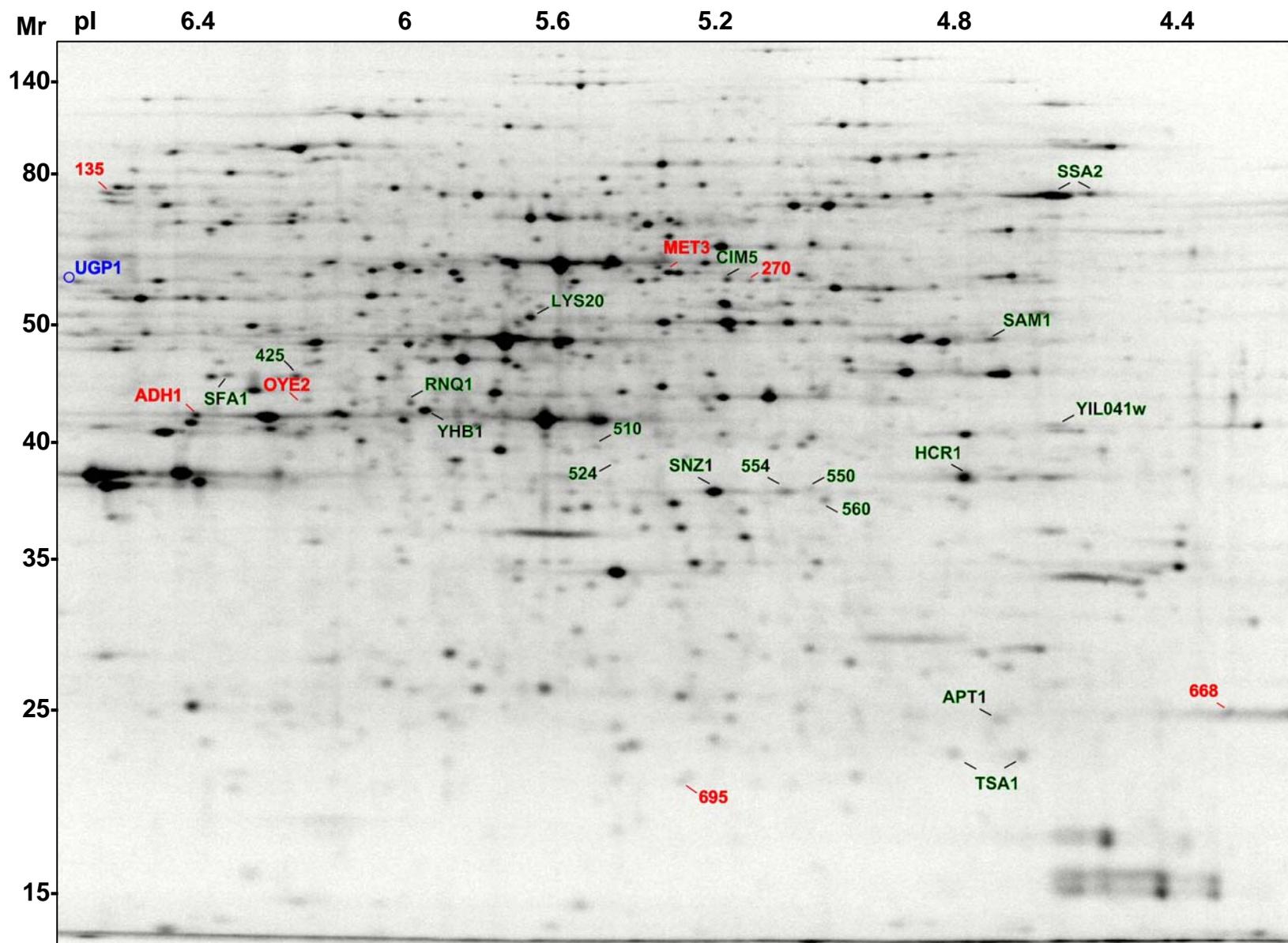
**Map 3 – 2D-map of proteins from *S. cerevisiae* control cells, induced after heat shock.**

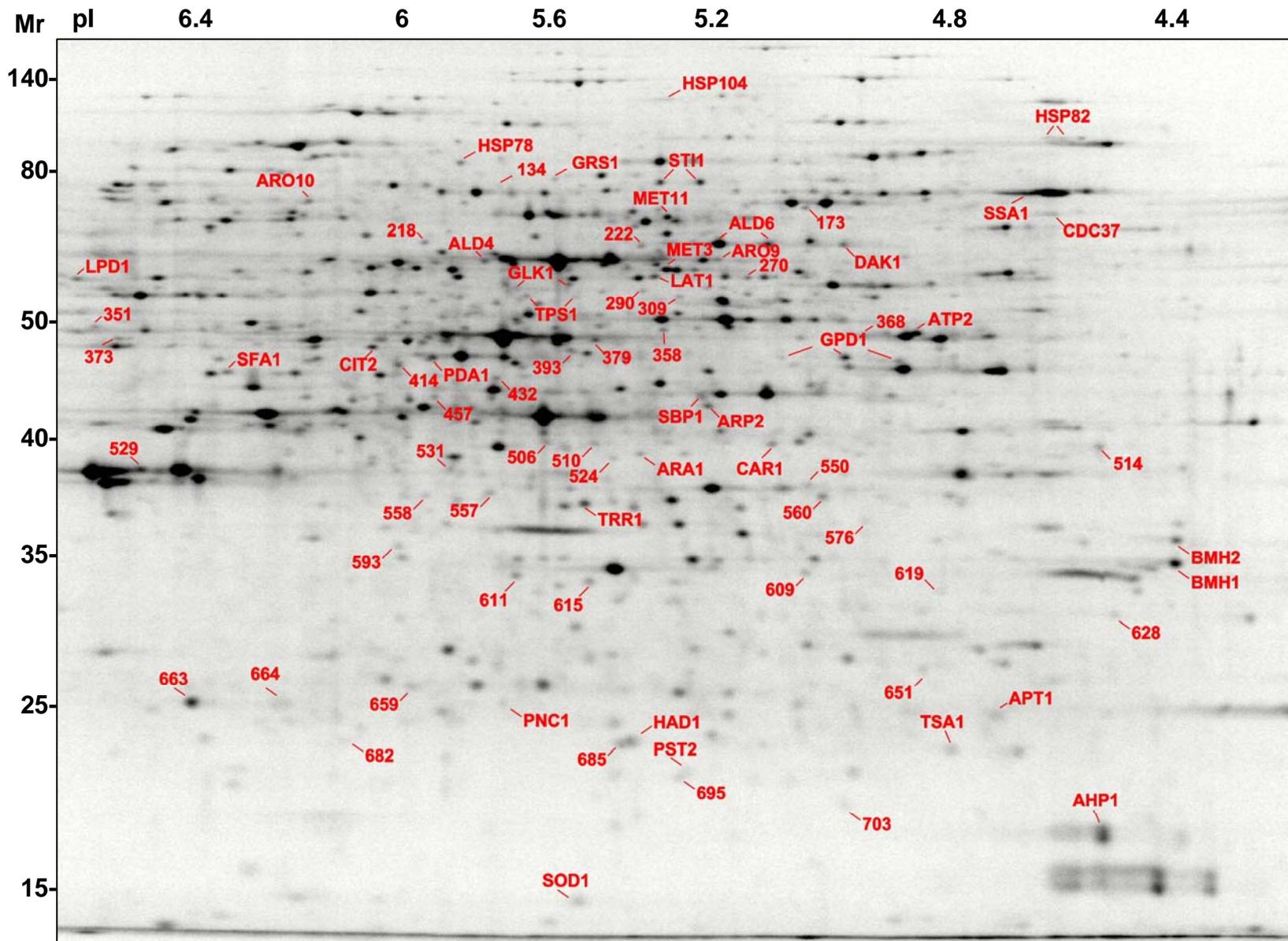
The map shows in black the spots corresponding to the control cells grown at 25°C and indicates in red the proteins whose expression level is increased by at least 2 fold in *S. cerevisiae* control cells after temperature upshift from 25°C to 37°C for 30 minutes. Proteins were labelled *in vivo* with [<sup>35</sup>S]-methionine (page 289).

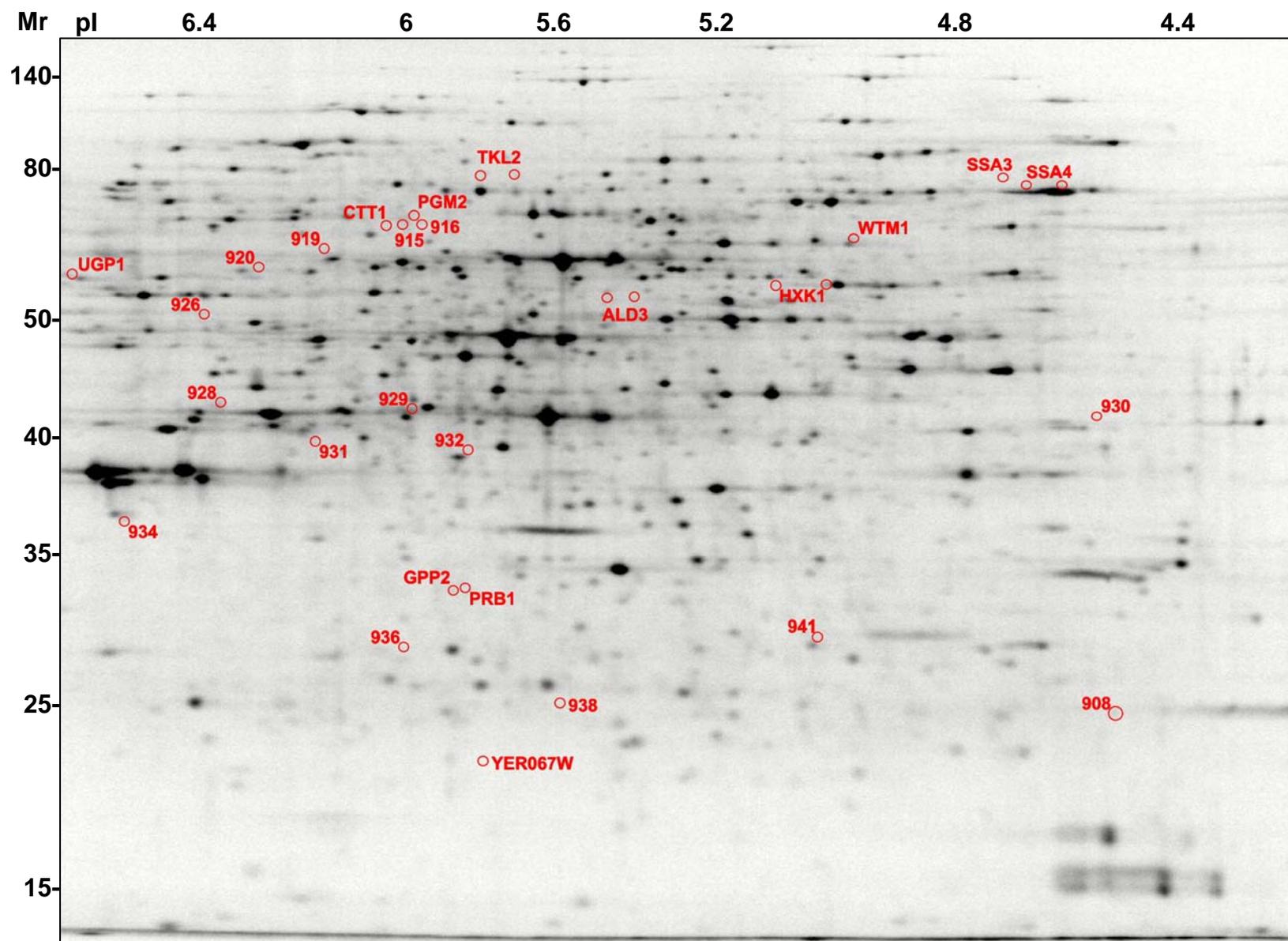
**Map 4 – 2D-map of proteins from *S. cerevisiae* control cells, uniquely expressed after heat shock.**

The map shows in black the spots corresponding to the control cells grown at 25°C and indicates in red the proteins that are uniquely expressed in *S. cerevisiae* control cells after temperature upshift from 25°C to 37°C for 30 minutes. Proteins were labelled *in vivo* with [<sup>35</sup>S]-methionine (page 290).









**Map 5 – 2D-map of proteins from *S. cerevisiae* control cells, repressed after heat shock.** The map shows in black the spots corresponding to the control cells grown at 25°C and indicates in green the proteins whose expression level is decreased by at least 2 fold in *S. cerevisiae* control cells after temperature upshift from 25°C to 37°C for 30 minutes. Proteins were labelled *in vivo* with [<sup>35</sup>S]-methionine (page 292).

**Map 6 – 2D-map of proteins from *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>, induced after heat shock.** The map shows in black the spots corresponding to the control cells grown at 25°C and indicates in red the proteins whose expression level is increased by at least 2 fold in cells ambiguously decoding the CUG codon after temperature upshift from 25°C to 37°C for 30 minutes. Proteins were labelled *in vivo* with [<sup>35</sup>S]-methionine (page 293).

**Map 7 – 2D-map of proteins from *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>, uniquely expressed after heat shock.** The map shows in black the spots corresponding to the control cells grown at 25°C and indicates in red the proteins that are uniquely expressed in cells ambiguously decoding the CUG codon after temperature upshift from 25°C to 37°C for 30 minutes. Proteins were labelled *in vivo* with [<sup>35</sup>S]-methionine (page 294).

**Map 8 – 2D-map of proteins from *S. cerevisiae* cells expressing the *C. albicans* ser-tRNA<sub>CAG-T<sub>33</sub></sub>, repressed after heat shock.** The map shows in black the spots corresponding to the control cells grown at 25°C and indicates in green the proteins whose expression level is decreased by at least 2 fold in cells ambiguously decoding the CUG codon after temperature upshift from 25°C to 37°C for 30 minutes. Proteins were labelled *in vivo* with [<sup>35</sup>S]-methionine (page 295).

