Intracellular detection and functional impact of protein synthesis errors in *S. cerevisiae* and *E. coli*

Connor Sampson

Abstract

Accurate translation of the mRNA code is vital for both proper cell function, and the production of functional recombinant protein products. Errors in translation, i.e. the incorporation of amino acids other than those that should be incorporated according to the genetic code, negatively affect the function of proteins. This thesis focuses upon two main areas.

1. Possible error detection pathways in S. cerevisiae are examined through the use of knockout strains and inhibitor treatments upon both optimised and dis-optimised GST genes. None of the surveillance pathway mutants used in this study showed detectable changes in the expression levels or in the detectable degradation products expressed from the recombinant GST genes, indicating that these pathways are not involved in detecting translational error products. However, the sensitivity of the western blotting techniques used in this study is limited, and possible alternate approaches are discussed.

2. An MS-based analysis of translational errors and their effects upon binding affinity in GST was carried out prior to this project. The results are examined herein through computational analysis, based upon factors such as 3D modelling, evolutionary conservation, and predicted tRNA ratios. Additionally, several translational error products were re-created as DNA mutants in order to allow experimental analysis of the effects of errors on protein function. These two approaches demonstrate the importance of protein folding upon error observation.

1. Introduction

When the Central Dogma of Molecular Biology was proposed by Francis Crick in 1958 it became rapidly apparent that the proposed pathways of information transfer were an oversimplification of true function, as seen in Figure 1.1 [1]. Errors play a major role in this system; for example DNA replication errors allow evolution, while mRNA translation errors affect, among other things, proteome function and speciation.

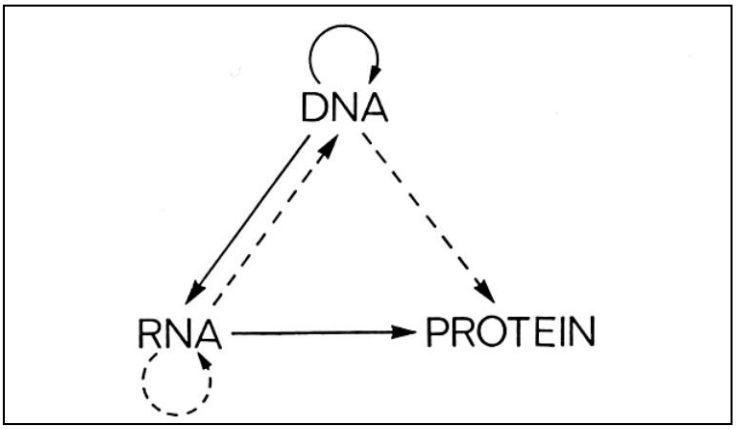


Figure 1.1: The Central Dogma of Molecula Biology, as drawn by Francis Crick. This thesis will concern itself with errors upon the "RNA→Proten" path of information transfer [1].

This thesis focuses upon translational errors; the loss of information within the "RNA → Protein" section of this diagram.

1. 1. Decoding Errors

Moving from the "4 letter" RNA code (A, C, G, U) to the "20 letter" amino acid code requires two accurate steps; the loading of amino acids onto their cognate tRNAs via aminoacyl-tRNA synthetases (aatRNAs) [2][3][4], and pairing the resulting aminoacyl-tRNAs (aatRNA) with their cognate mRNA codons [5][6]. As shown in Table 1.1 and Figure 1.2, both steps may be responsible for translational error and must be carried out faithfully for proper codon recognition to occur [6][7]. *In vivo*, tRNA molecules are misacetylated at a rate of around $\sim 10^{-4}$ and mRNA codons are misrecognised at a rate of around $\sim 10^{-4}$ [2].

	Translation Error	**tRNA Misacylation**	**Successful decoding**
Cognate amino acid bound to tRNA?	Yes	No	Yes
Cognate tRNA bound to mRNA anticodon?	No	Yes	Yes
Accurate decoding?	**No**	**No**	**Yes**

Table 1.1: Possible outcomes of a decoding event, note that both accurate tRNA loading and mRNA translation are required for correct amino acid insertion. These possibilities are shown diagrammatically in Figure 1.2.

Simple systems are often described by a negative speed/accuracy correlation, whereby proofreading allows greater error rates as overall speed increases. This makes intuitive sense and is the general case when examining direct molecular interactions [8], however this is not applicable to overall process of translation.

The transition from mRNA to Protein is a highly complex process, involving 20 amino acids corresponding to 20 tRNA species [9] (not including the non-canonical selenocysteine [10]

and pyrrolysine [11] which are synthesised through distinct mechanisms), containing a variable number of tRNA isoacceptors (41 in the case of *S. cerevisiae* [12]), stemming from multiple tRNA genes (274 in the case of *S. cerevisiae* [12]), decoding 64 possible codons. This thesis focuses upon specific missense errors, in which codons are recognised by incorrect aatRNAs; however these do not act in isolation and as such a range of major error types are included in this introduction.

Errors in the gene expression process, true to their complex nature, have a complex range of effects upon the cell. They can be detrimental to cell health [13], have been implicated in human disease [14], and are generally selected against through evolution [15], but can function as a cell stress response [7], confer cell benefits such as antifungal resistance [16], and be tolerated in normal growth [7].

It may be noted that certain translational alterations, commonly included under the title of translational errors, are, in fact, programmed cell functions; for example, many cells (including Bacterial, Fungal, and Mammalian examples) undergoing oxidative stress will load Methionine onto non-methionyl tRNA [17]. This class of translational error will not be discussed.

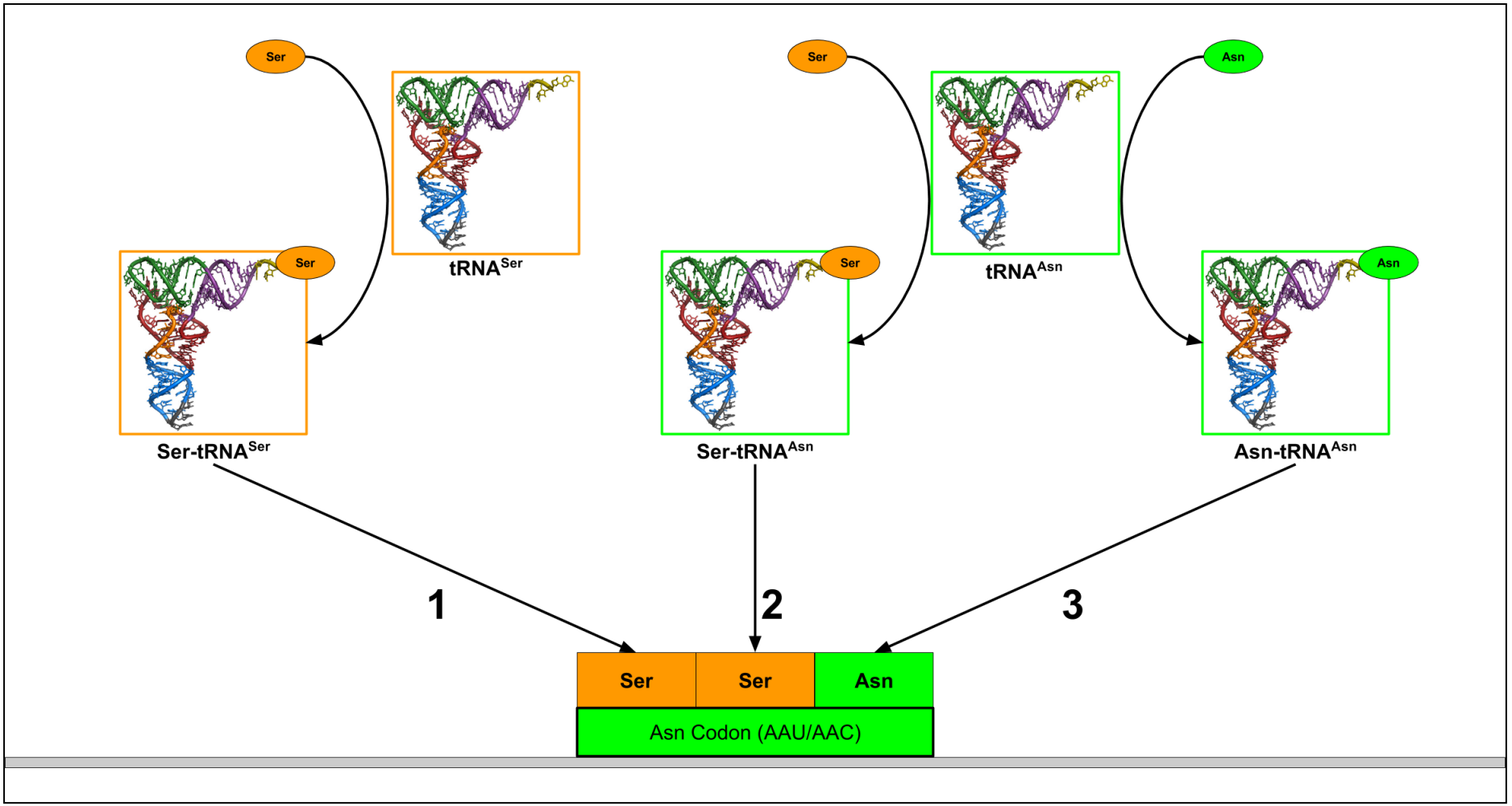


Figure 1.2: Possible outcomes from the decoding of an Asn Codon, as shown in Table 1.2, where errors at both the acetylation and the translation stage can be seen to insert a mismatched amino acid in the same manner. The loading of Ser onto $tRNA^{Asn}$ (2) is known to cause Ser insertion at Asn codons in industrial recombinant CHO cell lines [158]**.** ***tRNA image credit: Wikimedia Commons/Yikrazuul, based on PDB 1EHZ.***

1.2. Error Sources

1.2.1 Elongation Errors

Multiple forms of error may occur during the elongation of a nascent peptide chain, a general categorisation of which can be seen in Figure 1.3. Missense errors represent either the insertion of an incorrect aatRNA, leading to amino acid substitution, or the premature insertion of termination factors [18][19]. Processivity errors occur when the ribosome fails to make normal progress along the mRNA, taking the form of either premature termination [19], or a translational frameshift [20]. These error types are examined below.

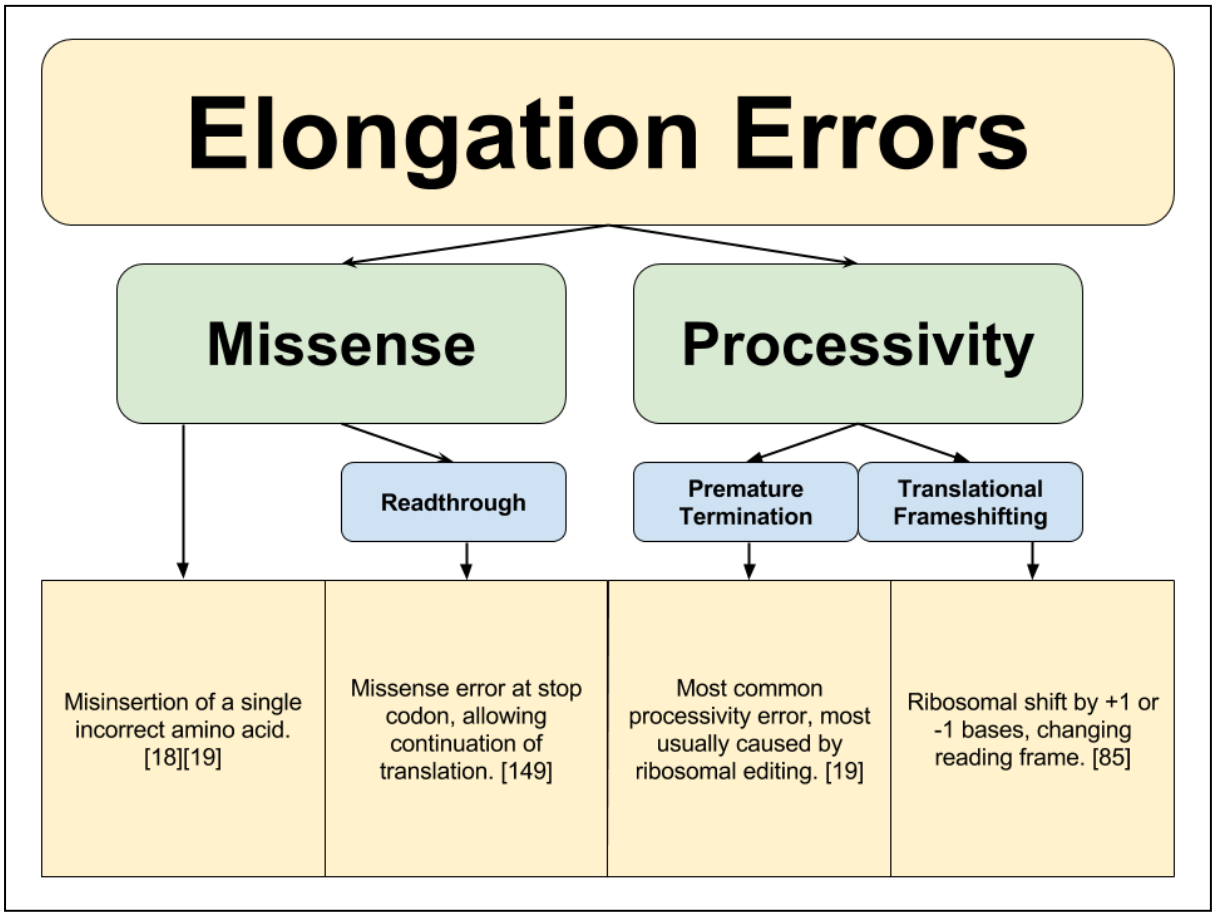


Figure 1.3: Forms of elongation error, as defined by Farabaugh et al 1999 [19] **.**

1.2.2 Missense Errors

Missense errors appear to be the most straightforward error type seen in Figure 1.3 however their precise causes and effects can be highly complex.

A missense error is the insertion of one amino acid in place of another in the developing polypeptide chain. This may be due to either tRNA misacylation (see *1.2.4 tRNA Misacylation*), or the insertion of a non/near-cognate aatRNA.

As the insertion of a an incorrect aatRNA is based upon codon mismatch, it is tempting to assume such errors may be understood entirely through the genetic code. A paper by G. Wu et al 2005 [21] attempts to do just this. Using a statistical model they assume that codons may pair if two out of three bases match, and that errors simply arise through similar tRNA anticodons competing equally for a given codon. Using this model, each possible error event for a codon is generated (i.e. all first base, second base, and third base mutations), and matched to the relevant amino acid. This results in a statistical distribution of error probabilities (e.g. for ACU there is a 6/12 chance of Threonine, 2/12 chance of Serine, 1/12 chance of Alanine, etc.). This may be applied to an entire mRNA and compared to experimentally observed errors. If this hypothesis of error occurrence were true, we would expect to see a strong match between predicted and observed errors [21].

The results, seen in Figure 1.4, show only general correlation, with significant divergence between predictions and experimental readings. This demonstrates that while codon similarity is an important factor in this error type, there is also dependence upon factors more complex than mere codon similarity.

A more accurate, and complex, model is presented by A. Fluitt et al 2007, incorporating a stronger basis in biological process [6].

It is well known that elongation rate *in vivo* is dependent upon the concentrations of suitable aatRNAs [22][23] indicating the rate limiting step in protein synthesis to be the “wait” period as a suitable aatRNA enters the ribosomal A site [6]. As aatRNA recruitment to the A site is thought to be dependent upon random diffusion, it is possible to estimate both the average recruitment time per aatRNA and the likelihood that an incorrect tRNA will reach the A site [6].

An arriving tRNA can be cognate, non-cognate, or near-cognate; where near-cognate species have an incorrect anticodon which nevertheless exhibits some complementarily with the mRNA codon, resulting in longer rejection times and a higher likelihood of misincorporation [24]. Near-cognate codon binding was originally assumed to depend upon correct pairing of the first two bases (where most discrimination is thought to occur), allowing third base error in the interaction. This assumption was used by A. Fluitt et al and other modelling studies [6][25] .

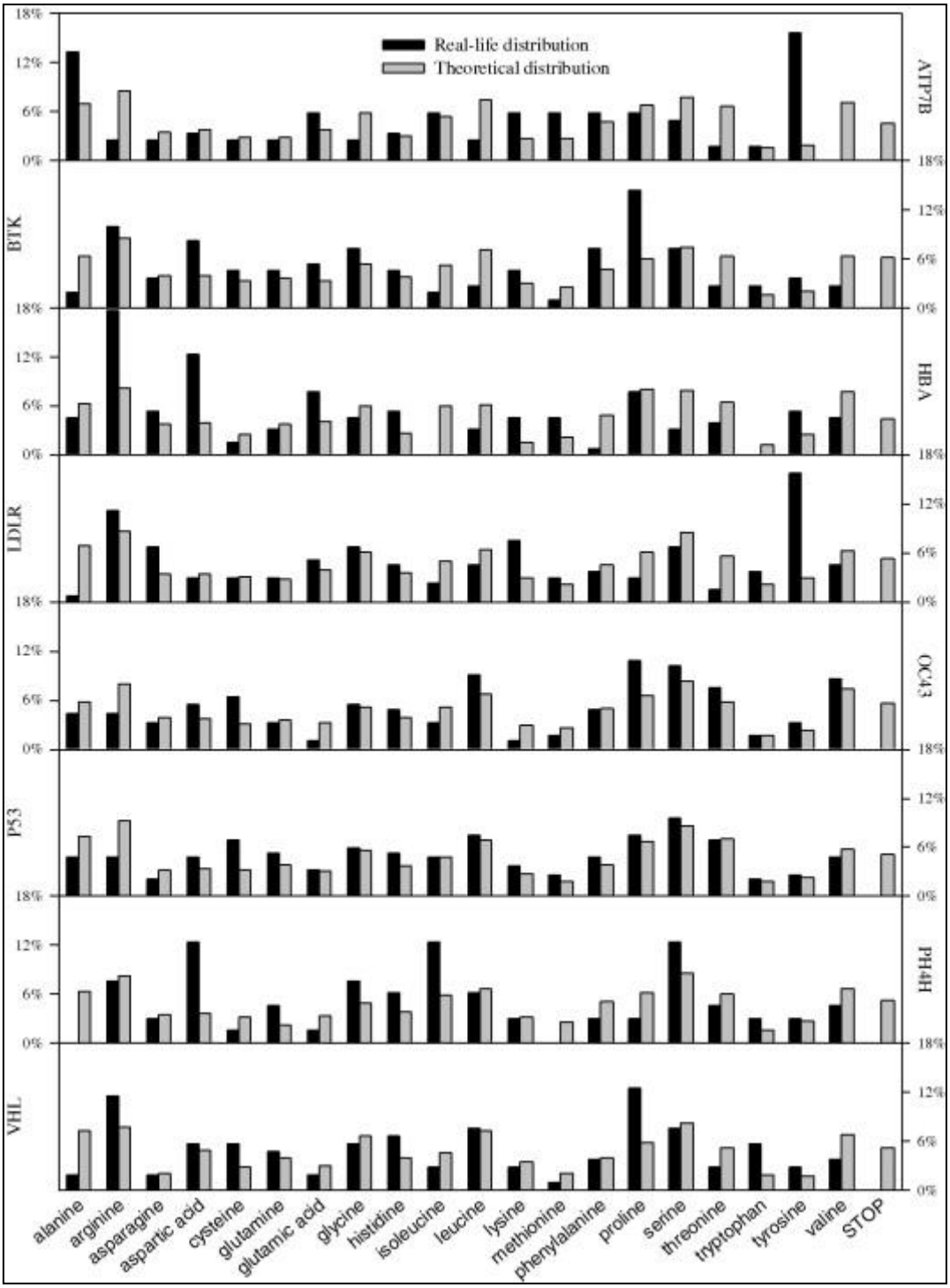


Figure 1.4: Predicted and actual mutation rates for a selection of high error proteins. Data generated assuming errors to be based on codon similarity alone, showing rough correlation with observed data. This indicates missense errors to be influenced by codon similarity, but dependent on further mechanisms. ***Diagram taken from Wu et al 2005*** [21].

Experiments based on a Luciferase reporter system updated this definition, wherein an active site Arg is mutated and only a missense insertion of the correct amino acid will produce a functional protein (a process known as missense suppression) [26]. This work has shown that formation of the min-helix between the codon and anticodon is vital for activation, and an unpaired third base will prevent this process. Instead a weaker transient mini-helix must be formed for near-cognate insertion. In this model a strong second base interaction is required, around which the first and third bases may form weaker hydrogen bond interactions with their anticodon counterparts, thus allowing the insertion of the near-cognate base [26]. This appears to be due to ribosomal monitoring bases A1492, A1493, and G530. It is considered in this paper that interactions with the mini-helix allows monitoring bases to alter confirmation from the Off to On positions [26], however recent work suggests that this monitoring triad does not, in fact, interact with the mini-helix codons. Instead this triad is responsible for water exclusion [27]. Under this model, explained in *1.3 Monitoring Bases A1492, A1493, and G530*, the mini helix itself is more indirectly vital in the acceptance of near-cognate aatRNAs. A non-cognate aatRNA, only binding with the first two bases, would have significant unpaired hydrogen bonds upon desolvation, while the near-cognate tRNA, with its weak hydrogen bonding in positions one and three, is far less vulnerable to desolvation energy barriers [26][27].

This being said, whilst the initial definition of near-cognate is now seen as incorrect in the work of A. Fluitt et al, their findings, based upon experimental data and mathematical modelling, are still highly relevant to understanding this area. It is seen that translation rates, due to time spent rejecting incorrect tRNAs, are dependent primarily upon relative aatRNA proportions within the cell, as opposed to absolute aatRNA levels. It is also seen that codons under high competition are more likely to suffer missense errors [6].

Therefore, it can be seen that a fast decoding (accurate) codon is dependent on a high relative concentration of cognate aatRNAs, while a slow decoding (inaccurate) codon is due to a low relative concentration of cognate aatRNAs [6][24], and that near-cognate competition is based upon mismatched codon pairs forming transient mini-helices, allowing the desolvation barrier to be overcome.

Through this process we see a positive speed-accuracy correlation, although it should be noted that the increase in both speed and accuracy is due the decreased likelihood of non-cognate and near-cognate aatRNA entry into the A site. This speed-accuracy relationship, whilst useful, is therefore not directly causative.

1.2.2.1 Missense Errors - tRNA balance importance

As seen above, cellular aatRNA balance is highly important for translational efficiency. Accordingly, we would expect to observe evolutionary pressure to increase translational speed and reduce errors (missense, along with associated errors such as ribosomal editing [28][29] and possible +1 frameshifting [30]) by correlating aatRNA levels to codon usage.

A comparison of *S. cerevisiae* and *E. coli* aatRNA and codon abundances indicate just such a pattern. In both species tRNAs levels correlate strongly with codon choice, whereby more

common codons are represented by a larger population of cognate tRNAs. These levels, however, differ greatly between the two species and as such a different balance of codons is found to be optimal in each case [31]. This demonstrates strong evolutionary pressure to maintain the tRNA/codon correlation with regards to cell codon requirements, minimising translational delays and errors through the managing of competition effects. Furthermore this evolutionary pressure is found to exist in both prokaryotic and eukaryotic cells, demonstrating a fundamental feature of translation.

In the above study tRNA abundances were directly sampled, allowing clear and reliable results. It is possible, however, to estimate tRNA abundances indirectly, whereby tRNAs with a higher isogene number will be more highly expressed (see *1.2.2.2 Missense Errors - tRNA level maintenance*). Using this approach, the tRNA/codon correlation can be observed within a wide range of unicellular organisms, further supporting the above conclusions [32].

It may be noted that optimisation of the aatRNA/codon balance is only possible for DNA sequences with only one host. For example an interspecies plasmid, 2μ, exists within multiple species [33] is and therefore is exposed to differing aatRNA balances [31]. As such this interspecies plasmid shows no codon optimisation, demonstrating the ability of cells to make use of non-optimised genes. Codon optimisation is therefore preferable, but not essential for gene expression.

Codon optimisation can also take the form of "fine tuning" within multiple cells of the same organism. Humans, for example, have a range of tissue types, each with their own translational needs. As such varying human tissues show varying tRNA levels to match their major gene products and, interestingly, varying codon preferences [34], showing each differential tissue to be treated as a separately optimised translational system acting under its own selective pressure to achieve maximal efficiency.

Previous studies, however, have noted that a simple codon / aatRNA analysis in eukaryotes shows a weak correlation in many cells, and have proposed that in many "more complex" cells, other factors demonstrate a higher selective pressure than aatRNA / codon balance [35], such as a gene length / codon correlation [36] Further examination, however, reveals the importance of tRNA modification in the reading of the genetic code.

tRNA modification plays a major role in codon recognition, especially modifications to the third "wobble position" codon base, known as base 34 [35][37]. Modifications here are used heavily by the three domains in order to adjust codon recognition by the tRNA pool (see *1.2.2.4. Missense Errors - tRNA Activation & Degradation*) [37]. These modifications, allowing the recognition of a wide range of codons by a smaller number of tRNAs [35], once taken into account, restore a high level of correlation between codon usage and tRNA balance across all domains, conserving codon optimisation as a highly conserved fundamental selective pressure, even in large, slow growing, organisms [35].

It may be noted that as codon optimisation contributes to both translational error reduction and decoding speed improvement, either effect may provide the dominant selective pressure. Examination of Intrinsically Disordered Regions, where folding does not take place, shows a

lowered level of codon optimisation compared to conventional, folding, protein regions. This suggests a lower selective pressure in these regions, where substitution errors are more easily accommodated, and as such shows clear contributions of translational errors to the system [38].

It is therefore seen that across domains, from unicellular organisms to mammalian tissue, codon optimization is a highly important, fundamentally conserved process, highlighting the importance of managing these errors as part of basic translational maintenance. Accordingly, it is important to understand what factors, both natural and artificial, may affect this tRNA balance and thus impact translational efficiency.

1.2.2.2 Missense Errors - tRNA level maintenance

As mentioned previously, each tRNA is encoded by multiple gene copies. As tRNAs are non-translated transcripts, abundance is largely dependent upon the number of genes present for each tRNA [32][39].

One method to confirm gene copy number of as a major determinant of tRNA levels in a cell is through observation of evolutionary change. As stated in *1.2.2.1 Missense Errors - tRNA balance importance*, there is strong evolutionary pressure for cells to correlate tRNA balance with codon usage, with values varying heavily between cells. Accordingly, in horizontal gene transfer the introduction of foreign genetic material is likely to introduce genes not adapted to the cell system (possessing a low Codon Adaptation Index / CAI [40]). In situations where this foreign genetic material places a significant burden on the cell, the tRNA/codon balance must be restored. The evolutionary pressure for this process derives from ribosome sequestration on the low CAI genes and similar mechanisms reducing overall cell growth [39][41].

The impact of low CAI gene introduction affects both the frequency of horizontal gene transfer and the likelihood that transferred genes will be fixed into the host genome [42]. A cell receiving a gene rich in low CAI codons may, however, modify the codon usage within that gene through codon selection, whereby unfavourable codons are mutated into codons better adapted to the host cell [39][43]. A strong indicator that this process is driven by translational efficiency as a selective force is the observation that more highly expressed genes show higher rates of codon adaption [43].

This mechanism of adaption shows the host tRNA levels as fixed and unchanging, however when non-optimised genes are inserted into the cell it is also possible for the cell to respond by altering tRNA gene copy numbers; directly altering the local tRNA balance [39]. This is plausible as tRNA gene duplication and deletion is a well established evolutionary process [44] indicating gene copy number to be the primary controller of tRNA concentrations within the cell.

It should be noted that while tRNA gene copy number and codon usage coevolve, it is suggested that genetic constraints on tRNA gene copy number make this a less favourable long term strategy when inserting new genes, and instead a likely forms a first step in a

proposed adaption cascade [39], as shown in Figure 1.5, demonstrating a complex evolutionary process focused on restoring translational efficiency. Another potential reason for codon adaption in place of long term tRNA duplication is that while a tRNA pool adaptation may increase overall efficiency, it decreases efficiency (and thus increases errors) in the original cell genes. Thus, for the introduction of discrete genetic material, the optimal long term strategy is codon adaption to match host DNA, driven by the evolutionary pressure for cells to maintain accurate and efficient translation.

Additionally, the effect of codon optimisation may be great enough to play a significant role in microbial speciation [43], demonstrating the extent to which translational errors and stalls determine the viability of an organism's genetic material.

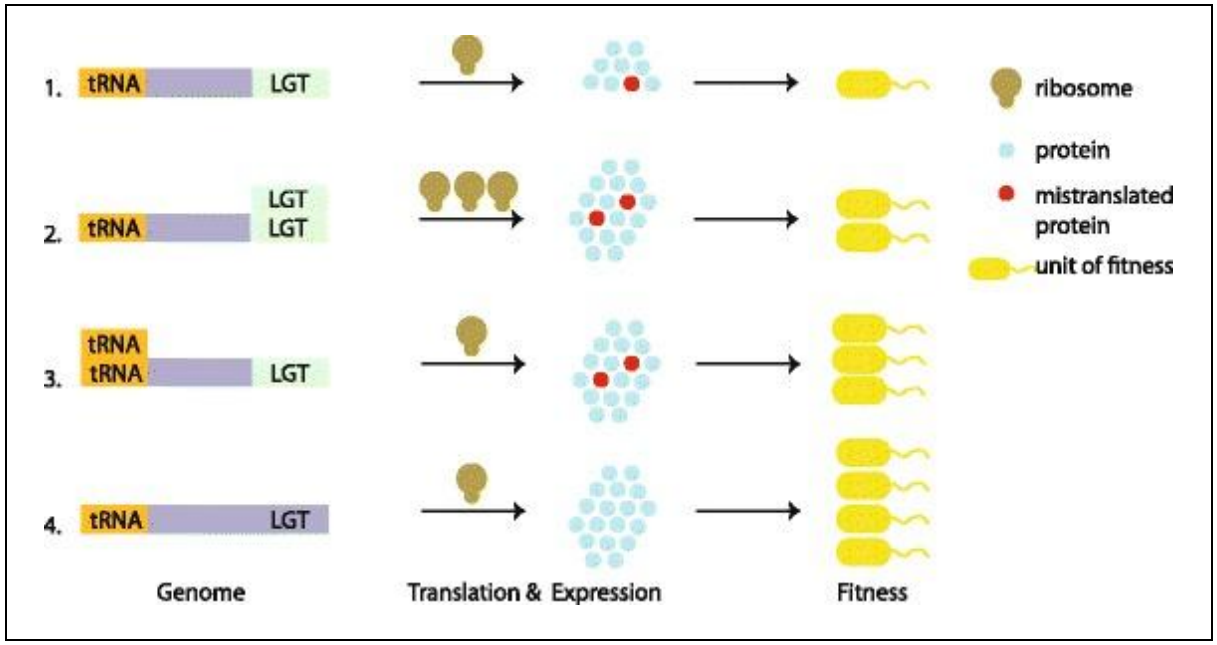


Figure 1.5: Adaptation to the introduction of low CAI codon genes, showing a proposed evolutionary pathway focused on the restoration of codon / tRNA balance, thus restoring efficient and low error translation to the system. 1. Lateral Gene Transfer introduces a gene which positively affects cell fitness, and is fixed into the culture. 2. Expression of the non-optimal gene is achieved through gene duplication, at the expense of inefficient translation and ribosome sequestration. 3. tRNA gene duplication restores some tRNA / codon balance and increases cell translation efficiency. 4. Codon adaption replaces non-optimal codons with optimal codons, allowing a return to the cell's original tRNA / codon balance, restoring optimal decoding to the system [39]. ***Image taken from M. McDonald et al. 2015*** [39]

It has therefore been seen that proper tRNA / codon balance is highly important within the cell, and that the tRNA levels are primarily regulated and altered through gene duplication, however deletion study has shown differing contributions of individual tRNA genes. The effect of a tRNA gene is determined by flanking DNA regions, likely due to pol III recruitment [45]. While this indicates an ability of the cell to fine tune tRNA levels, this mechanism is still based on seemingly unreactive genetic coding.

This study did, however, highlight some more dynamic homeostatic tRNA expression regulation. Deletion of tRNA genes often reduces the abundance of the corresponding tRNA, however in some instances the corresponding tRNA abundance remained stable or even increased, indicating a regulatory system similar to many protein coding genes [45].

Slight active variation in tRNA levels with respect to codon usage has been reported at various *E. coli* growth rates [46], however the increases seen were generally moderate. The study found an increase in high demand tRNAs at higher growth rates to match changing transcriptional needs, although the increases were not as great as would be expected for full tRNA / codon usage correlation [46]. It therefore appears that this pathway is a minor "fine tuning" event responding to growth in a programmed fashion, whereby tRNA pools may be non-adapting but moderately dynamic over organism growth. It should be noted, however, that the significantly different growth mediums used to achieve each growth rate in the above study represent different growth conditions, and as such it cannot be ruled out that alternate cell pressures may have affected the system in a manner confused with growth response.

To summarize the above points, tRNA levels are generally determined by the number of gene repeats for each tRNA species, wherein each tRNA gene contributes a genetically predetermined level of transcription. tRNA abundances are carefully maintained within the cell, and simple deletions of isogene repeats may not result in alterations of tRNA abundance. An alteration in codon levels requires an alteration of tRNA balance before codon adaption can restore the system, and this must occur through slow alteration of tRNA gene number. This is, therefore, a system based around consistency in translation quality as opposed to dynamic adaption and shows only a long term ability to respond to changing codon requirement. In the short term, a cell would be unable to react to sudden demands of low codon adaptation genes, resulting in low efficiency high error translation.

1.2.2.3 Missense Errors - Silk Glands

As with most biological systems, there is an exception to the normal mechanism. It should be noted that an example of strong dynamic tRNA is regulation is found within the silk glands of spiders such as *Nephila clavipe* [47][48]. During silk production the high levels of fibrin translation alter the codon demands in a tissue specific manner [47]. In order to main efficient, high accuracy translation, the levels of tRNAs cognate to spider silk are greatly raised.

The alanine specific tRNA in *Nephila clavipe* shows a novel regulatory system with two isoforms found within the genome; one constitutive and the other tissue specific [47][49]. Analysis of gene distribution shows distinct clustering of the tissue specific version, as opposed to the scattered constitutive version, and no amplification / rearrangement events surrounding increased tissue specific expression [50].

Most tRNAs are transcribed by Pol III, and utilise a Box A / Box B type (type 2) promoter, shown in Figure 1.6, with optional upstream motifs such as a promotional TATA box. This system gives rise to the normal behaviour discussed in *1.2.2.2 Missense Errors - tRNA level maintenance* [51]. The tissue specific Alanine tRNA, however, shows dependency on a far upstream Distal Sequence Element for promotion [48] as found in Type 3 promoters (Figure 1.6), an unusual feature of tRNA genes [51]. It can therefore be seen that this variable tissue specific regulatory system is both mechanistically and functionally distinct from ordinary tRNAs.

Ordinary tRNA transcription is focused upon stability in order to maintain low error rates and efficient translation. A system with highly changeable codon requirements may show adaptations such as the above discussed silk $tRNA^{Ala}$ or silk $tRNA^{Gly}$ [52] (not discussed here). It is notable, however, that dynamic tRNA expression requires a largely different regulatory system to that normally found in tRNAs, and gives an indication of the dramatic regulatory alterations required to adapt the behaviour of this system.

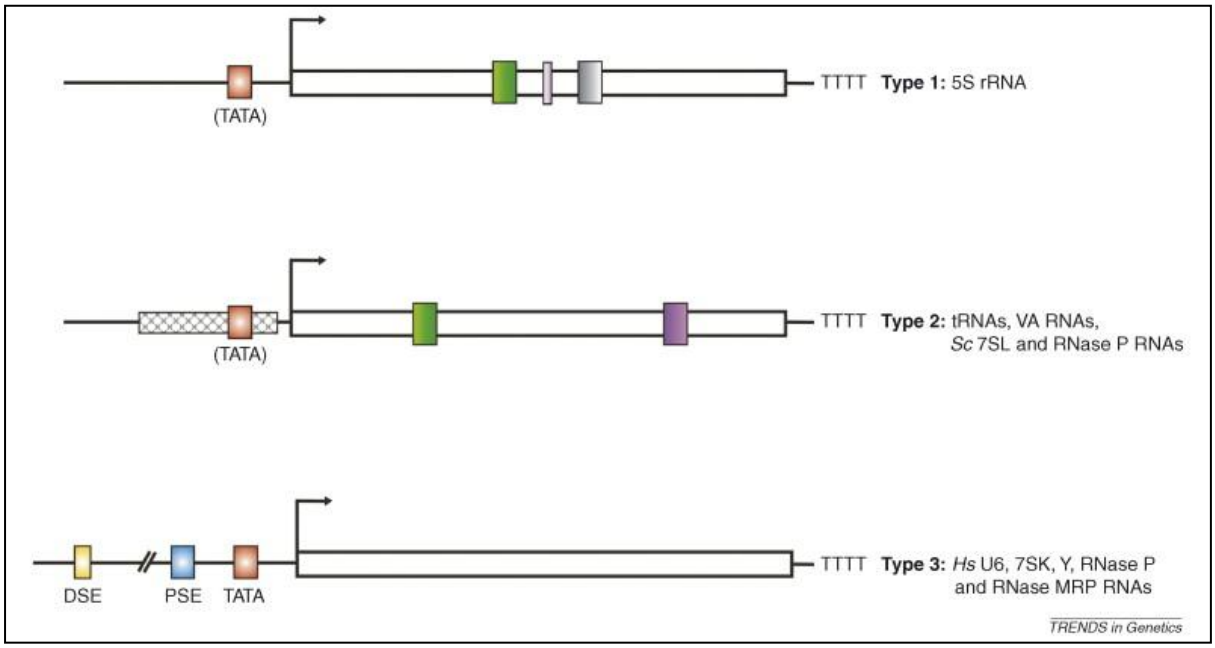


Figure 1.6: Major RNA polymerase (Pol) III promoter types, showing the Box A / Box B type (type 2) promoter normally seen in tRNA genes, and the DSE (type 3) found in the tissue specific tRNA Ala. Image taken from G. Dieci et al 2007. *"The approximate positions of upstream and internal control elements are indicated relative to the transcription start site (arrow). A, box A; B, box B; C, box C; IE, intermediate element; TTTT, transcription termination signal; PSE, proximal sequence element; DSE, distal sequence element; TATA, TATA box or TATA-like sequence. ... Parentheses indicate the fact that not all type 1 and type 2 promoters contain a TATA element. The upstream cross-hatched region in type 2 promoters reflects the species-specific conservation of composite sequence motifs upstream of tRNA genes."* [51]

1.2.2.4. Missense Errors - tRNA Activation & Degradation

Proper processing and charging of tRNAs is required before they become available for use and can count towards the cell's aatRNA balance. This is a complex process involving RNA processing [53] and modification of the nucleotide bases (increasing EF-1α·GTP binding affinity two fold) [54], in addition to charging with a correct amino acid (increasing EF-1α·GTP binding affinity 10^3-fold to 10^4-fold) [54]. As such, the active tRNA concentration within a cell depends not simply upon transcription, but also upon the downstream processes shown in Figure 1.7.

For degradation Figure 1.7 shows multiple quality control pathways, and a variable, stress related, degradation pathway; endonucleolytic cleavage [53][55].

Conditions giving rise to oxidative stress, particularly in the stationary phase, are known to generate tRNA fragments cleaved through the anticodon loop through as a eukaryotically

conserved process [56]. The generation of these half tRNAs is known to suppress translation within the cell, [57][58] however whole tRNA levels remain stable until after half tRNA levels have peaked [56]. As such tRNA cleavage does not affect translation through tRNA depletion, but rather through a direct effect of half tRNA fragments displacing translation initiation machinery (eIF4G/eIF4A and eIF4F) from mRNAs [58]. As such the process does not affect active aatRNA levels within the cell, demonstrating a mechanism of translational control fundamentally unrelated to aatRNA levels and error management.

Another important aspect of Figure 1.7 is the charging of tRNAs with amino acids, with respect to the amino acid loss through translation. Intuitively, a high enough demand for certain aatRNAs would deplete the system's acylation abilities for that species and alter the active aatRNA concentration, with implications for translational errors. Such a system has been considered, however simulations have shown that, generally, tRNA depletion does not affect active aatRNA species [59]. At extreme levels of translation, here simulated as 50% of ribosomes translating a single tRNA species, a non-optimised codon interacting with low abundance tRNA would be 10% less efficient than an optimised codon interacting with high abundance tRNA. This level is not achievable within *S. cerevisiae*, only found within certain biological circumstances and some modern recombinant systems, such as slow growing mammalian cells [59]. As such, tRNA charging and discharging is not a common contributor to aatRNA balance, and thus has little impact upon general translational errors outside of a specialised system.

A final area to consider is that of processing and modification. In yeast cells, oxidation may trigger a stress response which directly modifies $tRNALeu^{(CAA)}$ to , $tRNALeu^{(\underline{C}AA)}$ with a m^5C modification upon the wobble base [60]. This modification increases affinity for UUG codons, thereby increasing accurate decoding speeds upon the UUG codon and significantly upregulating UUG rich genes. In the current example this results in a change of ribosome composition, and a combination of altered the tRNA and ribosomal natures stimulate increased translation of stress response genes [60].

This adaptive process is demonstrative the extensive role of tRNA modification, especially wobble base modification, in the differential recognition of codons by tRNA molecules. This system is further seen in the differing decoding strategies, and corresponding tRNA codon modifications, found in the three domains of life, as examined in H. Grosjean et al. 2010 [37].

To quote the paper directly "Both decoding accuracy (limiting 'cross-box' misreading) and efficacy (increasing translation rates and avoiding frameshifts) are very much depending on specific tRNA modification enzymes transforming genetically encoded canonical A, C, G, U nucleotides of the precursor tRNA transcripts into a large variety of chemically altered derivatives with innovative structural and decoding potentialities" [37]. This is particularly clear in its effect upon aatRNA/codon balance within cells, as mentioned in *1.2.2.1 Missense Errors - tRNA balance importance* [35]. Accordingly tRNA modification is a major factor in decoding and fidelity.

It is therefore seen that while tRNA processing does not does not limit aatRNA levels in a normal system, processing may instead affect aatRNA behaviour. This may be through anticodon modification altering activity (as above) [60] resulting in effects similar to aatRNA concentration change, or by directly increasing the recognition of cognate over non/near-cognate codons [61] and reducing frameshifting [62].

Two relevant features are therefore seen in the processing pathway. Firstly, general processing and charging represent a robust system, designed to maintain tRNA levels effectively, leaving tRNA gene number as the major controlling factor. Secondly, base modifications are highly important for the maintenance of fidelity, showing a more dynamic and adaptable control system than gene number, allowing fine tuning of aatRNA activity in order to maximise efficiency of codon recognition under differing conditions.

Or, to generalise greatly: tRNA gene number controls tRNA balance on an evolutionary level, tRNA processing works as a non-limiting step to maintain the gene dictated balance, and tRNA modification allows fine tuning in a more dynamic manor. Together, this system allows precise and stable suppression of translational error, and a maintenance of translational efficiency. It does not, however, allow dramatic adaption of aatRNA activity in response to sudden changes of codon balance (e.g. the introduction of low CAI genetic material.)

1.2.3 Codon Context

The concept of codon bias has been much discussed above, however there also exists codon pair biases found in all domains of life, even when dipeptide effects are removed. As these are enriched in coding regions, they appear to affect the translation process [63]. Contextual codon biases have long been observed [64] along with the effect of a codon's neighbours on translational accuracy [65][66][67].

These codon pair biases, while occasionally not present in smaller genomes, or altered in certain organisms, are found throughout all domains of life, indicating a fundamentally conserved feature of the mRNA decoding process [63]. While these biases are more subtle than codon biases [63], they behave in a largely similar manner with exact biases varying between organisms [68] due to genetic drift [63] and species specific factors [69].

This system is further complicated by the interaction of individual bases in neighbouring codons. Positions "P-site 3" and "A-site 1" always show pair bias; however many other pairs between the P and A sites show additional biases in a range of species. As a result, base specific interactions influence codon selection. These biases are likely due to variation around the 45° mRNA kink between P- and A-sites, subtly altering the geometry of interacting aatRNA molecules [68].

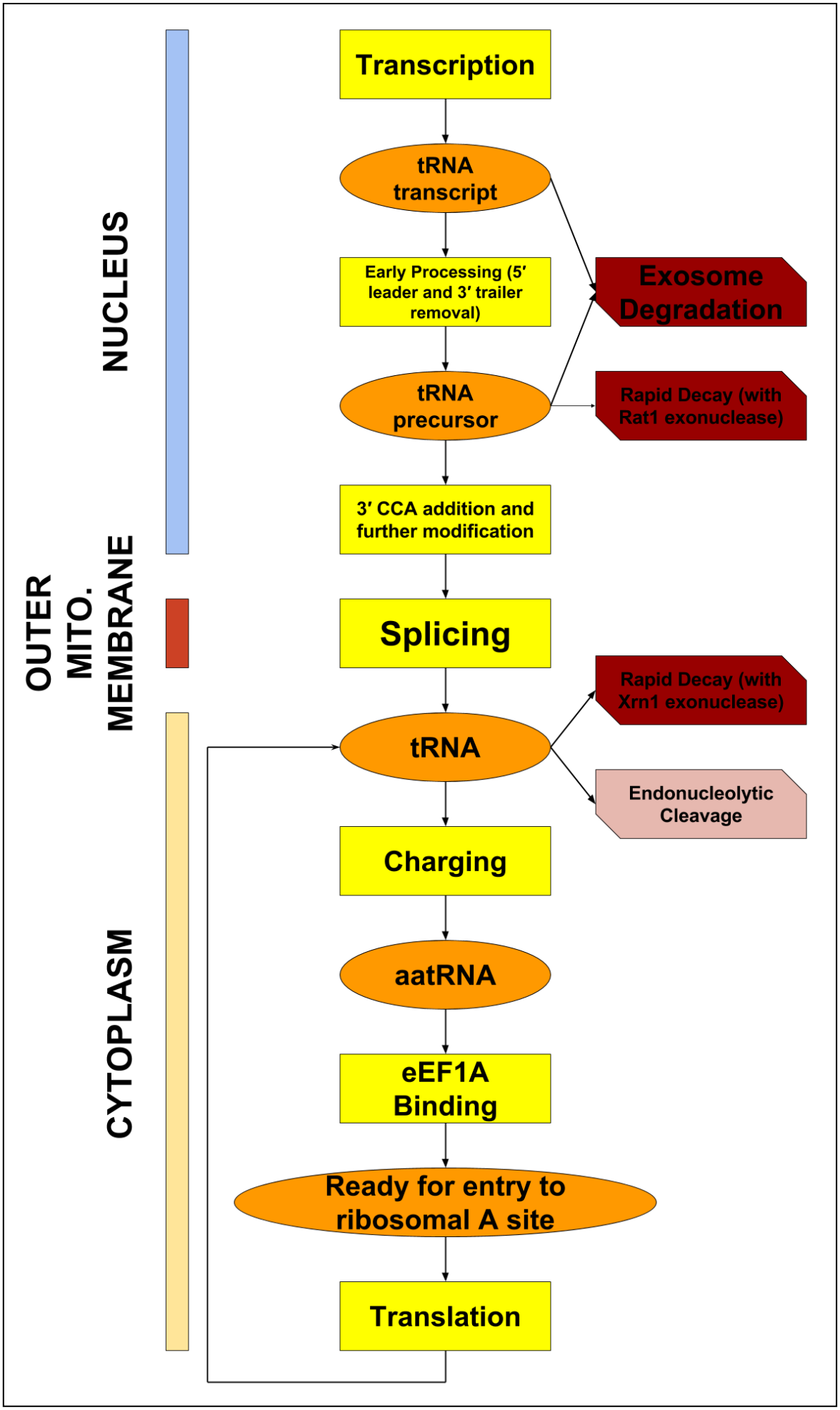


Figure 1.7: General overview of tRNA processing. The tRNA transcript undergoes extensive modification before a mature tRNA molecule is produced. Specific cases may not necessarily follow the above sequence, such as common tRNAs without introns [53] **or rare tRNA gene organisations such as split genes** [159] **and coding for the 3′ sequence upstream of the 5′ sequence** [160]**, both of which require specialised splicing processes. Dark red boxes show quality control degradation steps** [53][55]**. The endonucleic cleavage represents a variable stress response degradation as discussed in the main text** [53][56]**.**

Due to increased pressures in highly expressed genes, CAI and pair selection show a very slight positive correlation in Eukaryotic organisms, however in bacterial genomes a negative correlation is seen as high frequency codon pairs often contain low frequency codons, slowing transcription in high expression genes [68].

Codon pair selection is an important process in gene optimisation, affecting various errors, translation efficiency, and translational speed. It gives perspective to the importance of codon / aatRNA balance discussed here that pair selection is generally secondary to codon / aatRNA balance, representing a trade off between these two optimisation systems.

Additionally, as codon pair selection has the potential to greatly affect translation rates, cellular preference for individual codon selection in highly expressed genes may tentatively suggest a greater focus upon reducing missense errors than simply increasing the speed of translation.

1.2.4 tRNA Misacylation

A major source of error is the charging of tRNAs. While this project will alter substitution levels through aatRNA misincorporation (See *1.2.2 Missense Errors*), these misacylation events will also be present in all samples.

Aminoacyl-tRNA synthetases (aatRNA synthetases) make extensive contact with their substrate amino acid in order to maximise discrimination [2], however in many cases the intrinsic binding energies of amino acids to aatRNA synthetases are inadequate to give the required accuracy of translation due to the similarities of many amino acids [2][74]. This leads to the loading of incorrect amino acids onto tRNA molecules, resulting in translational error (previously shown in Figure 1.2), a phenomenon observed *in vivo* [7].

Reinforcing this model of random misaminoacylatioin is the observation that increasing the concentration of an amino acid in the growth media results in its increased misincorporation upon tRNAs through competition effects [75]. In response to this, proofreading systems are required [2][75]. These have a wide range of forms throughout life, but seemingly conserved functions [75].

The synthesis of aatRNA is a two step process, whereby the synthetase first binds the target amino acid and “activates” it with ATP, forming an aminoacyl-adenylate (aa-AMP) intermediate, then transfers this to the 3′ end of a suitable tRNA, releasing AMP in the process [2]. There are, however, alternate approaches to be found throughout life; as exemplified by the multiple pathways for Asn and Gln loading, which may be carried out either individually or simultaneously within an organism. These amino acids are loaded onto their tRNAs either through the direct actions of asparaginyl- and glutaminyl-tRNA synthetases, or through pre-translational modification, whereby Asp-$tRNA^{Asn}$ and Glu-$tRNA^{Gln}$ are modified to Asn-$tRNA^{Asn}$ and Gln-$tRNA^{Gln}$, demonstrating variation in this core process [76].

Editing, the controlled hydrolysis of the amino acid-tRNA bond, takes place upon every step of this process, within the same active site that catalyses aminoacylation, a separate editing

domain in the tRNA synthetase [2], or editing domains in separately expressed proteins [75][77]. The general pathways are shown in Figure 1.8, working to give an error rate of around 10^{-4} [2].

These multiple pathways can coexist within the same system, with pre and post transfer editing working alongside one another to give functional redundancy. In such cases one pathway usually dominates, however if the dominant pathway is repressed, the secondary pathway may compensate and prevent a rise in error levels [78].

Consistent with the redundancy often seen, inactivation of these pathways has severe effects on cell viability [79], genomic stability [80] and the ability to grow in the presence of varying amino acid concentrations [75][79] due to translational errors. This response is in keeping with the severity of decoding errors originating from other sources, particularly the missense errors focused upon in this thesis (see *1.2.2 Missense Errors*), and the associated evolutionary pressures against them, as amino acid misincorporation has the same effect on protein population regardless of source.

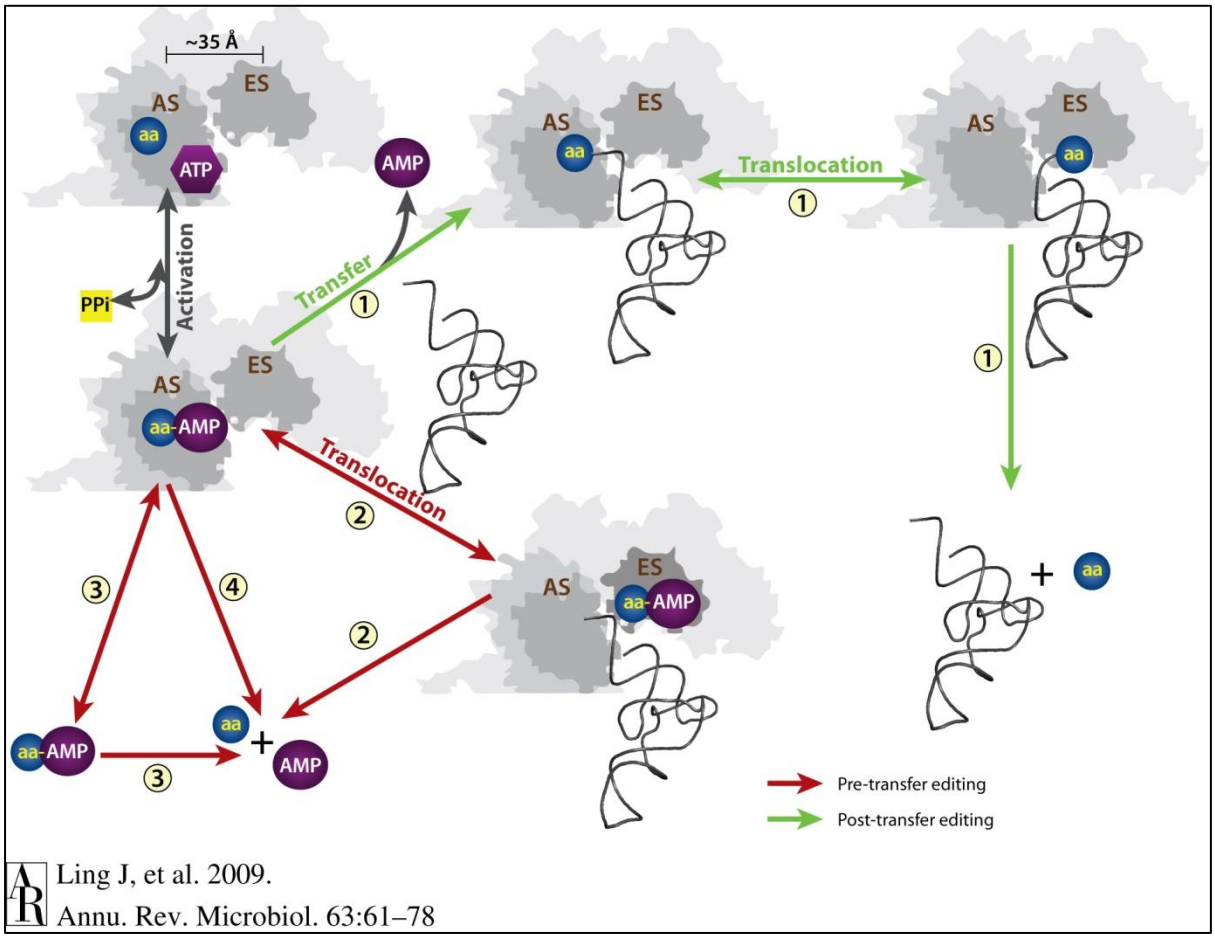


Figure 1.8: ***"Editing pathways against the noncognate amino acid (aa)."*** **from J. Ling et al 2009** [2] ***"(Pathway 1) Posttransfer editing. The misactivated aa is first attached to the tRNA and then translocated from the active site (AS) to the distinct editing site (ES) for hydrolysis. (Pathway 2) Translocation of aminoacyl-adenylate (aa-AMP). The misactivated aa-AMP directly translocates to the ES to be hydrolyzed. (Pathway 3) Selective release. After activation, the noncognate but not the cognate aa-AMP is expelled into solution and subjected to spontaneous hydrolysis. (Pathway 4) Active site hydrolysis. The noncognate aa-AMP is hydrolyzed at the AS before release. Pathways 2–4 collectively comprise pretransfer editing."*** [2]

1.2.5 Tautomerization

Since the elucidation of nucleotide structure and the role of base pairing in base recognition, tautomerization has been proposed as a potential error source [81].

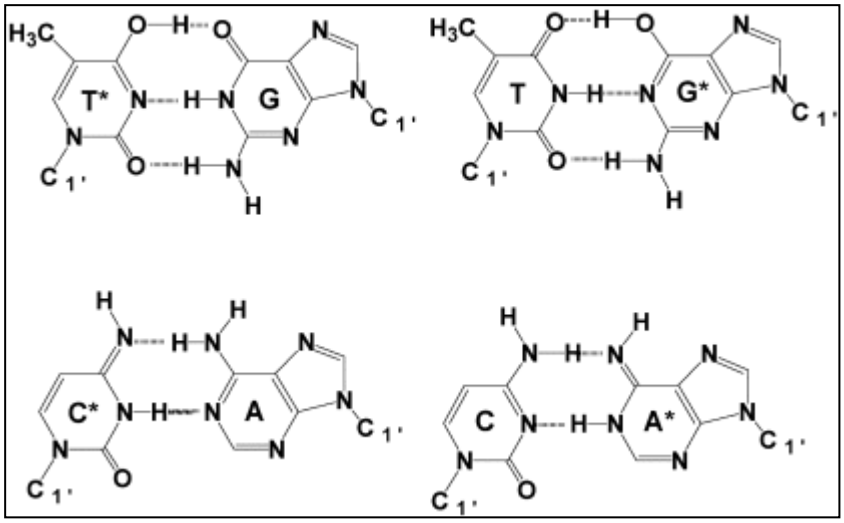


Figure 1.9: ***"Base-pairing schemes involving minor tautomers of the natural bases; top, T:G pairs involving keto-enol tautomers; below, C:A pairs involving amino-imino tautomers"*** **taken from V.H. Harris et all 2003** [161].

Tautomerization, the process by which two isomeric forms of a molecule exist in equilibrium, is a feature of the nucleotide bases, resulting in major and minor tautomeric forms. These minor forms are capable of creating non-Watson-Crick base pairs, as seen in Figure 1.9.

DNA replication has been confirmed to involve these G:T [82] and C:A [83] base pairs, demonstrating true biological effect [84]. As translation depends upon this process, tautomerization appears to be a likely error source.

Crystal structures have shown non-Watson-Crick pairing within the ribosome consistent with tautomerization, exemplified by a G-U pair appearing to follow non-wobble interactions; however *in vivo* such arrangements are highly unfavourable. Molecular dynamics calculations carried out by *Priyadarshi Satpati* and *Johan Åqvist* show that, for this example, tRNA-ribosome complexes at the proofreading stage disfavour the enol tautomeric form by a factor of 10,000 compared to the cognate case, and the crystal trapping of tautomeric base pairs is only achieved through the extreme conditions found in crystallographic study [85].

Crystallographic tautomeric pairing, however, sheds light on ribosomal proofreading against tautomeric and general non-cognate pairs *in vivo*. In the crystal structure of tautomeric pairs, a Mg^{2+} ion is seen in the region of rRNA monitoring bases A1492 and A1493, interfering with their normal function, as described in *1.3 Monitoring bases A1492, A1493, and G530* [85]. This further supports the idea that a functioning ribosome, with a discriminating 16s rRNA monitoring triad, would not be vulnerable to tautomeric errors under normal conditions.

1.2.6 Premature Termination

Premature termination is the end of translation and the release of peptide before the stop codon is reached. Incidents of Termination Factors inserting prematurely are exceedingly rare, and premature termination is most usually the result of "ribosome editing" [19]. If a non-cognate amino acid successfully enters into the ribosome and transfers to the P site, the interactions between the codon-anticodon pair are weaker than the comparative interactions for a cognate pair. This increases the likelihood of premature dissociation of the ribosome and release of the nascent peptide [19][28][29]. This acts as a reserve system to limit the effect of missense translational errors, whereby the incorrectly inserted aatRNAs, having passed all previous quality control mechanisms, may still be rejected. This is, however, a final resort as the termination of a nascent peptide represents a high energetic cost for the prevention of a single missense error.

As this process depends on non-cognate insertion, rates of ribosome editing are increased by conditions which encourage non-cognate insertion, such as amino acid starvation altering the levels of available tRNAs [28][29] (see *1.2.2.1 Missense Errors - tRNA balance importance*).

1.2.7 Frameshifting

A translocational frameshift is most simply seen as a “slippage” of the ribosome during translation. If the ribosome slips by a number of bases not divisible by three it enters a new reading frame (either +1 or -1 to the original reading frame), causing all following codons to be altered [20]. This has a large impact upon the developing polypeptide, and is likely to result in premature termination, as 3 out of 64 random out of frame codons will be termination codons [20]. These large errors, however, are particularly rare, occurring with tenfold lower frequency than other error types [86]. Frameshift events come in two forms; spontaneous and programmed; both of which are believed to follow the similarly cotranslational mechanisms for +1 and -1 slippages [19][86].

These errors were originally treated as special case of missense error, whereby a tRNA mistakenly associates with 4 mRNA bases, however this particular model, known as quadruplet translocation, has long since been abandoned in favour of unique +1 and -1 slippage mechanisms [19][86].

A -1 frameshift usually occurs upon a consensus sequence found in all domains [20][86], often seen in prokaryotes [87] and eukaryotic viruses [88][89], as shown in Figure 1.10. The initial step, found in prokaryotic systems only, is ribosome binding to the internal Shine Dalgarno sequence. These sequences, when found internally, are known to cause pauses in translation and allow accurate ribosomal positioning through complementarity with 16s rRNA, [90][91] and as such works to position the ribosome ready for programmed frameshift [20].

The frameshift itself depends on the slippery region and obstruction. The slippery region, usually with a sequence *X XXY YYZ* (where spaces indicate codon positions) [92] is designed

so that a -1 frameshift will allow repairing of non-wobble bases, and thus minimise the energetic cost of frameshift [92].

Competing models for the precise slippage mechanism have been unified into a “kinetic model” [86] viewing all pathways as possible slippage mechanisms, and indicating two distinct approaches to -1 slippage. The first (“Integrated” [92] and “9Å” [93] models) suggests that the mRNA obstruction prevents proper relative movement of the ribosome and mRNA upon accommodation of an incoming aatRNA upon the slippery sequence, due to GTP hydrolysis by eEF1a, thus creating tension in the spacer region (Figure 1.10). A -1 frameshift relieves this tension, relaxing the system and allowing time for the ribosome to unwind the obstruction [86][92][93].

The other possibility (“cotranslational” model) proposes that normal translocation, powered by eEF2 GTP hydrolysis, is obstructed by mRNA secondary structure, resulting in a +2 translocation in place of a +3 translocation, effectively giving a -1 frameshift [86][94][95].

Similar to -1 frameshift, +1 frameshift is found within both prokaryotic [96] and eukaryotic [30][97] systems, including mammalian cells [86][98]. The mechanisms and signal sequences, however, are far more divergent [86]. The core process, relevant to this thesis, relies on another "slippery sequence". The general principle involves slippage between slow and fast codons. For example, a yeast slippage site of CUU AGG C contains the slow codon AGG corresponding to the low abundance Arg-tRNA^{CCU}. This causes a pause in ribosomal translation. A +1 slippage, however, would give the GGC codon, corresponding to the highly abundant CCG-tRNA^{Gly}. In this way, slow decoding of the in frame codon can be occasionally overtaken by a fast decoding +1 frame codon [86][92][99].

This relation to tRNA competition is verified through increasing the abundance rare aatRNA (here Arg-tRNA^{CCU}), thereby lessening competition effects and reducing +1 shift efficiency by 50%. Further to this, complete depletion of the rare aatRNA increases frameshift efficiency to near 100% as the ribosome reaches the blockage codon [99] and, importantly, takes a +1 frameshift to an abundant tRNA codon instead of simply carrying out a missense error (see *1.2.2 Missense Errors*) or activating the NGD pathway [100][101].

Given the low rate of spontaneous occurrence, and the careful arrangement required for programmed initiation, frameshifting is best viewed as a purposeful mechanism for generating additional polypeptides from desired sequences, with little relation to other decoding errors. Additionally, as rare spontaneous examples are likely to see early termination, [20] the random error here is, in comparison to others discussed, a minor player in protein synthesis.

However, +1 frameshifting is shown to be highly dependent on tRNA competition for its action. It's dependence on relative tRNA levels, and rare codon initiation, indicates a preference for conditions similar to missense errors; although a study focused on missense errors would be unlikely to detect the elevated levels of +1 frameshift errors, due to the difficulty in detecting the "truncated nonsense" peptides produced by this process, and a

decrease in native protein level will likely be overshadowed by the decrease caused by decoding with slower, error prone, codons through delays and missense alone.

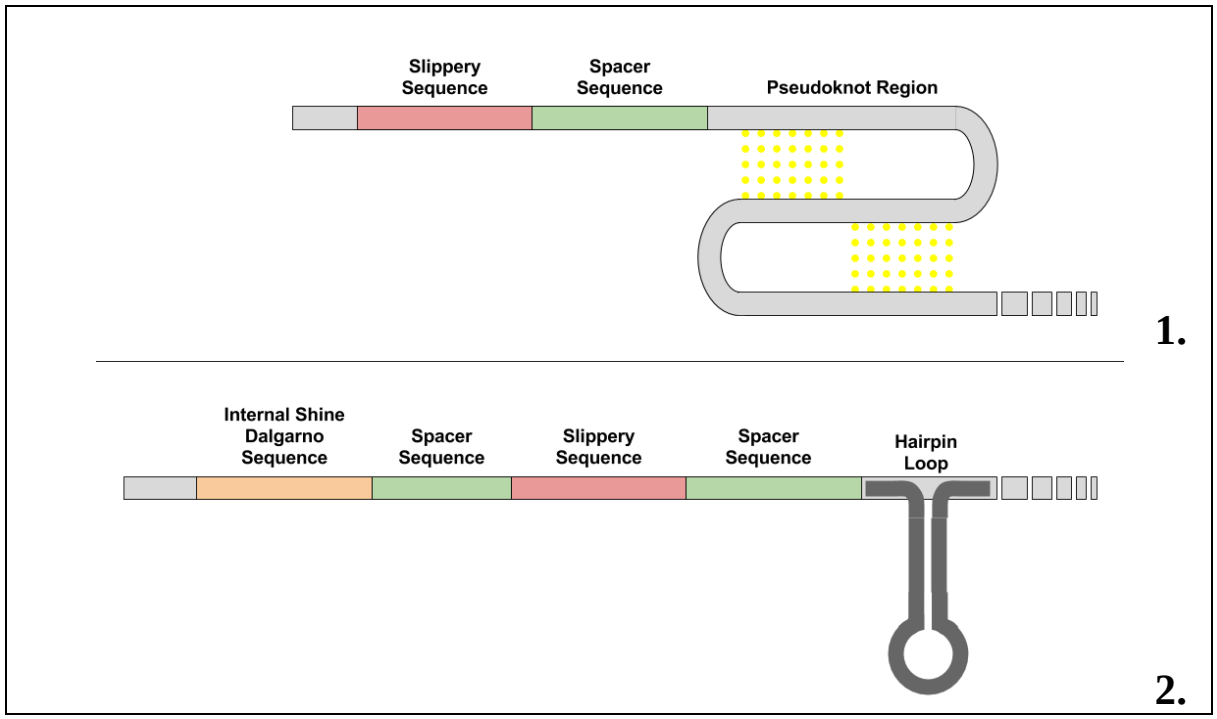


Figure 1.10: Programmed +1 frameshift regions. 1. Eukaryotic systems show a slippery sequence, followed by a pseudoknot region designed to impede ribosomal procession [86]**. 2. Prokaryotic systems show a similar orginisation, using a hairpin loop to slow ribosomal progression. An internal Shine Dalgarno sequence is also used to correctly position the ribosome for a frameshift event** [20]**.**

1.3. Monitoring Bases A1492, A1493, and G530.

Free energy simulations by Priyadarshi Satpati et al (2014) [27] show the key role of this 16s rRNA triad in translational fidelity. These bases undergo significant conformational change in the transition between ribosomal off and on states, as seen in Figure 1.11[27]. While not interacting directly with the codons, these bases act to exclude free water molecules from the area of interaction during transition to ribosomal On state. This prevents

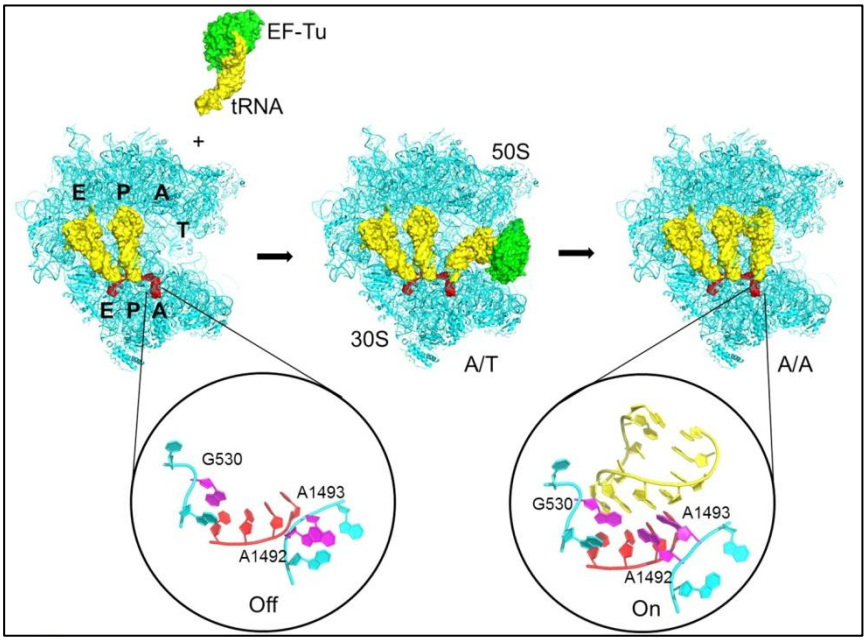


Figure 1.11: Conformational change of the monitoring bases during transition from ribosomal off to on state, showing their proximity to the codon-anticodon pair in the On state. These bases were originally assumed to interact with the codon-anticodon pair itself, but have been shown to act through solvent exclusion. ***Image taken from P Satpati et al 2014*** [27]

mismatched base pairs from forming hydrogen bonds with water, giving a high energetic cost for non-cognate base desolvation, disfavouring incorrect pairings. This affects the total energetic cost of transition from Off to On states, with extra cost of desolvation making the rejection pathway, returning to Off state, far more likely for a non-cognate base [27].

Further proof of this system can be seen in its deactivation. It is well documented that increasing Mg^{2+} concentrations correlate with an increase in error rates [102]. In sufficiently high Mg^{2+} concentrations, a Mg^{2+} atom can be seen to bind in the monitoring triad region, where it appears to aid transition of the region from Off to On states at a lower energetic cost, causing both higher error rate and overall translation speed [27][85]. Under these conditions the ribosome is able to form tautomeric errors not normally found within the ribosome [85] (see *1.2.5 Tautomerization*) as well as tolerating non-cognate insertions (see *1.2.2 Missense Errors*).

1.4. Ribosome Heterogeneity

An additional complication, as has been suggested by Ribas de Pouplana et al (2014) [103], is that ribosome heterogeneity may modulate translational fidelity. Ribosome heterogeneity means that ribosomal composition is not static, and may change in response to environmental factors in prokaryotic [104] and eukaryotic cells [105]. Ribosomal modifications [105] and mutations [106] have been observed to affect translational accuracy, lending support to this proposal. If true, the role of dynamic ribosomal heterogeneity as an environmental response may well alter the cell's approach to translational accuracy.

1.5. Translational Error in Context of General Misfolding

Highly expressed proteins demonstrate a slower sequence evolution than their lower expressed equivalents (E-R anticorrelation). This has been attributed to the cost of misfolded proteins on expression through the Translational Robustness Hypothesis [107]. This proposes that in the evolution of a highly expressed protein, even non functionally dependent residues will be under selective pressure to maintain the native structure, thus reducing the burden of misfolding on the cell. While the original hypothesis focuses on translational error induced misfolding [108] it has since been extended to include selection against error-free misfolding, another important factor in sequence evolution, through the Misfolding Avoidance Hypothesis [107].

Translational errors are predicted to represent a significant burden, with around 19% of average yeast proteins containing a missense error [107] and 20-65% of substitutions leading to inactivation, most commonly through misfolding [107][109][110]. To give perspective, in producing the required 1.26 million copies of PMA1, this error rate would give 63,000 missense misfolding errors, outnumbering >97% of yeast proteins. This demonstrates the need to reduce errors and decrease misfolding, even before non-error induced misfolding is considered [108].

In a similar manner to error based misfolding, error free misfolding may create a pool of non-native protein forms, and is selected against through the selection of high stability amino acids in each position, a pressure most clearly found in highly expressed proteins [108].

In the case of a highly expressed protein with optimised codons and amino acids, it is suggested that genetic mutations altering codon usage are likely to result in increased misfolding rates due to a rise in error based and error free misfolding; as such mutations are selected against [107][108]. This shows translational error to be a strong evolutionary restraint in the cell, favouring stability of highly expressed species.

In the Misfolding Avoidance Hypothesis the effect of error-free misfolding is of a comparable importance to mistranslation induced misfolding; i.e. post-translational and translational errors are similar in their impact. These competing factors result in a balance between optimal codon and optimal amino acid usage, causing both factors to be non-fully optimised [108]. Thus while translational error suppression is important, a cell must balance this against alternate sources of error, preventing perfect optimisation from occurring, and indeed implying that a sequence composed entirely of optimal codons is not the optimal sequence for efficient protein expression.

It may also be noted that the selection of suitable amino acids, increasing overall stability, may increase the ability of a protein to fold in the presence of translational errors, further complicating the relationship between these two factors. This effect has been proposed as a factor contributing to the higher stability of predicted ancestral proteins in comparison to their modern equivalents; allowing ancestral cells to function effectively in the presence of heightened translational error rates [111].

1.6. Overall Sequence Optimisation & Focus of This Thesis

Translational error is an inevitable and significant feature of protein creation. The design of an optimal codon sequence is a highly complex task involving the balancing of multiple error sources; such as matching codon choice to local aatRNA concentrations to minimise missense errors, selecting highly stable amino acids for each position to minimise spontaneous misfolding, managing codon context biases & resulting out of frame codons, balancing of base biases within neighbouring codons, avoiding frameshift sequences & unwanted mRNA secondary structures, and possibly accounting for further dynamic factors such as tRNA processing and depletion & ribosomal heterogeneity.

This being said, however, it is possible to create non-perfect "good" and "bad" sequences for a given protein based upon aatRNA/codon balance alone, as this system has a strong impact upon mistranslation, [6][24][31][32] represents a heavy source of misfolding [107][108], and has a significant effect upon cell function [31][32][42]. Clearly, there would be a large difference between minimum and maximum optimisation codon sequences based upon tRNA balance, providing a simple model system in which to experimentally probe the effects of translational error.

It may be noted that the choice of optimal codons as those with the highest tRNA abundance was challenged by P Shah et al 2010 [112]. It is well known that the ratio between cognate and non-cognate is the major factor determining translational error rates [18][26], however general selection for high abundance tRNAs is often seen as a suitable simplification. P Shah et al 2010 argues that correlation between tRNA abundances within genomes means that high abundance tRNAs often have high abundance near-cognate tRNAs, and as such would only provide increased translational speed, not accuracy [112]. This analysis, however, defined a near-cognate codon as a codon differing by a single base pair [112]. This represents a heavy simplification. Experimental work has shown that the binding of near-cognate bases relies on the formation of a transient mini-helix between the codon and anticodon; something which does not occur for all codons differing by a single base, and may occur with greater than one base difference [26]. Additionally, study of all codons one base removed from UUU, found only four out of fourteen codons to bind significantly above a non-cognate control [18]. As such, the analysis of P Shah et al 2010 will have taken into account an expanded and largely incorrect pool of near-cognate codons when determining competition effects and abundance correlation, increasing the likelihood of finding abundant codons under strong "near-cognate" competition. Finally, it is claimed that instead of error prevention, the tendency of cells to correlate codon usage with tRNA abundance is to increase translation speed, and reduce nonsense errors [112]. However, the observation that Intrinsically Disordered Regions, where folding does not take place, show reduced levels of codon/tRNA correlation (see *1.2.2.1 Missense Errors - tRNA balance importance*), can only be explained through recognition that missense errors, which may have reduced impact in these areas, are a key factor in driving cells to utilise codons with high tRNA abundances. As such, the codon optimisation technique used in this project, selecting for high abundance tRNAs, remains an appropriate and biologically relevant approach.

Previous work by Tobias von der Haar and Lyne Jossé created minimum and maximum codon optimisation sequences for Glutathione S-transferase (heron referred to as Max-GST and Min-GST respectively) in order to examine translational errors and their regulation through comparative study within transformed *S. cerevisiae* and *E. coli*, with the aim of identifying systems involved in translational error regulation.

The binding affinities of Max-GST and Min-GST, expressed in both *S. cerevisiae* and *E. coli*, were examined through the use of a glutathione affinity column; whereby GST extracts were passed through the column, and the percentage of protein bound recorded. It was expected that Min-GST in both species, due to the higher rate of random substitution errors, would display a lower level of binding than Max-GST. This behaviour was observed in *E. coli* samples, shown in Figure 1.12.1, demonstrating clear functional impact of codon optimisation upon protein activity. *S. cerevisiae* samples, however, showed no binding difference between Max and Min-GST, implying no difference in substitution errors in the mature proteins.

In addition to this, a mass spectrometry technique was created in order to detect translational error substitutions within mature proteins. Using this technique, translational error levels could be directly observed in *E. coli* and *S. cerevisiae* produced GST samples, as shown in

Figure 1.12.2. It can be seen that true error levels reflect the conclusions of the earlier binding study; *E. coli* samples show Min-GST to present higher error levels, while *S. cerevisiae* samples show equal error levels regardless of codon optimisation. *S. cerevisiae* cell cultures were also exposed to NAT, known to increase translational error rates [113], which resulted in observable rises in translational error, with Min-GST rising the greatest.

As such, this thesis attempts to identify possible error regulation pathways in *S. cerevisiae* which may explain the resistance of these cells to the production of additional translational errors in response to lower codon optimisation.

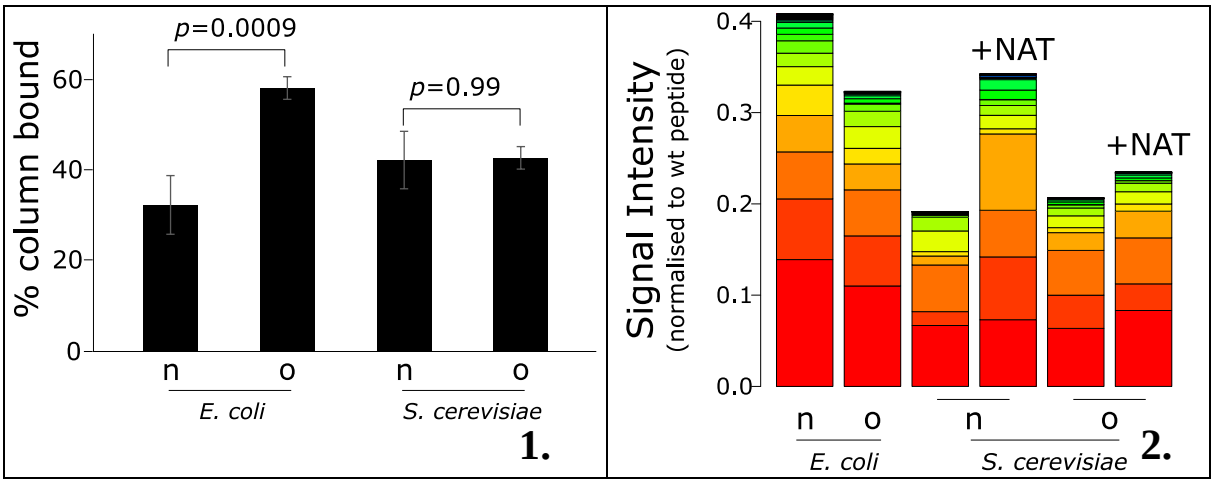


Figure 1.12: Effects of codon optimisation in the production of GST. Graphs provided by Tobias von der Haar. 1. Min/Non-Optimised GST and Max/Optimised GST produced in *E. coli* and *S. cerevisiae* were passed over a glutathione affinity column. The percentage bound to the column was recorded. In keeping with expectations, the higher levels of error substitutions in the non-optimised *E. coli* sample reduced its binding compared to the optimised sample; however *S. cerevisiae* samples showed identical binding, regardless of codon optimisation. 2. Mass Spectrometry was used to directly detect translational error substitutions within mature proteins. Min/Non-Optimised GST and Max/Optimised GST produced in *E. coli* and *S. cerevisiae* were examined. *E. coli* samples showed higher levels of substitution errors for non-optimised GST, while *S. cerevisiae* showed approximately equal substitution errors for both Optimised and Non-Optimised GST. NAT, known to increase levels of translational error, was added to *S. cerevisiae* cultures, resulting in detectable rises in error levels.

Additionally, MS analysis was carried out on *E. coli* Min-GST before and after binding to a glutathione column, allowing the relative binding affinities of various translational error substitutions to be observed. In this thesis, computational analysis is carried out upon this data set in order to gain a fuller understanding of the errors, and several of these errors are created at the DNA level as GST mutants in order to study their behaviour directly. These complimentary approaches will allow the behaviour of translational error proteins to be examined, and will primarily address two question: 1. What factors differentiate an observable translational error from a non-observable translational error? 2. Does the process of inserting a translational error substitution into the growing peptide have any effects upon the mature product beyond amino acid alteration (i.e. will the DNA mutants show similar behaviour to the original translational error proteins)?

2. Materials and Methods

2.1. Strains

2.1.1. S. cerevisiae

Protein expression - BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) [114].
Knockout mutants - *Δrqc1*, *Δrqc2*, *Δpep4*, *Δupf1*, *Δdom34*, & *Δdhh1* [115].

2.1.2. E. coli Strains

Protein expression - BL21 (*fhuA2* [*lon*] *ompT gal* [*dcm*] *ΔhsdS*).
Plasmid cloning - TOP10 (*F– mcrA Δ(mrr-hsdRMS-mcrBC*) *Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ*(*ara leu*) *7697 galU galK rpsL* (*StrR*) *endA1 nupG*).

2.2 Transformation

2.2.1. *S. cerevisiae*

The expression vector containing full length N-terminal $(His)_6$ tagged GST was pre-prepared by Dr Tobias von der Haar. 1ml BY4741 overnight culture was pelleted through centrifugation, followed by addition of 240µl 50% PEG 4000, 36µl 1M LiAc, 10µl ssDNA (preheated to 96°C for ten minutes to achieve denaturation), 2.5µl 2-Mercaptoethanol, 1µl plasmid DNA, and 69.5µl water followed by vortexing. Cells were then incubated at room temperature for 20 minutes, heat shocked at 42°C for 15-20 minutes, centrifuged at 2000rpm for 2 minutes, supernatant removed and resuspended in 200µl water. Cells were plated on selective Ura deficient agar.

2.2.2. *E. coli*

The expression vector containing full length N-terminal $(His)_6$ tagged eIF4AI was pre-prepared by Dr Tobias von der Haar. 100µl of competent cells were incubated on ice with 1µl plasmid for 20-30 minutes, heat shocked at 42°C for 60 seconds, and then returned to ice. 1ml of LB broth was added, and the cells incubated at 37°C for 1 hour with shaking. Cells were then plated on 50µg/ml kanamycin LB or 20 µg/ml chloramphenicol LB agar depending upon selection marker used.

2.3 Generation of Competent *E. coli* Cells

115ml LB was inoculated with 32 µl of overnight culture and incubated at 37°C. At OD_{600} = 0.5, 15ml of warm (37°C) glycerol was added. Incubation continued for 5 minutes, before transferring to a pre-chilled centrifuge bottle and keeping on ice for a further 10 minutes. Culture was centrifuged at 4000rpm, 4°C, for 10 minutes and supernatant discarded. Cells were resuspended in 100ml ice cold Mg-Solution (0.1M $MgCl_2$, 15% Glycerol), centrifuged at 3800rmp, 4°C, for 8 minutes, and supernatant discarded. They were then resuspended in 100ml ice cold T-Salts (0.075M $CaCl_2$, 0.006M $MgCl_2$, 15% Glycerol), incubated on ice for 20 minutes, centrifuged at 3800rmp, 4°C, for 8 minutes, and supernatant discarded. Resulting cells were gently resuspended in 5ml ice cold T-Salts and stored at -80°C. Cells were allowed to freeze overnight before use.

2.4. S. cerevisiae drug testing

Cells were grown in 5ml Ura- broths overnight, before transferring to a 30ml solution at 0.1OD_{600}. NAT cell cultures were grown in the presence of 3µg/ml Nourseothricin Sulfate. PMSF cell cultures were grown until around OD_{600} 0.6-1.0 before addition of 1mM Phenylmethylsulfonyl fluoride. MG132 cell cultures were grown until around OD_{600} 0.6-1.0 before addition of 100µM MG132 & 100µM SDS (to increase permeability and allow MG132 uptake [116]).

2.5 Protein Extraction

2.5.1. S. cerevisiae

Protein extraction was carried out as described by von der Haar 2007 upon 5 OD_{600} units of cell culture [117].

2.5.2. E. coli

5 OD_{600} units of cell culture were harvested, frozen cell pellets resuspended in 500µl Resuspension Buffer (200mM NaCl, 20mM Tris-HCl pH 7.5), and lysed through sonication.

2.6. Protein Precipitation

Methanol/Chloroform precipitation was carried out as described by von der Haar 2007 [117].

2.7. Immunoprecipitation

200µl protein samples were incubated with Anti-Ubiquitin antibody (Abcam Anti-Ubiquitin antibody [ab19247][118]) at 4°C overnight with rocking. 40µl washed Dynabeads® were added to samples, and incubated for 30 minutes at room temperature with rocking. Beads were removed through pelleting, and washed through three further suspensions and pelletings. Beads were resuspended in 40µl 3xSDS Sample Buffer, vortexed briefly, heated to 95-100°C for 10 minutes, and finally centrifuged for 1 minute at 13000rmp. The supernatant was taken for SDS-PAGE analysis.

2.8. Western Blots

2.8.1. SDS-PAGE

Gels were made according to Table 2.1, (12.5% separating and 4% stacking concentrations) samples boiled at 96°C in a volume of 4xSDS Sample Buffer (80mM Tris pH6.8, 20% Glycerol, 2% SDS, 0.05% Bromophenol Blue) and run at 180V in Running Buffer (25mM Tris-base, 192mM Glycine, 5.2mM SDS).

	Separating Gel	Stacking Gel
30% Acrylamide:Bisacrylamide 29:1	9ml	2ml
4x Lower Tris (1.5M Tris pH 8.8, 0.4 % SDS)	5.4ml	-
4x Upper Tris (1.0M Tris pH 6.8, 0.4% SDS)	-	3.5ml

Deionised Water	7.2ml	8.4ml
40% Ammonium Persulfate (APS)	80µl	80µl
Tetramethylethylenediamine (TEMED)	10µl	10µl

Table 2.1: Composition of polyacrylamide gels used. Quantities shown for a four gel batch.

2.8.2. Transfer

Nitrocellulose paper was soaked for 2 minutes in methanol, before soaking with gel in Transfer Buffer (20mM Tris, 200mM Glycine, 3.5mM SDS) for 10 minutes. Semi dry transfer was carried out at 9V for 30 Minutes.

2.8.3. Enhanced Chemical Luminescence (ECL) Detection

Blocking was carried out using TBS-M (50mM Tris pH 7.8, 150mM NaCl, 5% dried milk powder) for ten minutes, followed by overnight incubation in TBS-M-Primary Antibody (anti-GST produced in rabbit [G7781 SIGMA]) at 4°C. Membrane was rinsed once in TBS-M, and then incubated with a TBS-M-Secondary antibody mixture (anti-rabbit antibody, peroxidase linked [12-348 SIGMA]) for one hour at room temperature. Blot was visualised using ECL solutions; 10 ml Solution 1 (2.5mM Luminol, 400µM Coumaric Acid, 100mM Tris pH 8.5) and 10ml Solution 2 (0.064% H_2O_2, 100mM Tris pH 8.5) incubated with membrane for two minutes, before visualization with a SynGene G:BOX for cumulative exposures of 1-10 minutes.

If repeat detections were required upon the same gel, stripping was carried out using ThermoFisher Scientific Restore^{TM} Stripping buffer [119].

2.9. Gel Staining

2.9.1. Coomassie Blue

Gels were soaked in Staining Solution (40% Methanol, 8% Acetic Acid, 0.6µM Coomassie Brilliant Blue) for 30 minutes, destained in destain solution (40% Methanol, 8% Acetic Acid), and stored in distilled water.

2.9.2. SYPRO® Ruby Protein Gel Stain

Stain was used according to manufacturer instructions [120].

2.10. Cell Growth Assays (*S. cerevisiae*)

Growth rates were measured using a BMG labtech SPECTROstar^{nano} for BY4741 transformed with a control plasmid. Samples consisted of 1ml Ura- broth inoculated with 10µl overnight culture, and all drugs were added at beginning of incubation to levels used in previous experiments. OD_{600} measurements were made at 30 minute intervals with 400rpm double orbital shaking in a 24 well plate. Growth rates for each sample were calculated for the fastest increasing seven point run. All readings were adjusted according to three sterile Ura- controls.

2.11. His Tag Purification

500ml to 1000ml culture was used to generate samples. Yeast cells were lysed through 30 minute incubation at room temperature in the presence of 5ml Y-PER (Thermo Fisher Scientific, UK) per gram wet mass, with 8 M Urea added for denaturing conditions. *E. coli* cells were lysed through sonication of frozen pellet in Buffer A-Urea (50mM Tris pH 7.5, 250mM NaCl, 8M Urea) with the addition of one protease inhibitor tablet (Sigma Aldrich cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail[121]). Urea is used to denature before Nickel binding, as the tertiary structure prevents access to the N-Terminal $(His)_6$ Tag for this GST construct. Proteins were extracted through centrifugation for 15 minutes at 15,000 rpm. 1ml of Chelating Sepharose Fast Flow [122] was used per 2-3g of highly expressed protein. For poorly expressed Min constructs, lower volumes of sepharose are recommended to reduce contamination. Beads were washed in column with 10ml water, 2ml nickel solution (0.2M $NiCl_2$). 10ml water, and 10ml buffer A-Urea. Sample is then run over the column and washed with 10ml Buffer A-Urea. In dialysis based runs Buffer B-Urea (50 mM Tris pH 7.5, 100 mM NaCl, 100 mM EDTA, 5mM β-mercaptoethanol, 8M Urea) is used to elute in 1ml fractions. In non-dialysis runs the sample is washed with 10ml Wash Buffer (20mM Tris pH 7.5, 200mM NaCl, 5mM $MgCl_2$, and 5mM β-mercaptoethanol) to remove urea, and Buffer B used to elute in 1ml fractions.

2.12. Dialysis

Dialysis was carried out using Thermo Scientific™ Slide-A-Lyzer™ 10K MWCO Dialysis Cassettes [123] in dialysis buffer (20mM Tris pH 7.5, 200mM NaCl, 5mM $MgCl_2$, and 5mM β-mercaptoethanol). Sample was dialysed overnight at 4°C, then for two further hours in fresh dialysis buffer. This method is not recommended due to resultant protein precipitation. Instead urea should be removed upon the nickel-sepharose column, as described above.

2.13. GST Binding Assay

Purification fractions were analysed using coomassie blue stained SDS-PAGE gels, and fractions with high GST were pooled. Total sample volume was reduced to 1ml using 10,000 MWCO Pall Corporation Microsep™ Advance Centrifugal Device [124]. GST binding assay made use of Pierce™ Glutathione Magnetic Agarose Beads. 100µl slurry was washed with 2 x 1ml Dialysis Buffer, before addition of 0.5ml sample, 15 minute room temperature incubation with shaking, removal of supernatant, wash with 2 x 1ml Dialysis Buffer, elution through addition of 200µl SDS-Sample Buffer at 95°C for 10 minutes, and final separation through centrifugation at 13,000 rpm for 3 minutes.

2.14. MS Sample Preparation

Samples were run on SDS-PAGE gels, coomassie blue stained and destained as normal. Sample bands were cut from gel, diced into 1mm cubes, and transferred to microcentrifuge tubes.

Gel pieces were washed in the following series, removing liquid between each wash: 2 x 15 minutes 100mM NH_4HCO_3, 50% Methanol (to reduce staining); 2 x 15 minute 100µl 50mM NH_4HCO_3:acetonitrile (1:1); 15 minutes 100µl acetonitrile to shrink gel pieces; 30 minutes at 56°C ~50µl (gel covering) 10mM DTT in 50mM NH_4HCO_3; ~ 5 -10 minutes brief shrinking in acetonitrile; 15 minutes at room temperature in dark 55mM Iodoacetamide in 50mM NH_4HCO_3; 2 x 15 minutes 100µl 50mM NH_4HCO_3, 30 minutes at 4°C 25mM NH_4HCO_3 10% acetonitrile, 10ng/µl trypsin (checked after 15 minutes, adding more trypsin solution if all had been absorbed by gel pieces); overnight in ~10µl (minimum to cover gel and avoid diluting sample) 25mM NH_4HCO_3 10% acetonitrile.

5µl acetonitrile was added and samples sonicated for 15 minutes in an ultrasound water bath. Samples were spun down and supernatant collected. 10µl 50% acetonitrile, 5% formic acid was added and samples sonicated another 15 minutes before final supernatant was collected and pooled with previous supernatant. Samples were stored at -80°C before analysis.

2.15. Computational

2.15.1. Image Analysis

Gel band intensities were measured using ImageJ [125], and distances were measured using GIMP [126].

2.15.2. Molecular modelling

Protein structures were obtained through the PDB [127] and visualised using PyMol [128]. Conservation values were generated using the ConSurf server [129][130][131][132][133]. Amino acid changes ranked according to the BLOSUM62 Substitution Matrix [134].

2.15.3. Statistical Analysis

Statistical Analysis was carried out using Microsoft Excel (2007). Relative tRNA levels were taken from existing literature [135][136].

3. Results

3.1. *S. cerevisiae* Error Control

Previous MS experiments established the presence of translational errors in GST production, and their differing behaviours in *E. coli* and *S. cerevisiae*. It was observed that while low optimisation codon sequences increased error levels in *E. coli*, the same behaviour was not observed in *S. cerevisiae*, where error levels were stable regardless of codon optimisation (as shown in Figure 1.12). As such we wished to establish whether translational error products were processed by proteolytic pathways such as the surveillance pathways in *S. cerevisiae*, explaining the differences observed in these systems.

3.1.1. Production Levels

In order to confirm clear differences between Min-GST and Max-GST, levels of production were compared in *S. cerevisiae* transformants through multiple western blots, as exampled in Figure 3.1.1. These were used to estimate the relative GST levels detectable in transformed strains, with Max-GST seen as seven fold more abundant than Min-GST in Figure 3.1.2, demonstrating clear biological effects of the sequence optimisation.

Due to these differing abundances, loading was normalised to give main bands of equal strength in Figure 3.1.1, allowing examination for degradation products. The additional low MW band observed was also present in control samples; however higher MW bands were observed as more abundant in non-control samples, representing possible GST-Ubiquitin conjugates.

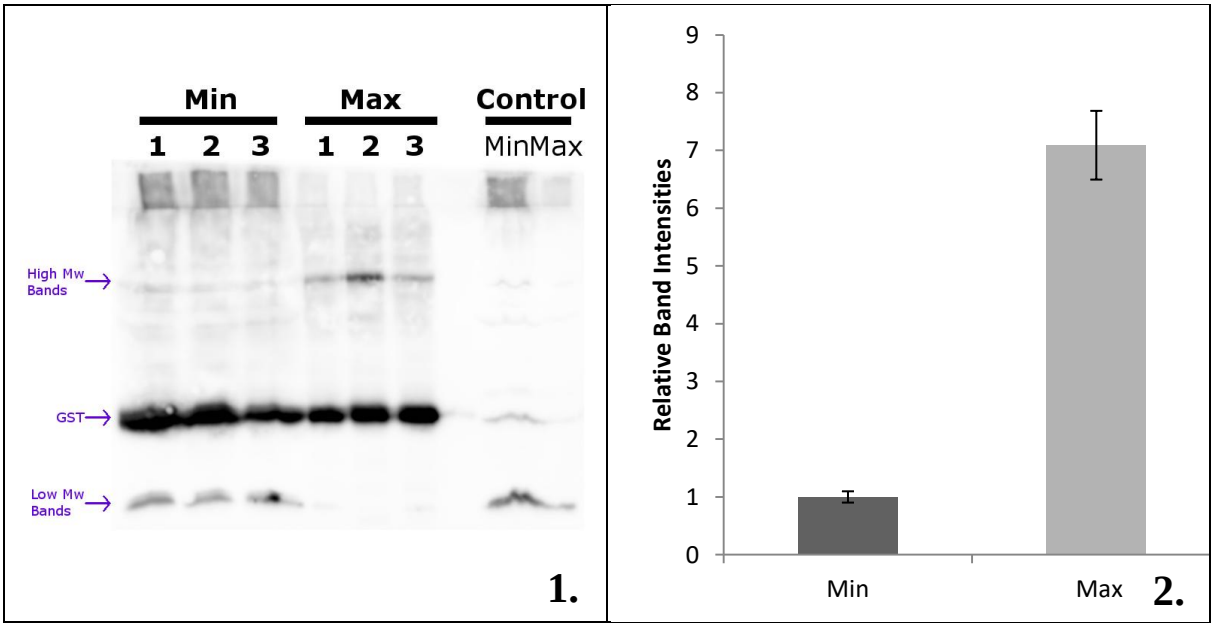


Figure 3.1: Max/Min levels comparison. 1. Min-GST and Max-GST western blot. Due to the greater abundance of Max-GST, loading levels are adjusted to give a roughly even intensity of GST both cases (7.5 Min : 1 Max). Min = Min-GST. Max = Max-GST. Control cells are transformed with blank plasmid and loaded equally to Min or Max. 2. Relative band intensities for Min-GST and Max-GST, adjusted for equal loading levels, with 95% confidence intervals shown. n = 6. p = 9.54×10^{-6}.

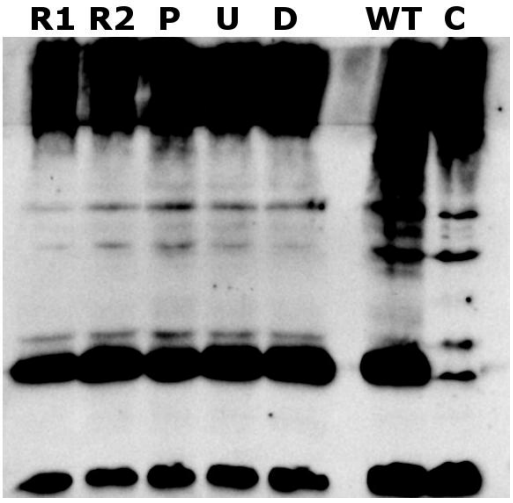
3.1.2. Knockout Screening

While no breakdown products were observed in WT samples, it is possible that breakdown products are simply removed too rapidly for detection. In order to enrich potential breakdown products, and possibly alter total GST levels, deletion mutants are examined for likely pathways. For this purpose, Min-GST was transformed into a range of *S. cerevisiae* knockout mutants.

One group of pathways which could, in principle, affect the levels of translational error products are the surveillance pathways. Surveillance pathway knockouts, described in Figure 3.2.1, were compared through western blot for Min-GST in order to examine the possible role of mRNA surveillance pathways in translational error response. The result, shown in Figure 3.2.2, shows no clear alteration in the levels of GST, or build-up of breakdown products in any knockout strain, suggesting no role in breakdown of this high error product.

Knockout	Function	Pathway
Rqc1 & *Rqc2*	Components of the Ribosome Quality Control Complex. Responsible for binding stalled ribosomes and triggering degradation of nascent polypeptides [137].	No Go Decay & Non Stop Decay [101]
Dom34	Involved in the dissociation of stalled ribosomes and mRNA cleavage, *Pep4* both in coding regions and the 3' UTR [100][138].	No Go Decay [100] & Non Stop Decay [138]
Upf1	Involved in termination at premature stop codons [139].	Nonsense Mediated Decay [139]
	Vacuolar protease required for the activation of multiple vacuole hydrolases [140].	*Vacuole decay control* [140]

1.



2.

Figure 3.2: Examination of mRNA surveillance pathway knockouts. 1. Knockout strains tested and their relevant functions. 2. Western blot showing expression of Min-GST (large central band) in surveillance pathway knockouts, with wild type [WT] and empty plasmid [C] controls. No significant variation is seen between the mutant samples, failing to indicate a key role for any of the pathways examined. R1 = *Δrqc1* R2 = *Δrqc2* P = *Δpep4*. U = *Δupf1* D = *Δdom34* WT = Non-mutant control. C = Blank plasmid control.

During the course of these experiments we became aware of the role of *dhh1* as a conserved eukaryotic protein known to target non-optimized mRNA based on ribosome accumulation, acting independently of other major mRNA decay pathways [141]. It was decided to test this

mechanism in addition to those in Figure 3.2. A knockout strain was examined through western blot for Min-GST and Max-GST expression, as seen in Figure 3.3. No significant difference in GST band intensity can be seen in response to dhh1 removal. Higher MW bands do appear to increase in concentration in *Δdhh1* cells, however coomassie blue staining (not shown) indicates this to be due to an increase in general protein levels, not specific to GST.

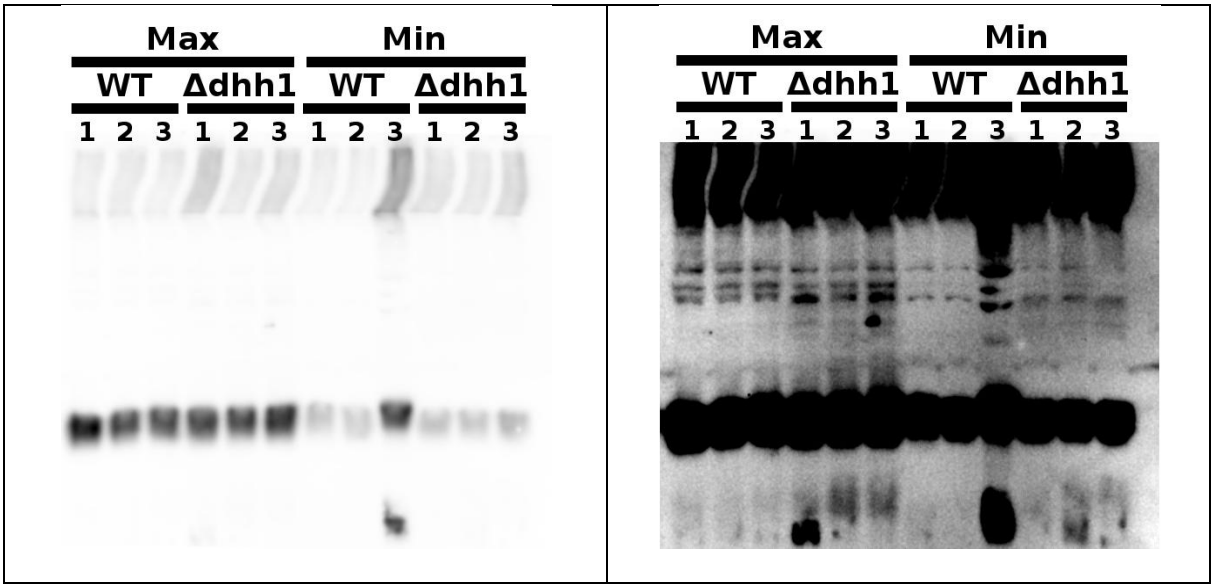


Figure 3.3: Differing levels of exposure for anti-GST western blots, comparing the effects of *dhh1* knockout upon Min-GST and Max-GST expression. No significant difference is observed, however *dhh1* mutation leads to a slight increase in overall protein levels, as confirmed by coomassie blue staining (not shown). Max = GST-Max. Min = GST-Min. WT = Wild Type. *Δdhh1* = *dhh1* knockout.

3.1.3. PMSF & MG132 Assays

As the above pathway knockouts did not demonstrate involvement in GST translational error response, the canonical protein turnover pathways were next examined for involvement. PMSF is a common serine protease inhibitor, known to affect several vacuolar proteases responsible for the degradation of long lived proteins while maintaining proteasome function [142]. MG132, on the other hand, is an inhibitor of the proteasome [142]. Using these drugs it is possible to identify the decay pathway of a sufficiently turned over protein [142].

Additionally, Nourseothricin (NAT), an antibiotic known to induce miscoding errors in translation [113], was used to increase error levels in both Min-GST and Max-GST, potentially amplifying error dependent processes.

In order to confirm suitable drug doses for use in these experiments, maximum growth rates were examined under drug conditions, as shown in Figure 3.4. Each condition shows a slight significant reduction in growth rate, indicating true biological effect upon cell function, while maintaining cell viability. This should, therefore, allow the following experiments to represent drug effects upon functional cells.

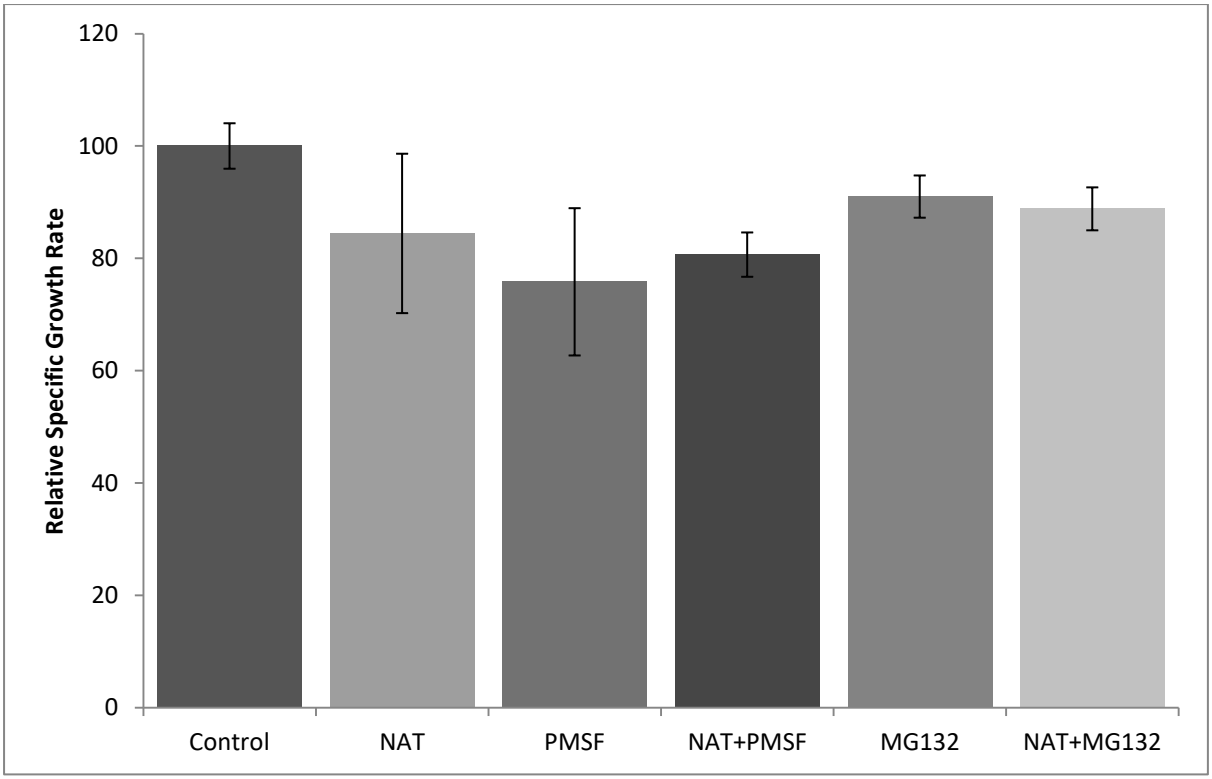
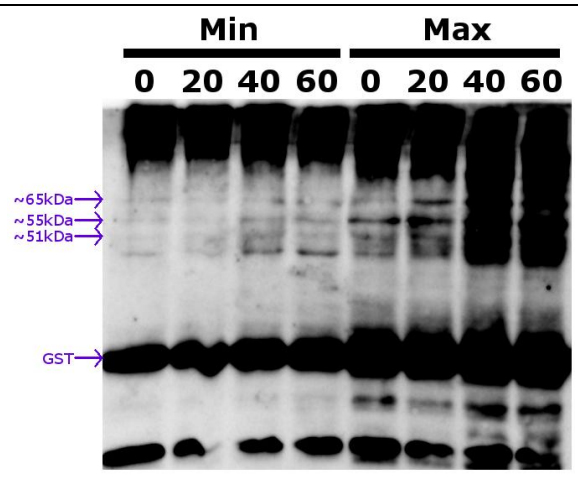
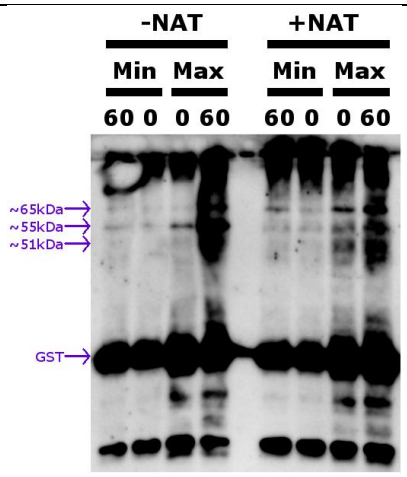


Figure 3.4: Maximum growth rates for yeast BY4741 under drug assay conditions, with 95% confidence intervals shown. All drug combinations produce a small but significant decrease in growth rate, indicating a biological effect on sufficiently viable cells, supporting the use of these concentrations throughout the project.

As such, Min-GST and Max-GTS expressing *S. cerevisiae* cultures were exposed to the drug doses tested above. While no breakdown products were revealed, PMSF and MG132 were seen to have time dependent and differing effects upon banding patterns in anti-GST western blots, particularly upon higher MW bands, as shown in Figure 3.5. The concentration of higher MW bands is cumulatively increased by NAT and PMSF, and decreased by MG132.

	Time Series	With NAT addition.
PMSF		

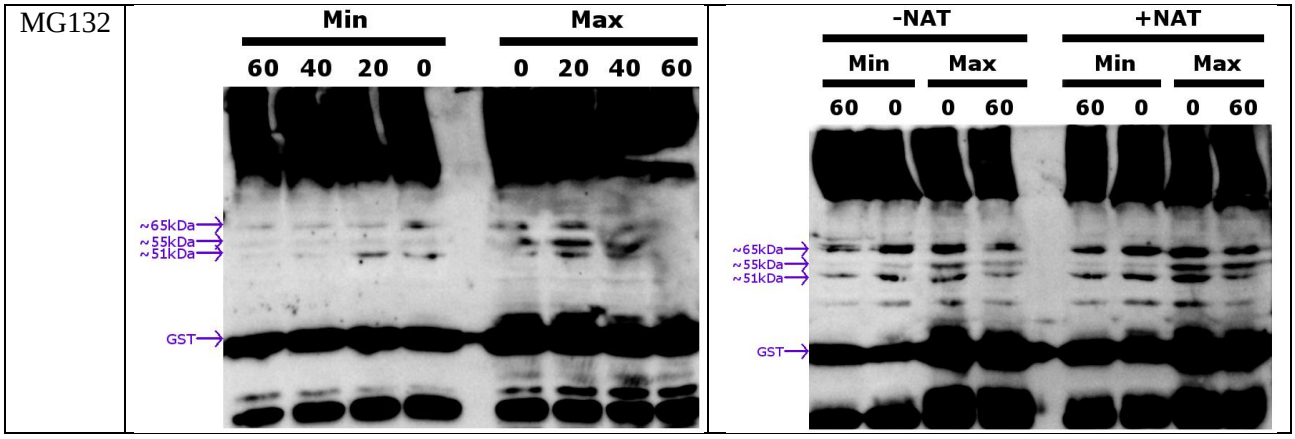
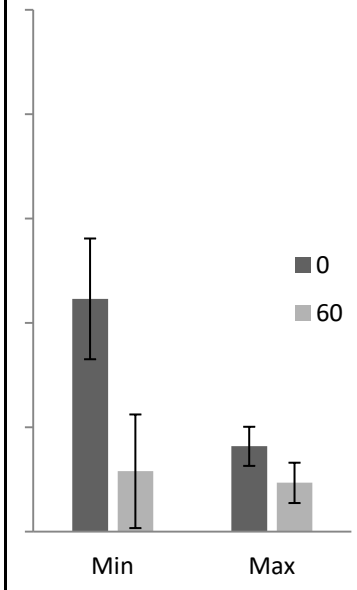
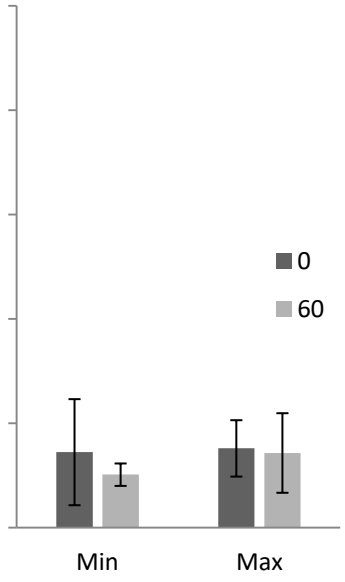
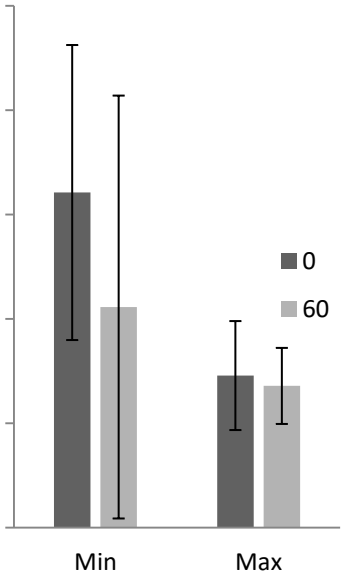
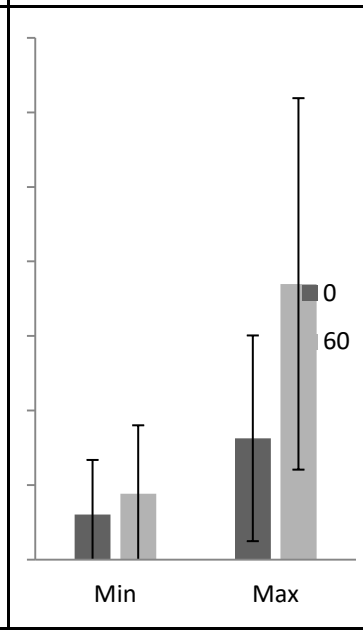
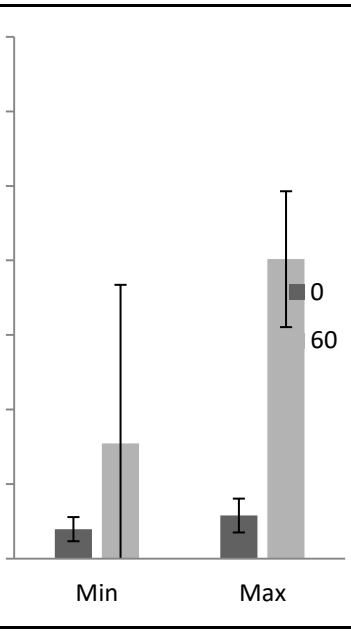
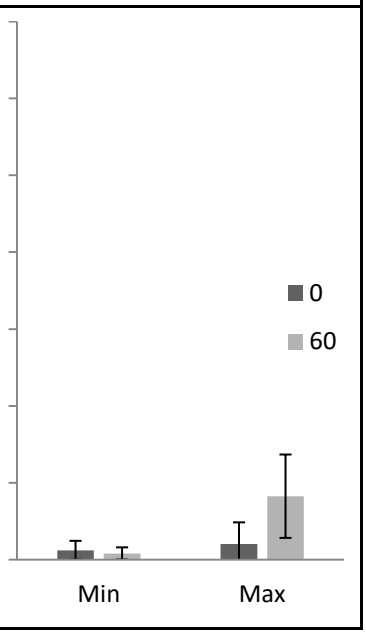


Figure 3.5: Anti-GST western blots showing the effect of MG132 & PMSF treatment, both in the presence and absence of error inducing NAT, upon Min-GST & Max-GST expressing cells. MG132 & PMSF are seen to alter band intensities over time, suggesting true biological response. The relative intensity of Min-GST & Max-GST (large central band) appears unaffected by drug treatment, however higher MW bands appear to show considerable variability. Both PMSF and NAT lead to an increase in upper MW band intensity, while PMSF leads to a reduction. These results indicate significant drug responses suitable for further study. Bands of interest are indicated.

Additional blots were carried out in order to test the significance of the high MW band intensity shifts observed in Figure 3.5, with the results shown in Figure 3.6. When particular bands are examined relative to GST, changes are considerably more moderate, however MG132 appears to have a specific effect upon the ~51kDa band, differing from the general and less significant actions of PMSF and NAT. It should be noted, however, that while PMSF lacks strong significance, the increased abundances seen would be expected for mono-ubiquitinated proteins during inhibition of vacuolar breakdown, and the MWs involved are similar to that which would be expected for Ubq-GST.

Min-GST and Max-GST do not appear to respond in a significantly different manner in response to drug treatment, with the possible exception of the MG132 ~51kDa effect.

	Relative band intensities normalised to GST		
	~51kDa	~55kDa	~65kDa
MG132			
PMSF			

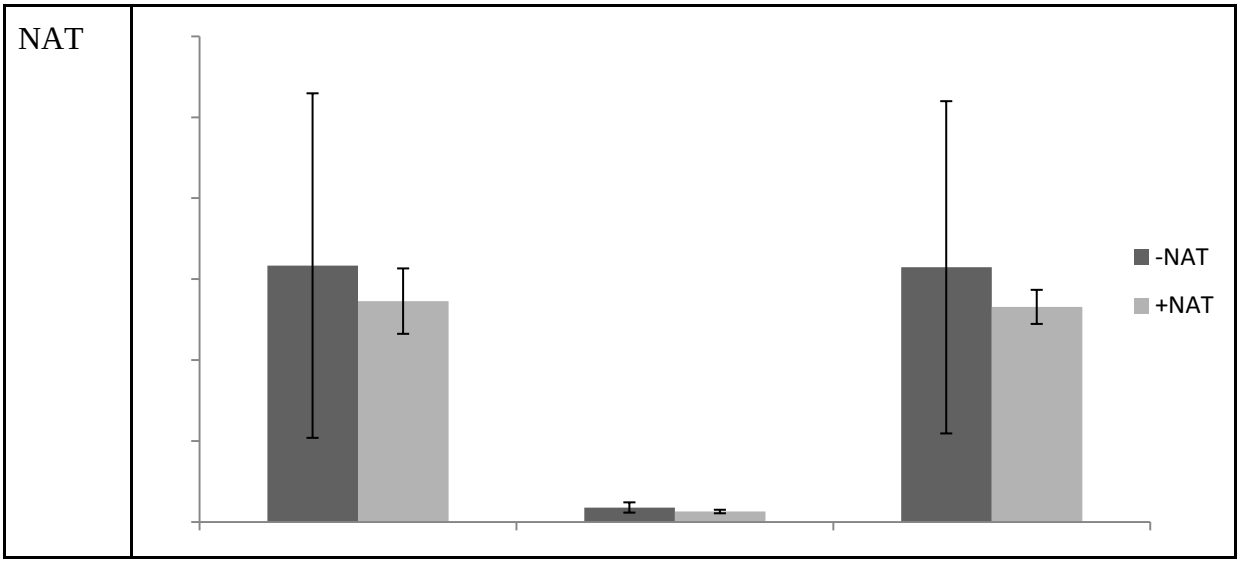


Figure 3.6: Band intensity analysis for high MW bands observed in anti-GST western blotting, with 95% confidence intervals shown. Time since MG132/PMSF addition is indicated in minutes. NAT examined for Min-GST. Units are relative within drug tests. MG132 is observed to significantly decrease the ~51kDa band, reducing levels in Min equal to Max, suggesting a specific target alongside the more general lightening of intensities. PMSF appears to increase all band intensities, however the high level of error leaves only the ~55kDa band increase as significant. NAT, when used without additional drugs, has a minimal effect on specific band intensities.

3.1.4. Immunoprecipitation

It has been suggested that the upper MW bands observed may include Ubiquitin-GST. As mono-ubiquitination results in vacuolar degradation (unlike the proteasomal degradation signalled by poly-ubiquitination) [143], this would likely respond to PMSF treatment. All upper bands were indeed observed to increase in presence of PMSF, and as such immunoprecipitation was carried out to examine this possibility.

Immunoprecipitation for Ubiquitin was followed by both anti-GST and anti-Ubq blots. As both the immunoprecipitation and primary antibodies were produced in rabbits, a control blot, in which a membrane was exposed to secondary antibody only, was carried out, shown in Figure 3.7.1, followed by chemical stripping and reblotting for Ubq, shown in Figure 3.7.2. A standard anti-GST blot was also carried out for comparison in Figure 3.7.3.

The antibody used for immunoprecipitation gives a very strong signal in secondary only blotting, and as such can be clearly seen in the IP precipitant bands for all samples. Anti-Ubiquitin blotting does not reveal any true ubiquitinated bands within the IP precipitant, or the original whole cell sample, indicating the assay to be too insensitive to detect the low levels of ubiquitinated proteins constitutively present in the cell. Anti-GST blotting failed to detect any higher MW bands in the IP precipitant, suggesting that either ubiquitinated species were not among the bands examined in Figure 3.6, or that levels were too low for detection.

Given that ubiquitinated proteins will be enriched in the IP precipitant, and that the high MW bands examined in Figure 3.6 were seen to be easily detectable in Figure 3.5, it can be reasonably concluded that Ubq-GST is not among the bands examined, and is not a majorly present protein in either Min-GST or Max-GST cells.

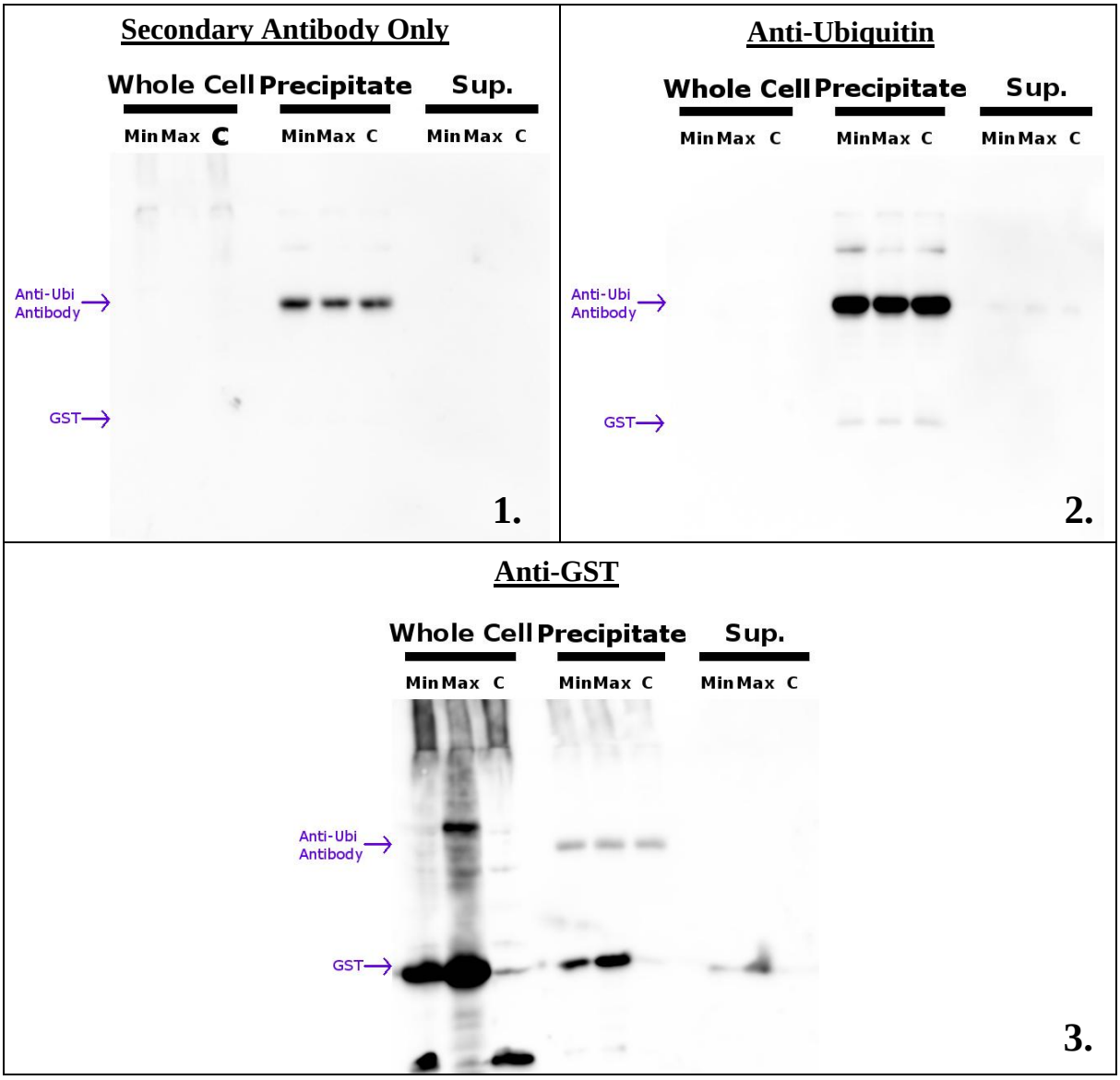


Figure 3.7: Western blots showing whole cell sample, immunoprecipitation precipitant for Ubiquitin, and the immunoprecipitation supernatant (depleted in Ubiquitinated species). 1. Secondary antibody only blot, showing the antibody used in immunoprecipitation to bind directly by the secondary antibody, causing it to appear in all further blots. 2. Stripped membrane from A, re-blotted normally for Ubiquitin, showing only faint bands in the whole cell lane for Min-GST and control, likely representing unrelated polyubiquitinated species. 3. Blot for GST on a separate membrane, showing the usual high MW bands in whole cell samples, but no additional high MW bands in the IP extract, suggesting the presence of no Ubq-GST. The presence of native GST in the IP extract is likely due to the difficulty in excluding this abundant protein. GST is also detected in the IP supernatant, although at a lower level suggesting a degree of protein loss throughout the experimental procedure.

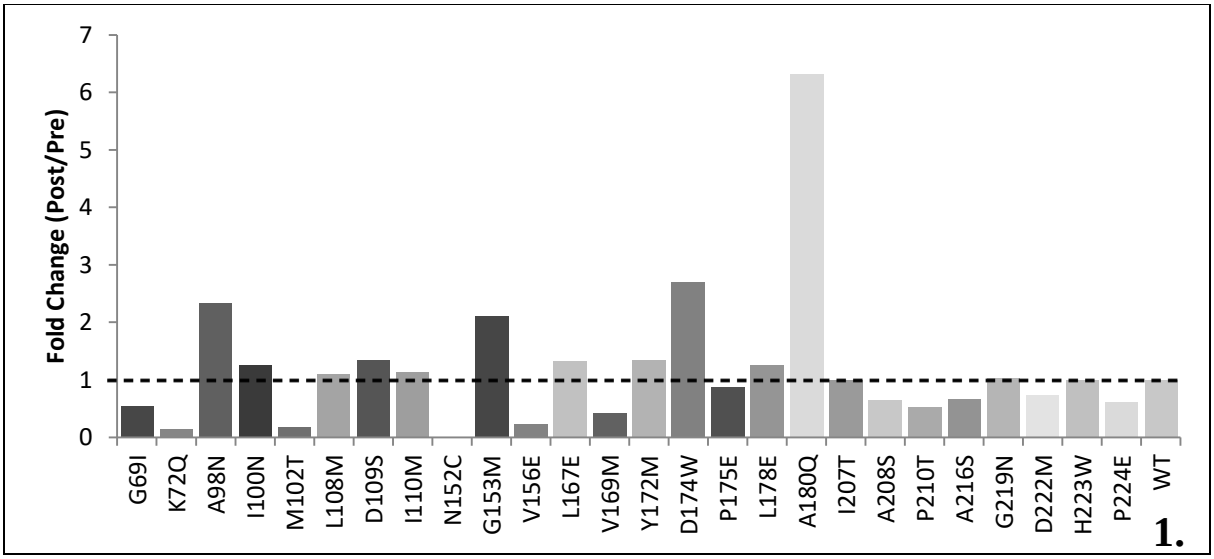
3.2. Computational Analysis of Observed Translational Errors

3.2.1. Observed Fold Changes

In addition to the *in vivo* work in *S. cerevisiae* detailed above, work was carried out upon *E. coli* produced GST in order to better understand the effect of translational errors upon protein function.

Prior to this Msc, Mass Spectrometry was used to examine the impact of various translational errors upon GST binding affinity. GST with a high error rate due to low codon optimisation (Pre) was passed over a glutathione affinity column, and bound protein eluted (Post). Translational error substitutions before (Pre) and after (Post) column treatment were investigated through mass spectrometry with the assumption that substitutions affecting binding affinity should be depleted following passage over the column. The relative fold changes observed (Post/Pre) are normalised to WT and shown in Figure 3.8.1.

The majority of samples varied consistently between a fold change of 6.3 (A180Q) and 0.13 (K72Q); however N152C is seen as immediately anomalous, giving a fold change of 0.00049. This dramatically differing fold change implies a separate mechanism to the other mutants examined. Computational analysis reveals the mutation to introduce a cysteine residue upon the protein surface, within a slight crevice which would prevent GST homodimerisation, but potentially allow direct interaction with GSH itself (along with a neighbouring glutamine residue), as modelled in Figure 3.8.2. If some form of binding were to occur, perhaps through interaction with the GSH carboxylate groups, that was not broken in normal elution, this substitution would be lost in normal analysis; showing strong depletion and explaining this anomalous data point.



1.

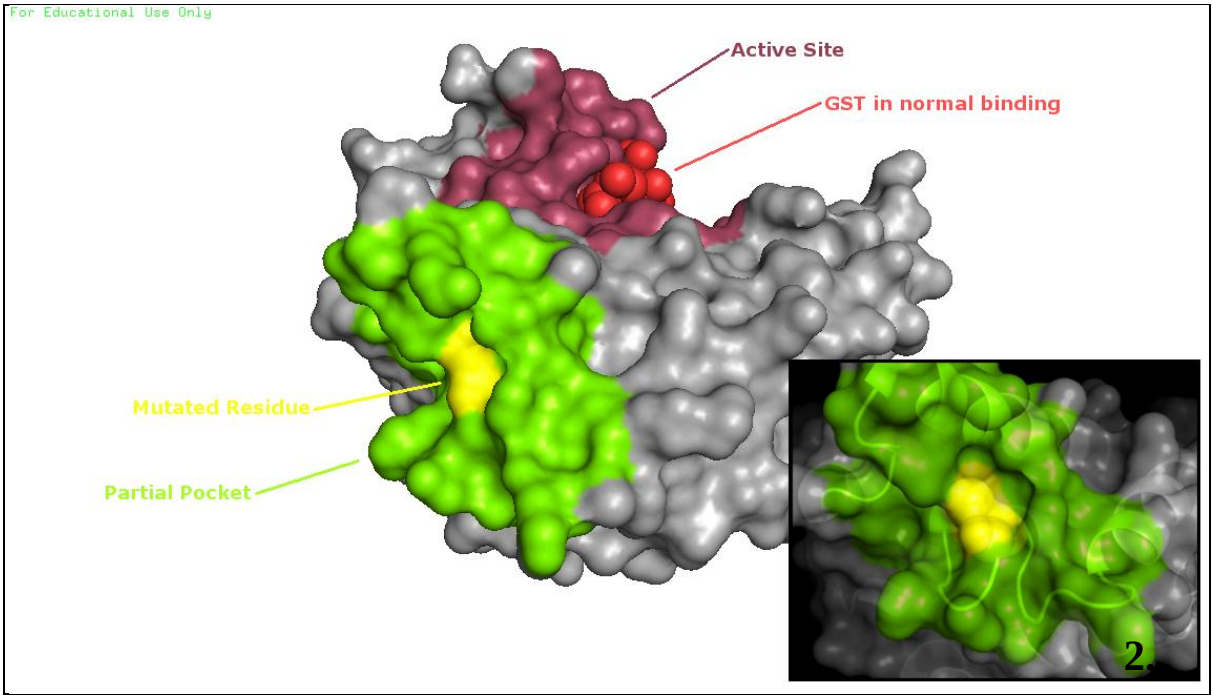


Figure 3.8: Mutations observed through the mass spectrometry experiment carried out prior to this experiment. 1. Relative fold changes of translational errors observed through mass spectrometry. His-tag extraction of Min-GST (pre) was incubated with glutathione magnetic beads. The beads were removed and bound GST molecules eluted for mass spectrometry analysis (post). Results are normalised to WT, shown as a dotted horizontal line. Higher fold change indicates stronger binding affinity. 2. Substitution N152C, known to give anomalously low fold change readings, shown to present a cytosine residue upon the surface within a cleft which, while preventative to homodimerisation, would allow direct interaction with GSH molecules. Insert image shows the mutant cytosine position more clearly.

3.2.2. Substitution Positions

In order to examine this dataset, a series of computational analyses were carried out. As substitutions closer to the active site are expected to affect binding more greatly, distance between the substitution and the active site (measured between the alpha carbon and GSH for consistency) was compared to the observed fold change in an attempt to demonstrate the direct effect of substitutions upon behaviour. The initial plot, shown in Figure 3.9.1 shows a reasonable level of correlation for this linear model, demonstrating substitutions closer to the ligand to more greatly reduce binding, although the residual plot shows a non-random distribution, indicating this linear fit not to be the true model.

The effect of substitution may be considered as an influence which spreads over the 3D area of the tertiary structure, with substitutions closer to the active site likely to cause a greater disruption of binding. A relationship of this kind may be predicted to follow the common inverse-square law, as the disruptive influence (charge alterations and steric effects on neighbouring residues) affects a 3D surface defined at each distance by $4\pi r^2$. As such it can be expected that:

$$Effect\ on\ binding \propto \frac{1}{distance^2}$$

As fold change is measured as Post/Pre, a large disruption of the active site would result in a lowered binding, as such:

$$distance \propto \frac{1}{\sqrt{disruption\ of\ binding}}$$

However as here we here use fold change as a measure of binding effectiveness (post/pre), not binding disruption, it is effectively inverse to the disruption of binding (pre/post).

$$\frac{1}{\sqrt{disruption\ of\ binding}} = disruption\ of\ binding^{-0.5} \propto \left(\frac{Pre}{Post}\right)^{-0.5} = \left(\frac{Post}{Pre}\right)^{0.5}$$

$$\therefore distance \propto \sqrt{\left(\frac{Post}{Pre}\right)}$$

This is plotted in Figure 3.9.2, showing a stronger correlation, and less dispersed residuals, indicating this model to be a better fit to observed data. Substitutions closer to the ligand are seen to reduce affinity in a manner consistent with the inverse square law, with p = 0.0117.

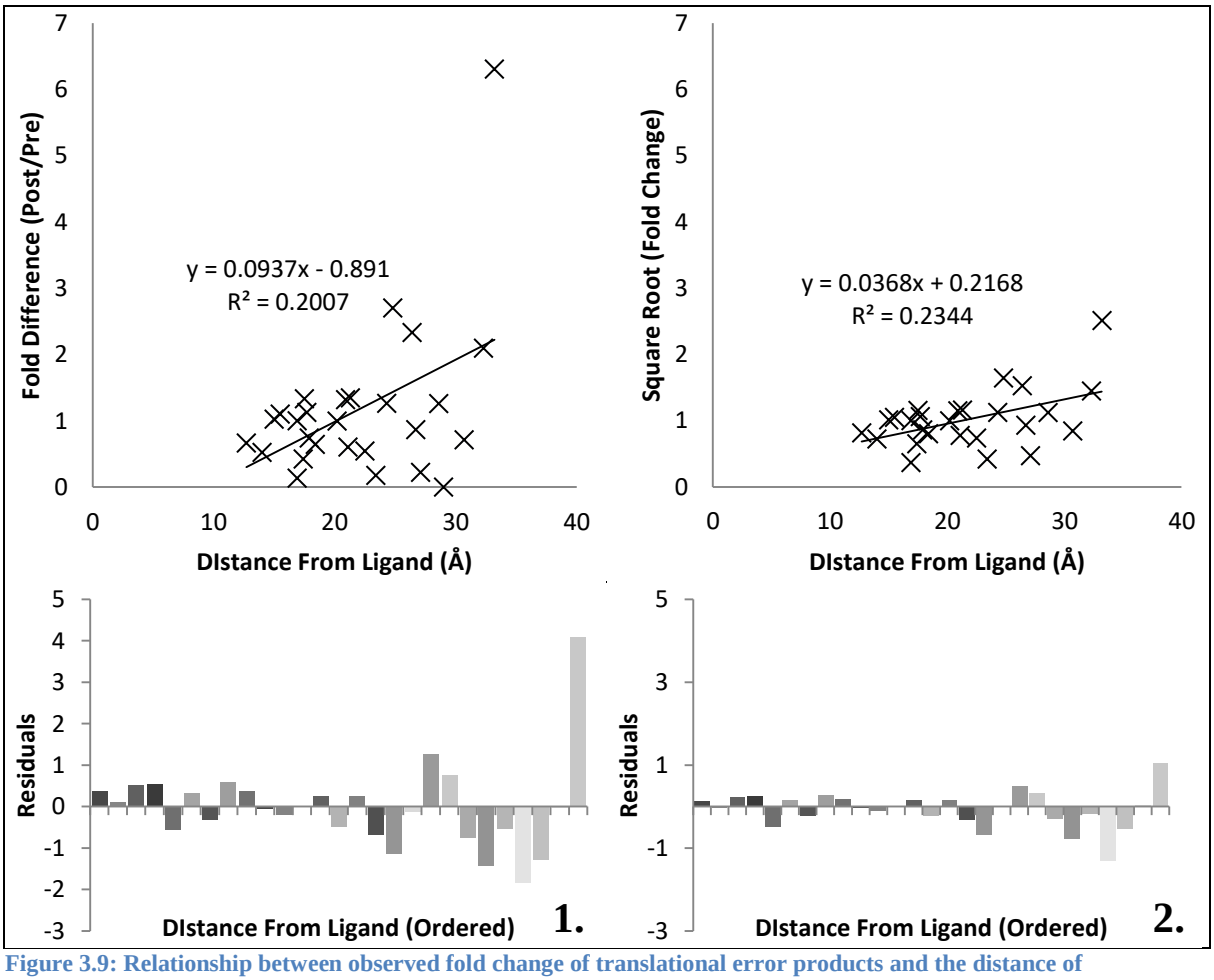


Figure 3.9: Relationship between observed fold change of translational error products and the distance of substitutions from the ligand. 1. Fitting assuming linear relationship between distance and disruption of active site, demonstrating general correlation with p = 0.0211, but significantly asymmetrical residuals (shown in order of increasing distance), indicating this model not to be optimal. 2. Fitting assuming inverse square relationship between distance and disruption of binding affinity, demonstrating stronger correlation with p = 0.0117, and more even residual distribution (shown in order of increasing distance), indicating a better fitting model.

Comparative effects of buried and surface residue substitutions were also examined, due to their likely differing effects upon protein stability. The result, shown in Figure 3.10, indicates no significant difference with either method of surface residue classification.

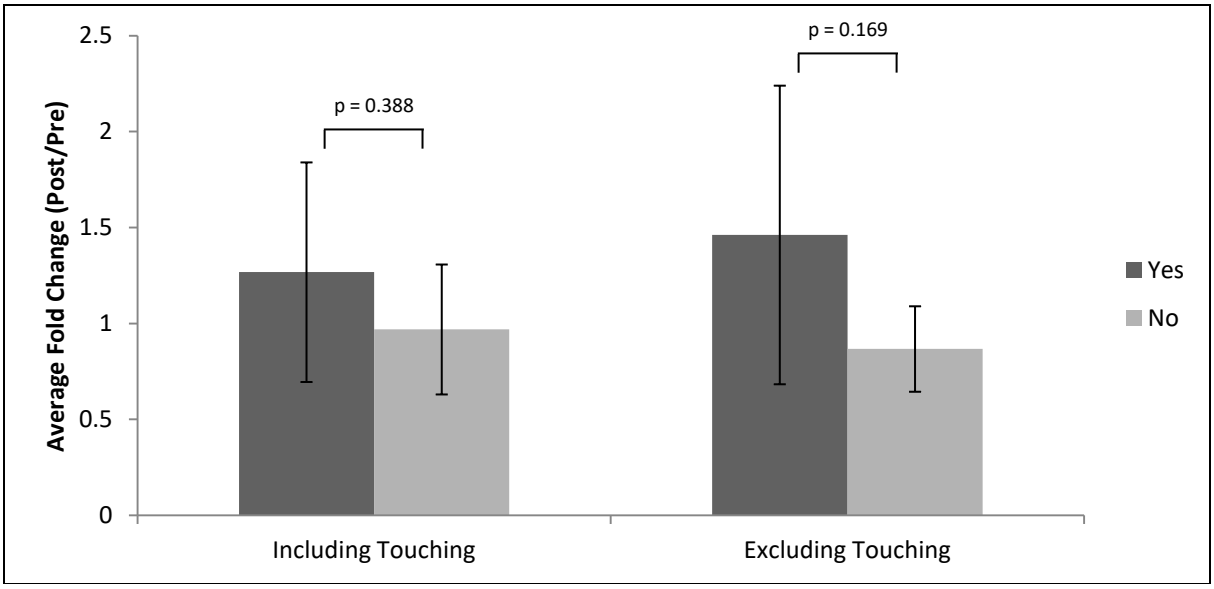


Figure 3.10: Differentiation between surface and buried residues for fold change. Two classifications are examined for surface presentation; including residues simply touching the surface as surface residues, or including only residues fully exposed upon the surface as surface residues. 95% Confidence intervals are shows, as well as t-test p-values between groups. Both classifications fail to find any difference between the groups' effect upon fold change.

3.2.3. Evolutionary Variability and Substitution Scores

Another approach to computational analysis made use of the ConSurf server to calculate the evolutionary variability of each residue, predicting the ability of each position to accommodate substitution errors. When plotted against 1/fold change (i.e. depletion), as seen in Figure 3.11.1, a non-linear relationship was observed, indicating that substitutions upon residues of medium variability lead to the highest observed reductions in binding affinity. An illustrative trend line is given, however it is low fitting and unlikely to represent the true biological mechanism (see discussion - *4.2.3. Influences of Evolutionary Variability and Favourability of Substitution*). The severity of observed substitutions are also indicated in Figure 3.11.1 through circle size; larger circles represent less favourable substitution according to the BLOSUM62 Matrix [134]. No clear influence of substitution severity can be seen.

While no apparent relationships to BLOSUM62 score were seen in Figure 3.11.1, clear plots of fold change and conservation against raw BLOSUM62 were created for the observed translational errors, as seen in Figure 3.11.2 and Figure 3.11.3 respectively. Neither shows significant correlation, indicating that raw BLOSUM62 scores do not have a clear impact upon the binding affinity of translational errors, and the variability of specific residues has no influence on the BLOSUM62 scores observed.

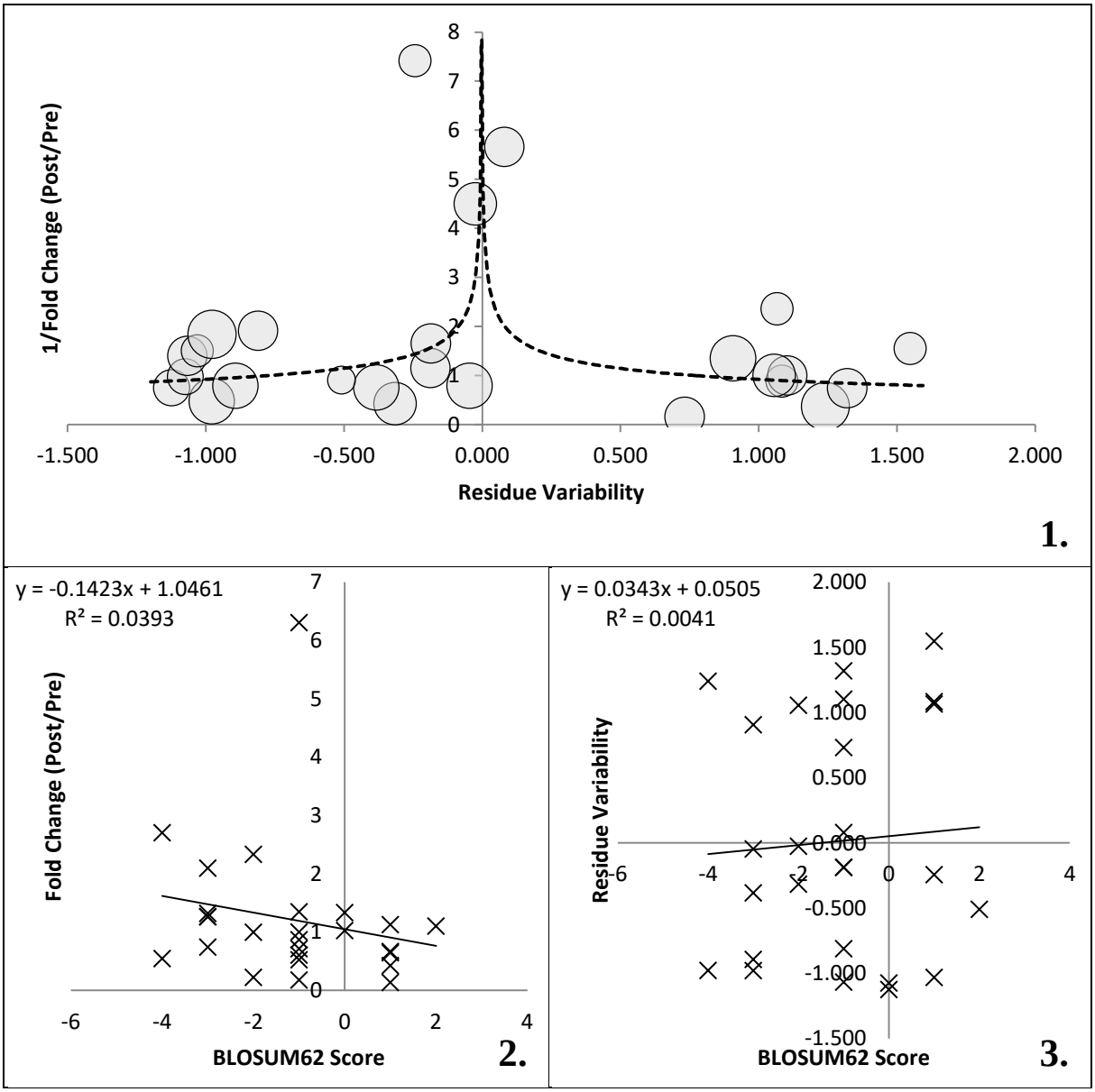


Figure 3.11: Conservation analysis for observed translational errors. 1. 1/Fold change (binding disruption) against residue variability as determined by the ConSurf server, indicating substitutions upon residues of medium variability to result in greatest reduction of binding affinity. Size of circles indicates severity of amino acid substitution observed in that position according to the BLOSUM62 Matrix. For illustrative purposes a regression of y=0.9468(|x|^-0.306) is plotted, with p = 0.116. Scale; 0 = average evolutionary rate. 1 = 1 standard deviation difference. 2. Fold change against raw BLOSUM62 for observed translational errors, demonstrating no significant correlation; p = 0.331. 3. Residue variability against raw BLOSUM62 for observed translational errors, demonstrating no significant correlation; p = 0.756.

3.2.4. tRNA Ratios and Frequencies of Substitution

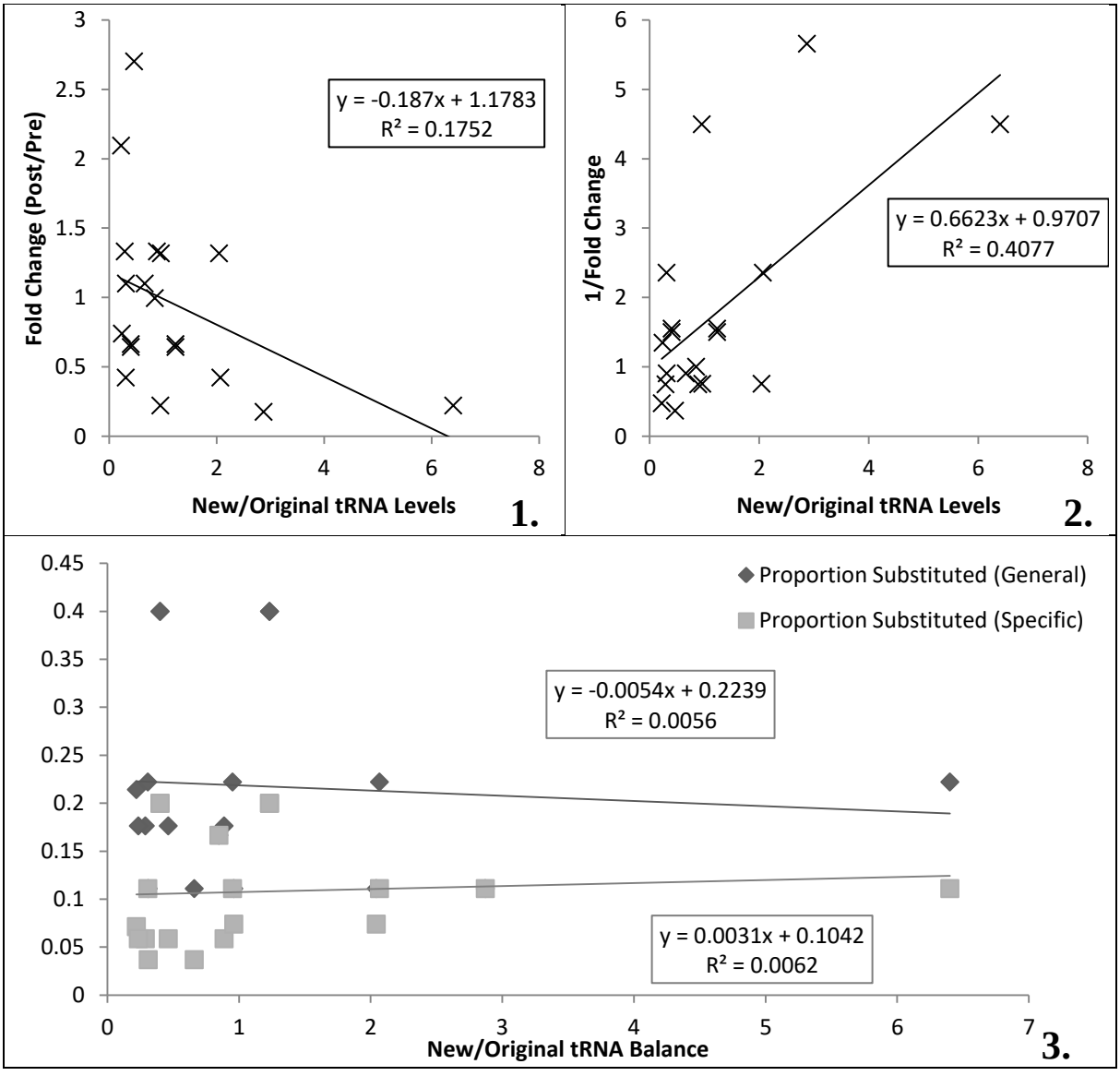
As translational error frequency is highly dependent upon the difference in relative tRNA abundance between the erroneous and correct amino acid (see *1.2.2.1 Missense Errors - tRNA balance importance*), this ratio is calculated for observed translational mutants based on available literature data, and compared to various other factors in Figure 3.12. This ratio is referred to as New/Original in figure labels, and shows "abundance of tRNAs delivering the observed amino acid"/"abundance of tRNAs delivering the expected amino acid". In cases where multiple tRNAs represent the same amino acid, all possible tRNA abundances were plotted. While this will reduce the strength of observed correlation, the plotting of all

possibilities avoids manual selection of tRNA ratios in order to increase correlations, and thus minimises bias.

Comparison of fold change to the tRNA ratio in Figure 3.12.1 demonstrates some correlation, appearing to follow an inverse relationship. The suitable transformation, in Figure 3.12.2, shows a very strong correlation, indicating that substitutions involving high ratios of new/original tRNA levels lead to the highest disruption of binding.

The proportion of each amino acid observed as substituted within the protein, either generally (e.g. the proportion of Tyrosine positions substituted), or according to a specific substitution (e.g. YxM), is compared to the relevant tRNA ratio for each substitution. The results, in Figure 3.12.3 show no relationship of either substitution measurement to tRNA proportions within the cell.

These amino acid substitution proportions were also compared to observed fold change in Figure 3.12.4, where specific substitutions showed negative correlation at p=0.0416, indicating that rarer substitution pairs result in higher levels of disruption.



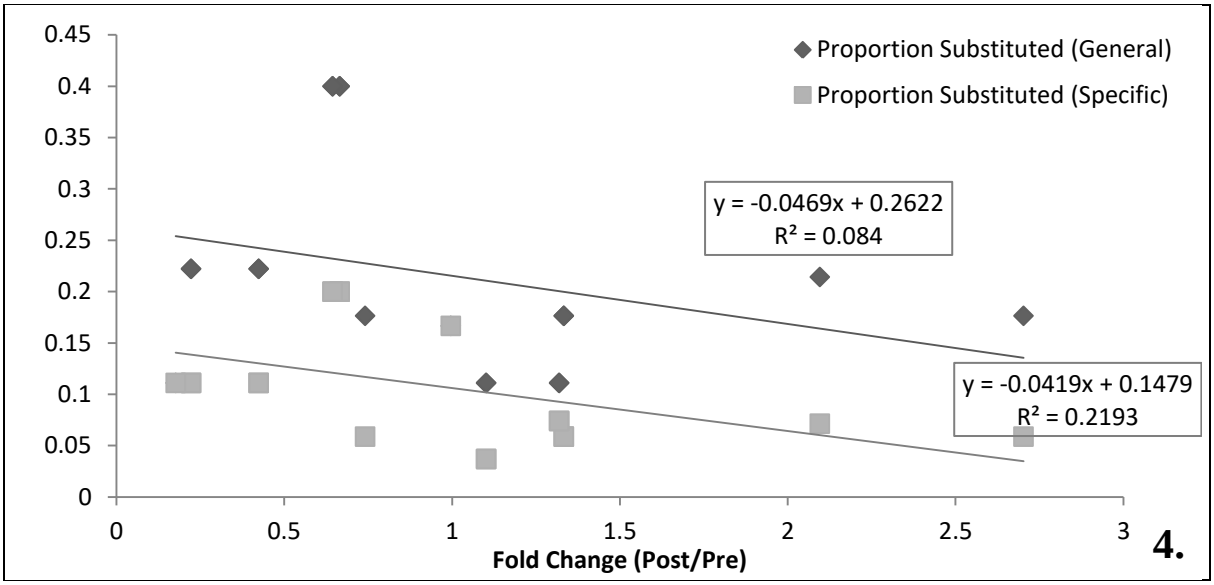


Figure 3.12: Examining possible effects of the tRNA abundance ratio ("New tRNA abundance" / "original tRNA abundance"), and the level of substitution observed for differing amino acids. 1. Fold change against tRNA abundance ratio produces a reasonable correlation (p = 0.0727), but appears to show an inverse relationship. 2. 1/Fold change against tRNA abundance ratio showing a very strong correlation (p = 0.0029), indicating that substitution with an amino acid of higher tRNA abundance leads to a greater disruption of binding affinity. 3. Proportion of various amino acids within the protein observed to be substituted, either generally or according to a specific substitution (e.g. YxM), against the tRNA abundance ratio. No correlation is seen (p = 0.7604 and 0.7482 respectively), indicating the likelihood of observational translational errors for each amino acid not to be based on tRNA abundance. 4. Proportion of various amino acids within the protein observed to be substituted, either generally or according to a specific substitution (e.g. YxM), against the observed fold change. While specific substitutions show no clear correlation (p = 0.227), general substitutions show significant correlation (p = 0.0416).

3.2.5. Analysis of observed and non-observed substitution positions.

While translational error substitutions were detected for many residues within the protein, a majority of residues showed no substitution.

Experimental examination of an observed substitution compared to a non-observed substitution was carried out in *3.3. Experimental Analysis of Translational Mutants* for Y172M compared to Y15M; however computational analysis was also carried out for a range of possible YxM substitutions based on a Consurf multiple sequence alignment for 150 homologues.

The evolutionary variation of several tyrosine residues can be seen in Figure 3.13.1, showing that the observed substitution position, Y172, has a significantly higher evolutionary variability than all non-observed substitutions.

It was seen that Methionine was present in position 172 for 29.3% of homologues, compared to only 3.3% containing Tyrosine in this position, as shown in Figure 3.13.3. This is in strong contrast to all non-observed substitution positions, wherein no homologues contained methionine.

In order to test whether these results represent a general case, all positions were examined.

For each amino acid, residues in MS detected regions were defined as observed or non-observed substitution sites. Average variability for each set was calculated, and the differences between observed and non-observed (i.e. "observed" - "non-observed"), shown in

Figure 3.13.4, demonstrating a general trend for higher variability in observed regions to be found for a wide range of amino acids.

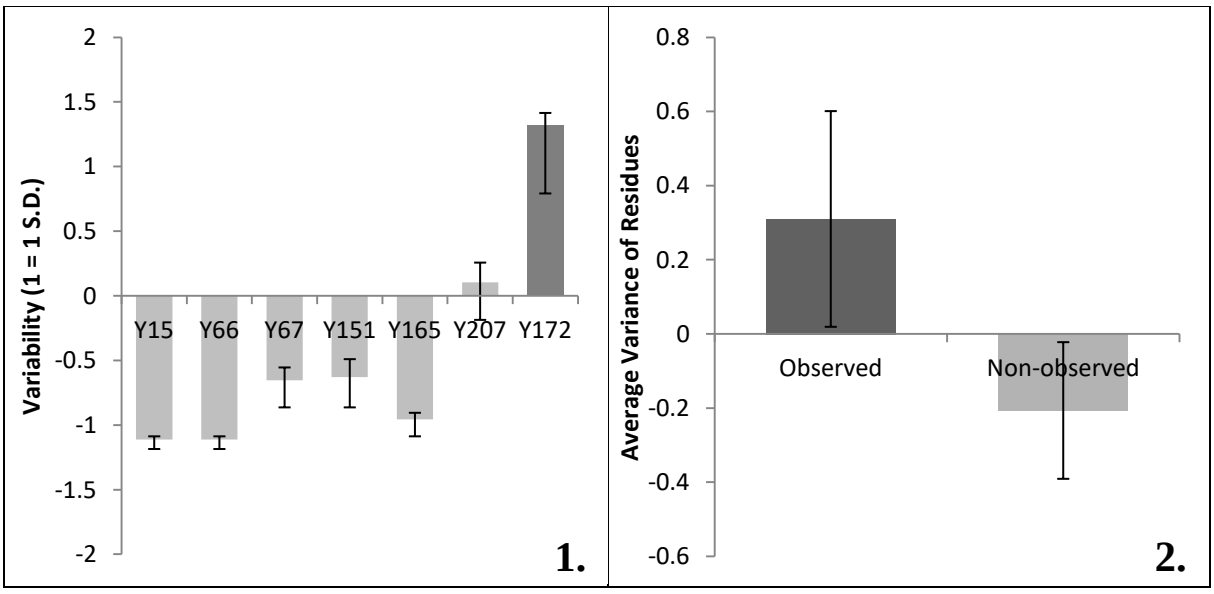
As this trend appears to be common for a range of amino acids, not simply a feature of the initially examined YxM substitution, statistical comparison of the evolutionary variability of MS detected substituted and non-substituted positions was carried out, shown in Figure 3.13.2. It can be seen that the two groups show significant difference, with p = 0.00418, demonstrating variability to be an important factor in differentiating these groups.

The above analysis of YxM also found that for the observed substitution, Methionine was present in many homologous versions of GST. In order to compare this with the larger data set of all substitutions, the question was asked: Are substitutions seen to introduce amino acids which also present in homologous versions of the protein more often than would be expected by chance alone?

The number of matches between homologous sequences and random substitutions expected by chance alone is calculated in Table 3.1, and shown to be largely similar to that observed in MS data. As such, while the presence of a substitution in homologous sequences indicates a stable alteration, this is not a clear factor in the detection of translational error substitutions.

Number of alternate amino acids found in homologues	1029
Number of residues in protein	218
Average homologue alternatives per residue	4.72
Total substitutions observed	75
Each substitution changes an amino acid to one of 19 alternatives. This has a 1/19 chance to match 1 random homologue alternative.	
Chance of match for each substitution	4.72/19 = 0.248
Expected number of matches	0.248*75=18.6
Actual number of matches	15
P value from chi squared	0.332

Table 3.1: Expected number of matches between translational error substitutions and alternative amino acids found within homologues. The expected 18 matches is not seen to be significantly different to the observed 15 matches, indicating no relationship beyond chance alone. (Homologues generated through 150 sequence MSA in Consurf.)



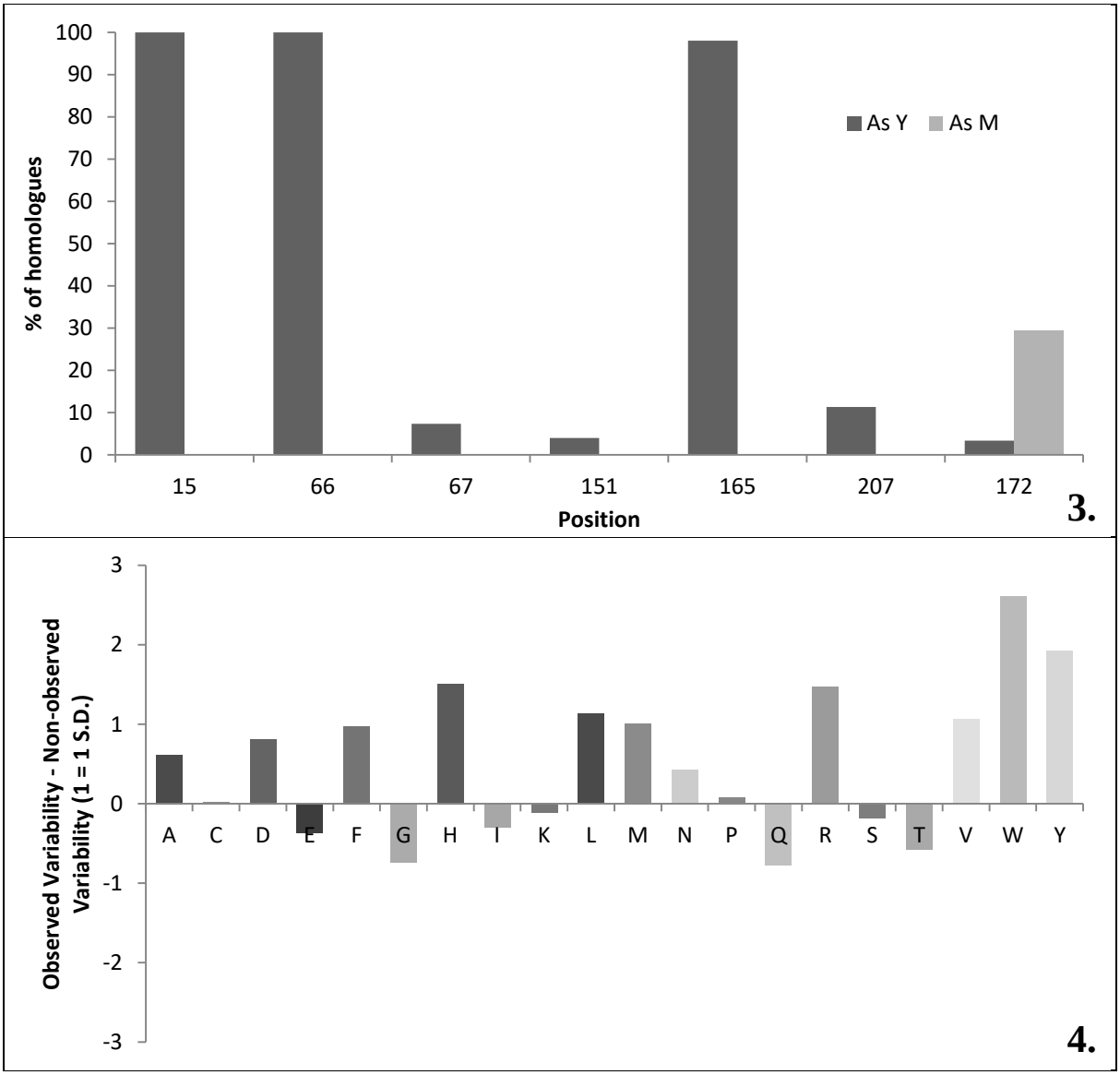


Figure 3.13: MSA based analysis of the observation of translational error substitutions in mature GST. 1. Evolutionary variability for several Tyrosine residues, demonstrating the observed substitution, Y172M, to occur on a position of higher variability than all other, non-observed, substitutions examined. 50% confidence intervals are shown. 2. Average variability for all positions with observed translational error substitutions, and all positions with no observed substitutions. 95% CIs are shown, demonstrating a significant difference between these groups with p = 0.00418; however, it may be noted that the standard deviations for observed and non-observed positions are 1.05 and 0.95 respectively. 3. Homologue occurrences of Tyrosine and Methionine in the positions examined for YxM substitution. It can be seen that whilst Tyrosine is present at varying frequencies in all positions, Methionine is present only in the 172 position, where the YxM error substation was observed. 4. Differences in variability between the positions of observed translational error substitutions, and positions with no observed errors, showing that for the majority of amino acids, there is a greater variability in translational error positions.

3.3. Experimental Analysis of Translational Mutants

3.3.1 Mutant Production

The degree of depletion of substituted proteins observed through the MS experiment was used to make predictions about the functionality of substituted proteins. In order to test these predictions, "pure" populations of a selection of translational error substitutions were created through genetic mutation. These mutants were expressed in *E. coli* and purified for individual analysis.

The MS experiment also indicated that substitutions only occurred upon certain codons; for example tyrosine to methionine substitutions occurred on some tyrosine codons, but not upon the same codon at other sites. As the MS assay only detects mature proteins within the cell, it was hypothesised that only substitutions which allowed basic protein folding were observed, whereas more destabilising substitutions led to protein degradation. In order to test this hypothesis, the observed substitution Y172M was created as a mutant, alongside the non-observed substitution Y15M, allowing direct comparison.

The translational mutants used are summarised in Table 3.2. Conservation values are generated through the ConSurf server and mapped to the PDB model 1UA5 from Kursila et al (to be published), [144] allowing mutations to be examined.

First, effects on protein stability were examined through western blotting of transformed cell cultures, shown in Figure 3.14. All mutants give far lower abundances than the WT, directly demonstrating lower levels within active cells. The non-observed mutant, Y15M, shows around half the abundance of its observed counterpart, Y172M, indicating a significantly lower presence within the cell. In fact, Y15M shows a lower abundance than all observed translational mutants, suggesting low stability to differentiate it from this group; i.e. the lowered stability explains why Y15M was not observed in the MS experiment, while Y172M was. Coomassie staining (not shown) revealed similar abundances to those seen here.

Mutant	Role of Residue	Conservation	Predicted effect of mutation	MS Fold Change (Post/Pre)
Y15M	*Positioned near the active site, the residue is seen to hydrogen bond with the unpaired NH group of a neighbouring helix, acting as a functional N cap. It is mentioned in literature as a binding residue for GSH* [145].	*Very high conservation residue, located within high conservation area due to presence in Active Site.*	*Removal of capping function introduces a polar helix terminal within the tertiary structure, close to the active site. Likely to affect local stability and protein activity. Does, however, fit with little steric hindrance.*	*Not Observed*
V156E	Buried hydrophobic residue.	Medium conservation residue, located within area of medium conservation.	Would introduce charge into hydrophobic core, however it is possible to expose charged group on the surface through slight steric clashing.	0.222
Y172M	Exposed on surface as hydrophilic residue.	Very low conservation residue, located within area of low conservation.	Able to lose charge without greatly affecting surface, and is able to occupy a highly similar space to the original residue. Unlikely to result in strong disruption.	1.347
A180Q	Exposed on surface as part of a largely hydrophilic region.	Low conservation residue, located within low conservation area.	The introduced charge can likely bond to water without difficulty, and shows no steric hindrance. Unlikely to result in strong disruption.	6.307

P210T	Buried hydrophobic residue.	High conservation residue, located within low conservation area.	Introduces dipole into hydrophobic region and replaces proline, a structurally unique amino acid; although shows good steric fitting. Likely to affect stability.	0.523

Table 3.2: Summary of experimentally examined mutations and their expected effects based upon conservation data from the ConSurf Server, and 3D modelling within PyMol.

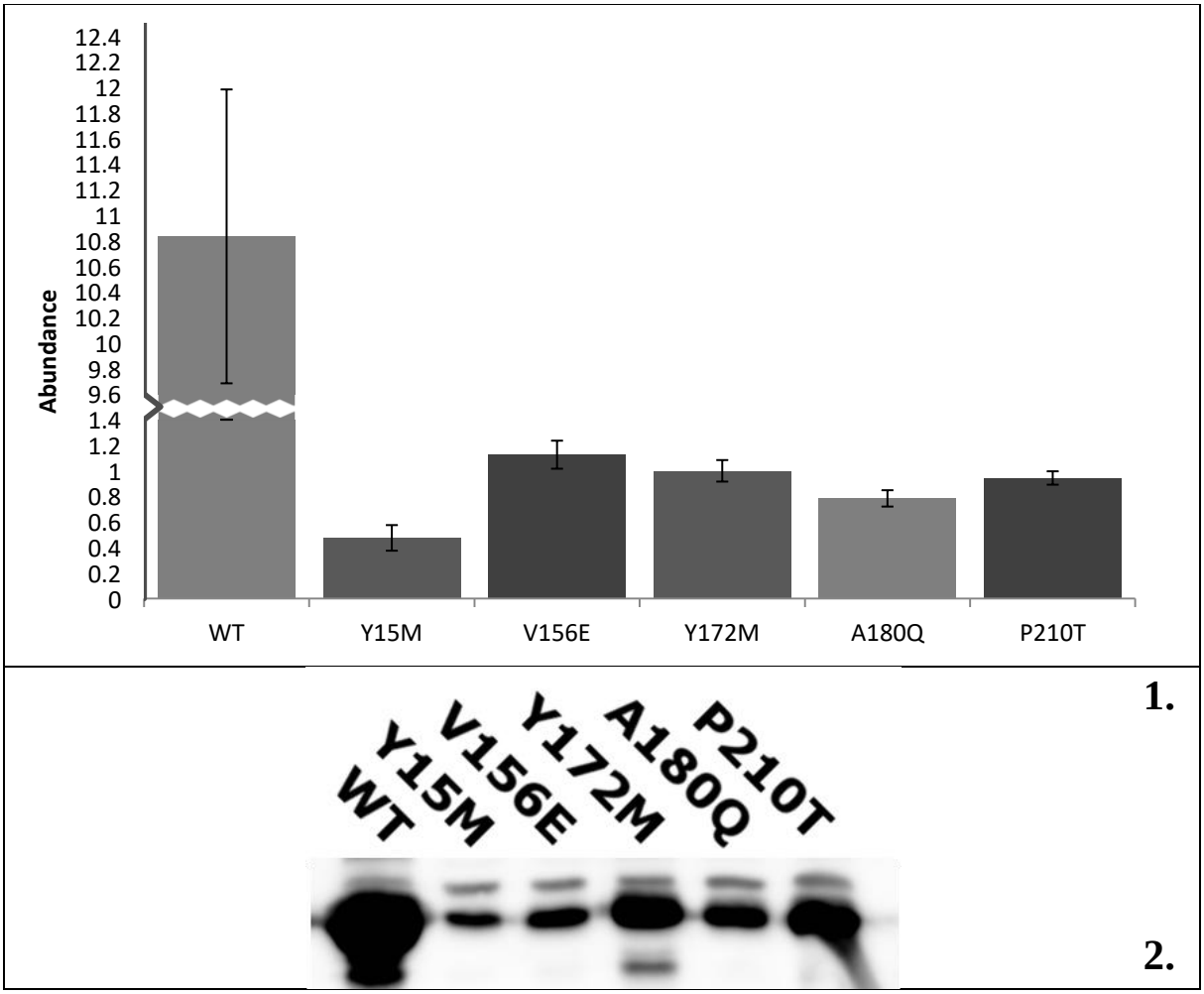


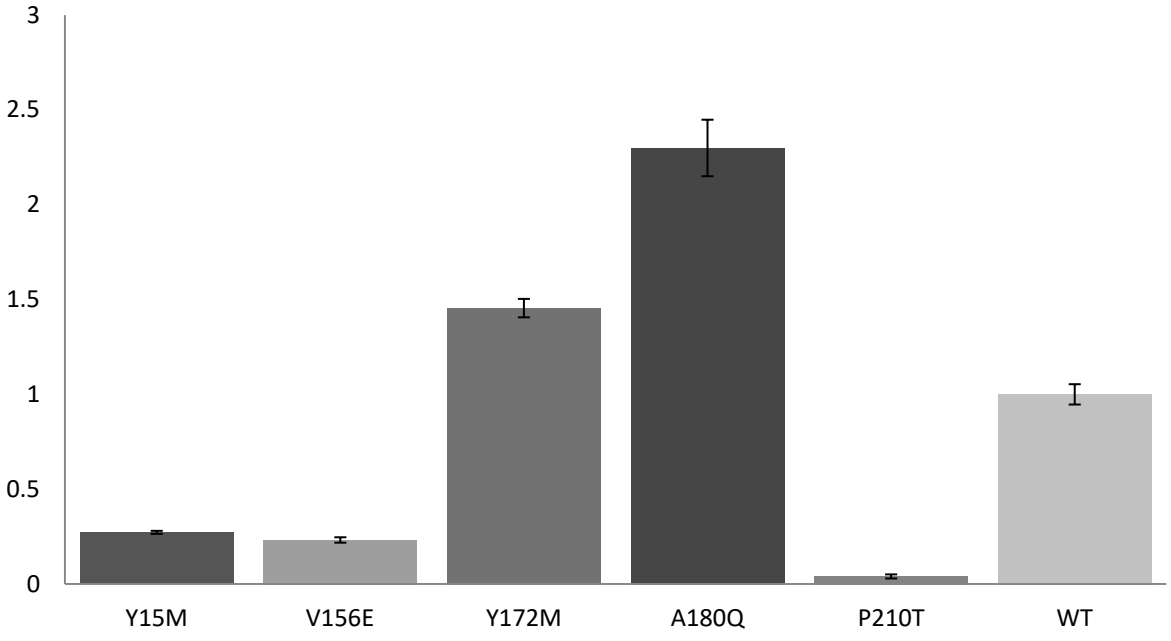
Figure 3.14: Abundance of various GST mutants. 1. Abundance of mutants quantified over multiple repeats with various load levels, normalised to Y172M abundance, 95% confidence intervals shown. All mutants show far lower levels than the WT (note the split y axis), with the non-observed Y15M giving around half the abundance of the observed Y172M. 2. Example western blot showing differing levels of GST present for each mutation.

3.3.2. Mutant Binding Affinities

The effect of these mutations upon binding affinity was examined similarly to the original MS experiment. Purified samples (Pre) were incubated with glutathione beads, the supernatant removed (Sup), and bound protein removed from the beads (Ext). These samples were then run on SDS-PAGE gels and visualised through coomassie blue staining, as exampled in Figure 3.15.3. Band quantification allowed an estimation of relative binding affinities.

The results, shown in Figure 3.15.1, show a range of binding affinities, both greater and lesser than the WT sample. The non-observed mutant, Y15M, shows a far lower affinity than its observed counterpart, Y172M, but a higher affinity than V156E and P210T. As such it does not appear that binding affinity is involved in determination of which translational errors are observed.

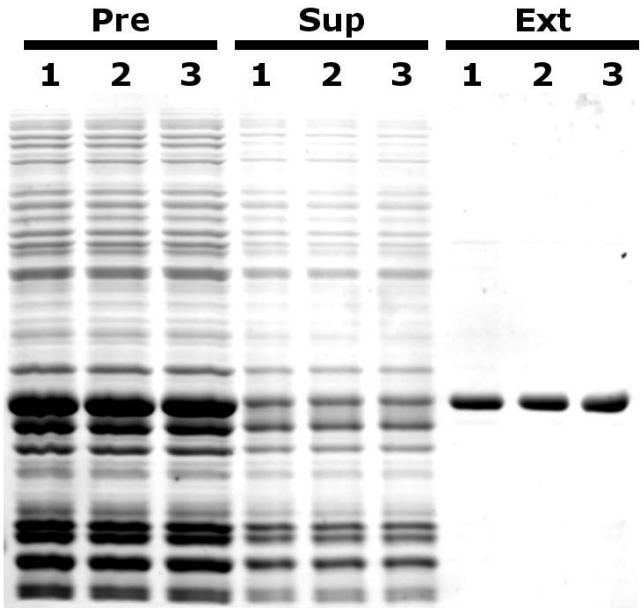
Figure 3.15.2 shows that while fold changes for DNA mutants and translational error mutants do not correlate perfectly due to experimental differences, the two experiments show general agreement; increase or decrease of affinity in relation to WT is maintained across experiments. Thus, while numerical values vary, the functional effect of both mutants types are maintained.



1.

	DNA mutant fold change	Translation error fold change
Y15M	0.272	
V156E	0.232	0.222
Y172M	1.455	1.347
A180Q	2.299	6.307
P210T	0.040	0.523
WT	1	1

2.



3.

Figure 3.15: Results of the GST binding assay. 1. "Extract on beads" / "Sample pre-assay" normalised to WT, where stronger binding mutants show higher values, with 1 SD shown. A range of repeatable binding strengths can be seen, with Y15M, V156E, & P210T binding more weakly than the WT, and Y172M & A180Q binding more strongly. Additionally, while the non-observed mutant (Y15M) has a far lower affinity than its observed counterpart (Y172M), other observed mutants show lower affinities (V156E & P210T). 2. Comparison between the fold change shown in part 1 (DNA mutant fold change), and the fold change observed in the original MS experiment for each mutant (translational error fold change). Due differences in experimental procedure there is not an exact match between the datasets, however general agreement can be seen; increase or decrease of affinity in relation to WT is maintained. 3. Example gel used, here for A180Q, showing triplicate samples for; pre-assay (Pre), the supernatant removed from the beads with lowered GST levels (Sup), and the extract removed from the beads (Ext).

4. Discussion

4.1. *S. cerevisiae* Error Control

4.1.1. Production Levels

The maximum optimisation codon sequence chosen for this project contained codons representing high abundance aatRNAs, and the minimum optimisation codon sequence contained codons representing low abundance aatRNAs. As a result, *S. cerevisiae* expression of Max-GST was observed at seven times the level of Min-GST in Figure 3.1. This difference likely represents multiple cumulative factors; higher production of Max-GST due to faster codon decoding speed [22][23], higher error rates in Min-GST due to increased aatRNA competition leading to production of unstable proteins [6][24][108], (also see *4.3.1. Stability and Production of Translational Mutants*), possible breakdown pathways, and the possibility of Min-GST translational error proteins failing to be recognised by the anti-GST antibody used.

While the relative contributions of these factors are difficult to ascertain, the results show the levels of folded GST molecules available in exponentially growing cells for both sequences, as western blotting (opposed to general protein staining) depends upon antibody binding to the folded protein. As only correctly folded proteins will be of use to the cell, western blotting provides highly biologically relevant quantifications, and here confirms the evolutionary importance of codon optimisation for high demand proteins, which has been much discussed in previous literature [39][41][107][108]. Additionally, this clear biological difference validates the use of these constructs within this project.

No clear breakdown products were observed, however high MW bands were present in non-GST samples. As such, the following knockout and inhibitor studies were carried out in an attempt to observe breakdown products, and understand the nature of the high MW bands.

4.1.2. Knockout Screening

The knockout studies, examining deactivation of the surveillance pathways in Figure 3.2 (No Go Decay [100][101], Non Stop Decay [101][138], and Nonsense Mediated Decay [139]) and a selective mRNA degradation pathway in Figure 3.3 (*dhh1* [141]) failed to observe any clear change in Min-GST production or breakdown products, indicating no clear role for these pathways in the breakdown of translational error products. It may be noted that *dhh1* mutation lead to a general rise of protein levels due to the increased mRNA levels within cells, although demonstrated no clear selective effects.

Failure to demonstrate a role in translational error does not, however, give grounds to claim no influence of these pathways. It is possible that deactivation of these pathways lead to an increased proportion of produced protein containing translational errors without being detected by these experiments, as discussed more fully in *4.1.5. S. cerevisiae Error Control: Conclusions and Limitations*.

Further work may be required to determine the effect of deactivating these pathways on translational error detection. If random errors are stable enough to detect, repeating the MS experiment upon mutant cell lines expressing Min-GST could provide insight into changing error profiles. If the random errors were not stable enough to reliably detect, experimental analysis would be more difficult, although alternative approaches are discussed in *4.1.5. S. cerevisiae Error Control: Conclusions and Limitations*.

4.1.3. PMSF & MG132 Assays

PMSF was used to inhibit vacuolar degradation [142], and MG132 to inhibit proteasomal degradation [142]. If error products were largely degraded through one of these mechanisms an increase in Min-GST relative to Max GST, or a build up of breakdown products, would be observed [142]. NAT, known to increase translational error rates [113], was added in order to increase the magnitude of possible effects .

No clear effect upon GST levels or breakdown products were observed in Figure 3.5, although as in the knockout studies above, this does mean that error levels were not altered. If the level of translational mutants destined to be degraded through one of these pathways were low compared to overall GST production, or the error protein was easily diverted into the ordinary GST turnover pathway, no clear difference would be observed.

Examination of upper band variation relative to GST intensity in Figure 3.6 suffered from large confidence intervals due to dividing one variable band by another, thus combining the errors inherent to western blotting. As such it is difficult to state with significance many of the drug effects, or rule them out with high confidence; however large enough variation could still be observed.

NAT appears to have little effect upon upper MW band intensities, although due to the high error levels, it is possible that effects were simply undetected in this experiment. PMSF is seen, although with low significance, to increase the intensity of all upper bands due to the deactivation of general vacuolar degradation. MG132, on the other hand, showed differing effects upon differing bands as a result of proteasome inhibition. The bands are discussed separately below:

The ~51kDa band in MG132 appears to show higher abundance for Min-GST samples than Max-GST, but is reduced to equal levels for both Min-GST & Max-GST in the presence of MG132. While this suggests a mechanism which acts selectively upon the high error version, resulting in abundant higher MW bands in the absence of MG132, it should be noted that no such effect is observed in the absence of PMSF or NAT. Due to the high errors involved, it is impossible to either validate or reject this possibility based on conflicting data.

The ~55kDa band is a strong candidate for a GST derivative destined to be broken down through vacuolar decay as is seen to consistently match GST levels prior to drug treatment, give no response to MG132, and greatly increase in abundance in response to PMSF treatment. Both Min-GST and Max-GST, however, behave in an identical manner suggesting

that this derivative, possibly Ubq-GST, has no relationship to the translational errors here examined.

The ~65kDa band shows very wide confidence intervals in MG132 study, and very weak signals in PMSF studies, preventing clear conclusions being drawn as to its nature.

The drug levels used in this experiment are demonstrated in Figure 3.4 to be appropriate. The slight significant decreases in growth rate indicate a true biological effect upon the cells, but sufficient cell growth that all readings will be representative of viable cells. The drug testing discussed above can therefore be accepted as relevant to general cell function beyond these tests. Due to the low level of growth inhibition, however, there must be sufficient activity within the cells despite inhibition (e.g. some vacuolar degradation with PMSF, some proteasomal activity with MG132, and non-fatal error levels with NAT). As such, effects of drug addition are likely to be subtle and difficult to establish through western blot. It may be preferable in future studies to add far higher levels of these drugs for short times (around 1 generation) in order to fully deactivate the target systems and observe more dramatic alterations in protein levels, producing signals which are likely to be more easily detectable through western blot.

As the alteration of high MW bands in response to drug treatment was seen as significant in some cases, and indicated the possible presence of mono-ubiquitinated GST molecules, further experimentation was carried out in order to establish if these bands truly represented GST, or simply cross reactions.

4.1.4. Immunoprecipitation

The possibility of mono-ubiquitinated GST production was raised, first through the presence of suitably high MW molecules in anti-GST western blots, and then through the observation that PMSF vacuolar decay inhibition lead to an increase in the abundance of these high MW molecules. In order to investigate this, immunoprecipitation was carried out targeting Ubiquitin, creating the three samples seen in Figure 3.7; whole cell extracts containing all proteins at their native levels, IP precipitant enriched for ubiquitin containing molecules, and the Supernatant from IP extraction depleted in ubiquitin containing molecules.

The antibody used for ubiquitin binding was produced in rabbit, causing the secondary antibody to bind directly to the immunoprecipitation antibody in all blots. This was observed in Figure 3.7.1, where only secondary antibody was added, showing strong bands in the IP precipitant. These bands were of similar MW to the high MW GST products being examined for, and thus carry a chance of obscuring positive results.

Ubiquitin blotting failed to reveal any clear bands in the whole cell extract, or the ubiquitin enriched IP precipitant, indicating that any ubiquitin containing molecules present were in concentrations too low to detect using this antibody (as whole cell extract is known to contain ubiquitin molecules, regardless of the effectiveness of the IP itself).

GST blotting revealed the expected bands within whole cell extract, and levels of native GST within the IP precipitant. Due to the high levels of GST present it is reasonable that the IP process failed to exclude all of this abundant protein, however it may be noted as strange that the IP precipitant appeared to contain higher levels of GST than the supernatant itself - perhaps due to the concentrating effect of resuspending the beads in a lower volume of sample buffer causing contaminating GST molecules to appear enriched. GST blotting failed to reveal any higher MW bands in either the IP precipitant, or the supernatant. Failure to

detect the high MW proteins in the supernatant, where general protein levels appear low as a result of experimental procedure, may be a result of lowered concentration. The IP supernatant, however, will be enriched for ubiquitin containing proteins, and as such failure to detect the normally clear high MW bands in the ubiquitin enriched sample through GST blotting gives a clear indication that these bands do not contain significant levels of a GST-Ubiquitin conjugate.

This experiment may benefit from concentration of the supernatants in order to better observe protein levels within it. Drug treatment would also be of use, with MG132 inhibiting proteasome decay for poly-ubiquitinated proteins and PMSF inhibiting vacuolar decay for mono-ubiquitinated proteins, leading to an overall increase in the concentration of ubiquitin containing proteins [113]. Additionally an alternate IP antibody which does not react with the secondary blotting antibody would be of use, as high exposure of the current western blots causes the IP antibody to obscure a significant area of gel (not shown) .

4.1.5. *S. cerevisiae* Error Control: Conclusions and Limitations

The dramatic differences in GST levels between minimum and maximum optimisation samples indicate clear differences between the constructs. This was discussed as representing multiple factors; differing production rates, stability, breakdown, and binding to the anti-GST antibody. While this demonstrates the levels of useful, properly folded, protein within the cells, it caused problems for later analysis; to what extent are differences in GST abundance due to differing error rates, and what proportion of errors are stable enough to detect? Altering error levels will only be observed as effecting GST levels if errors make significant contribution to overall abundance.

Knockout and drug inhibition experiments screened for a degradation pathway which would, when deactivated, allow maturation of translational error products, significantly increasing the abundance of Min-GST relative to Max-GST, or produce observable decay products.

As no decay products were observed under any conditions, either due to their non-existence, rapid degradation, or low abundance, this line of observation failed to demonstrate the role of any decay pathway. As such, the initial question, why error levels were seen as depended upon codon optimisation in *E. coli* but not *S. cerevisiae*, cannot be answered based upon current data.

The observation for alteration in GST levels relies on two assumptions: 1. The lower levels of observed Min-GST are significantly due to the degradation of error products, not simply factors such as lower production rates, and as such the effect of error product degradation is detectable through western blotting. 2. If error products were given the opportunity to mature, they would largely fold correctly and be detectable by western blotting.

Assumption 1 is questionable as western blotting here has been unable to clearly detect the known effect of *dhh1* deletion upon this system. *Dhh1* is involved in the selective degradation of mRNAs with high numbers of ribosomes present, such as the slow non-optimised mRNA in this experiment. As such its deletion should show a significant increase in Min-GST levels compared to Max-GST [141]. As this was not seen, it demonstrates that while overall levels of Min-GST and Max-GST vary greatly, differential effects due to codon choice may not always be able detectable with these methods. Assumption 2 is questionable

in light of the role of stability in determining whether a translational error is observable (*see 4.2.6. Computational Analysis of Observed Translational Errors: Conclusions* & *4.3.3. Analysis of Translation Mutants Initially Identified Through MS: Conclusions*), and existing studies showing that random errors within proteins are strong contributors to misfolding [107][108]. If translational error products were allowed to mature only to misfold and degrade they would simply represent an additional energetic cost to the cell, without altering observable protein levels.

As such it is quite possible that both the knockout and the drug based experiments would be unable to observe effects upon translational error. Accordingly, failure of knockout mutants and degradation drug inhibition experiments to detect any clear effects does not rule out there role in translational error regulation. Supporting this possibility, the role of protease function in enabling cell growth in the presence of high levels of translational errors has been observed in previous literature [7], further indicating that the current techniques may have simply lacked the correct sensitivity.

An alternative, *in vitro*, approach to monitoring the levels of translational errors for differing systems would be through an enzymatic assay of the produced and purified enzyme, assuming that lower error samples will show higher activities. This approach would, depending upon the assay used, allow detection of small differences between enzymatic efficiencies, however it would be unable to take into account translational errors which were unable to fold stably.

A more complex, *in vitro*, approach could focus on the cellular effects of translational errors as an indirect indicator. Previous studies have proposed that the pool of non-functioning protein and needless ribosome sequestration represent a growth burden upon the cell, and thus evolutionary selective pressure promotes translational accuracy in high expression genes [39][41]. As it may be argued that over-expressed GST is no more useful to the host cell than a misfolded protein, there may be little difference between production of a functional and non-functional product in this experiment due to the equally futile ribosome sequestration. In future experiments, a minimum-optimised gene encoding a protein required for growth (such as URA3 in Ura- media) could be useful to study, whereby correct folding increases cell growth, and misfolding places a burden upon the cell. Differing production levels for optimised and non-optimised codon sequences would make comparison of differing transformants difficult; however analysis of the effects of differing growth conditions would be entirely possible. For example if a minimum optimisation URA3 culture experienced growth reduction, while control cells did not, it would indicate effects upon translational error management. Another approach to this problem could be to produce the target protein under regulation so that only a suitable level of protein will be produced, minimising the effect of differing production rates, and placing additional burden upon cells producing multiple errors (as non-stable errors consume cell recourses, and stable errors decrease the effectiveness of the required protein pool). Under these conditions it may be possible to detect the effect of deactivating translational error related pathways through varying growth rates and cell health, as has been previously seen for protease deficiency in tRNA misacylating cells [7].

Aside from this, high MW bands were also examined in detail through drug assays and immunoprecipitation. While not entirely conclusive, the performed IP experiment suggests that the higher MW bands are not ubiquitinated-GST species, and there is currently no

compelling evidence to suggest relation to GST errors. While these bands appear upon anti-GST blots, they are likely cross reactions, and possibly a GST dimer molecule, reacting to the drug treatments according to their own regulatory systems. As such there may be little use in further studying their behaviour.

4.2. Computational Analysis of Observed Translational Errors

4.2.1. Translational Error Binding Affinities and Novel Gain of Function

Mass spectrometry analysis revealed a range of translational errors within the Min-GST product. Through analysis of binding to glutathione magnetic beads, and observing the depletion of specific misincorporation products, the relative binding affinities of varying translational mutants could be estimated, and seen in Figure 3.8.1 to include affinities both higher and lower than the WT. It may be noted that due to the difference between binding affinity and enzymatic activity, both increases and decreases in binding affinity likely represent decreases in enzymatic activity, as further discussed in *4.3.1. Stability and Production of Translational Mutants*.

One translational error, N152C showed a highly anomalous result, appearing to give a binding affinity several orders of magnitude lower than other observed substitutions. Computational analysis found this substitution to place a Cys residue upon the surface within a slight crevice, as shown in Figure 3.8.2, where disulphide homodimer formation would be prevented by steric effects, but direct interaction with GST itself would be possible. This matches the observed readings, where alternate binding to the magnetic beads appears to prevent elution and result in an apparent depletion of the GST mutant.

This result demonstrates possible novel function emerging through translational error, altering the activity of the original protein. Through error, the protein appears to bind the substrate alternately in a site distinct to the active site. While this is a relatively minor gained function, its existence raises the possibility that error products in other proteins may result in other gained functions. It is plausible, for example, that errors within the active sites or regulatory regions of other proteins could result in altered function, essentially producing a minor product of translation. The general concept of programmed error to produce a minor product is not unprecedented. Many examples of less specific errors exist, such as increased error rates in bacteria conferring antibiotic resistance through reduced drug target susceptibility [146], or mismethionylation in many species to protect against oxidative damage through the introduction of additional methionine groups [17][147][148]. Errors have also been known to produce specific protein products, such as common stop codon readthrough in drosophila to access further coding regions [149]. In light of already well studied uses of translational errors, it is not implausible that cells may take advantage of missense errors of this kind, perhaps through the use of low accuracy codons, in order to produce minor products with useful functions. Detection of such an error, however, would be difficult, given the wide range of substitution possibilities within cells, genetic drift's effect upon exact codon choices, and the difficulty in predicting protein function. As such, this function is likely to remain purely speculative.

4.2.2. Direct Influence of Substitution Upon Binding

Distances from the site of substitution to the bound ligand within the tertiary structure were seen in 3.9.2 to follow, with p = 0.0117, an inverse square relationship of:

$$Disruption\ of\ Binding \propto \frac{1}{Distance^2}$$

This is a common relationship for radiating influence in a 3D sphere, and here describes the influence of substitution effects such as charge differences and steric alterations, demonstrating that the disruption of binding is related to the position of the translational error, and as such is wholly dependent upon the substitution itself. In other words, translational error proteins are affected by the substitution only, and appear to contain no other disruptive features resulting from their flawed production.

Differences in binding affinities between surface and internal substitutions were also examined, however as this distinction is more likely to affect protein stability, a parameter distinct to binding affinity, no significant change was observed, as seen in Figure 3.10. Additionally, as strongly destabilising mutations will be removed due to failure to fold correctly, both surface and internal substitutions are "pre-selected" for moderate stability, [7][107][108] likely further reducing differences between these two groups.

4.2.3. Influences of Evolutionary Variability and Favourability of Substitution

Plotting of residue variability against reduction of binding affinity, shown in Figure 3.11.1, shows a distinctly non-linear relationship. This is likely the result of two competing factors. Firstly, substitutions in high variability regions are likely to be accommodated with little effect upon binding affinity. Secondly, many substitutions in low variability regions are likely to affect protein folding and stability, leading to misfolding and degradation through the heat shock pathway [7][107][108], and as such only particularly benign substitutions within these regions will be observable. Due the complexity of this relationship, a clear, biologically relevant, regression equation cannot be produced based on current data.

BLOSUM62 scores for observed substitutions were used as an indicator of likely impact (with less evolutionarily favourable substitutions expected to have a more detrimental effect upon the protein) [134]. BLOSUM62 scores of substitution in Figure 3.11.2 & 3.11.3 were not observed to correlate significantly with either the observed fold change, or the variability of the original residue. This lack of correlation is explained by two main factors. Firstly, BLOSUM62 score usage introduces a high level of simplification, whereby local effects and structural features are not taken into account, and as such the level of correlation would inevitably be weak, even with a perfect data set. Secondly, the translational error set observed is pre-selected for stability, as discussed in the above paragraph. Largely destabilising errors will not be observed, leaving only relatively stable examples of all substitutions, preventing BLOSUM62 scores having a significant effect upon the system.

4.2.4. Disconnect Between tRNA Abundance Ratios and Likelihood of Error

The non-optimised codon sequence of Min-GST suffers from high levels of errors due to aatRNA misincorporation. It is known that higher abundance aatRNAs are more likely to replace lower abundance aatRNAs (see *1.2.2 Missense Errors*), governing the frequency of these events. As such, the tRNA ratio of substitution can be examined (Abundance of new tRNA / Abundance of original tRNA), noting that aatRNA ratio is reliably represented by tRNA ratio (see *1.2.2.2 Missense Errors - tRNA level maintenance*).

It should also be noted that aatRNA misincorporation is not the only source of substitution errors, with misacylation errors also producing substitutions (see *1.2.4 tRNA Misacylation*) unrelated to tRNA abundances. Additionally, only near-cognate aatRNAs may incorrectly insert into the ribosome, limiting which substitutions are possible (see *1.2.2 Missense Errors*).

The tRNA ratio of substitution was used instead of simple tRNA abundances, as such ratios are known to carry more biological relevance. For example, an abundant tRNA will not necessarily show low errors if it is accompanied by several equally abundant near cognate tRNAs. Through examination of substitution ratios, this analysis avoids these complications [112].

Importantly, all examinations of this tRNA ratio, seen in Figure 3.12.1 - 3.12.3, show ratios of above and below 1; i.e. translational errors are observed for both favourable and non-favourable tRNA ratios. While the frequency of substitution at each position is known to be dependent upon tRNA ratio, all substitutions will occur to some extent. As a result, detection of substitutions in mature proteins will be strongly influenced by the stability of these substitutions, in addition to their likelihood of occurrence. This is demonstrated in Figure 3.12.3, where the probability of detecting a substitution for each type of amino acid is seen to be independent of the associated tRNA ratio.

Due to the nature of the mass spectrometry experiment used, relative abundance of error products cannot be commented on, giving a potential focus for future experiments.

A strong correlation between tRNA ratios and the depletion of binding affinity is observed in Figure 3.12.2. This correlation suggests that substitution of a lower abundance tRNA with a higher abundance tRNA results in greater reduction of binding affinity. As these substitution ratios are the most common *in vivo* (see *1.2.2 Missense Errors*), this observation carries implications for cell function. As cells attempt to optimise codon usage to match high abundance tRNAs, codons corresponding to low abundance tRNAs will represent low abundance amino acids within the cell, likely maintained within the protein due to difficulty in replacing the amino acid with a more common equivalent [107]. A substitution of one of these rarer amino acids is likely to result in replacement of a key residue, and thus lead to greater destabilisation of the protein.

Finally, binding affinity is compared to the probability of detecting a substitution for each type of amino acid. The proportion of each amino acid seen to be substituted shows no

relationship to fold change, however the proportion of each amino acid to follow a particular substitution, such as Y to M, does show negative correlation with fold change. It is likely that more rarely observed substitutions are more structurally unfavourable, and as such when these substitutions are observed they are strongly selected to have minimal effect upon the protein, and therefore have less impact upon the highly conserved active site.

4.2.5. Relationship Between Evolutionary Variability and Translational Error Detection

Multiple sequence alignment based analysis of the YxM substitution revealed the relationship between evolutionary variation and translational error observation, with the observed error sites varying more greatly than other positions, as shown in Figure 3.13.1.

Variation is a measure of the frequency at which residues are replaced between homologues, i.e. a high variation residue will commonly change between homologues, whereas a low variation residue is more likely to be conserved. Lower variation residues are often important for maintenance of the active site, or general protein folding. As such, the substitution of these low variation positions is seen to be evolutionarily unfavourable, matching the behaviour of YxM translational errors.

Further analysis found the relationship between evolutionary variation and the likelihood of observing translational errors to be maintained for all positions detected through the MS assay, as shown in Figure 3.13.2. It is hypothesised that evolutionary variation and translational error observation show this relationship as substitutions negatively affecting protein stability will be selected against in evolution, and error proteins containing these substitutions are likely to be rapidly degraded *in vivo*. This strongly indicates that the detection of translational errors is based upon folding selection, whereby substitutions in less variable regions are unlikely to be tolerated, preventing protein folding and leading to rapid degradation.

It may be noted that this relationship contains very high standard deviations. The average evolutionary variance scores of for observed and non-observed substitutions are 0.310 and -0.207, with respective standard deviations of 1.05 and 0.95. As such, there is a high level of overlap between these two groups with many non-observed substitution positions showing higher evolutionary variation than their observed equivalents, as seen for specific amino acids in Figure 3.13.4. This means that while there is a clear relationship with variation, it cannot be used to predict the locations of translational error observation.

This high level of deviation exists due to the simplifications involved in using variance for this analysis. Factors which are not accounted for include the limitation of translational errors to near-cognate tRNAs, the possibility that many evolutionary variations may be linked, the role of amino acid conservation in maintaining activity instead of stability, and the complexity of differing point mutations. This level of simplification also explains why, generally, there is no increased likelihood for observed error substitutions to match evolutionary variants.

4.2.6. Computational Analysis of Observed Translational Errors: Conclusions

Computational analysis of mass spectrometry data proved to be a useful addition to this project, resulting in several key conclusions.

Firstly, clear relationship between the position of substitution and resultant effect on binding affinity was established, indicating that translational errors effect proteins though the substitution only, and that the proteins are otherwise unaffected by the error in their production.

Secondly, residue variability appears to hold a complex relationship to the effect of substitutions upon binding affinity, whereby substitutions upon medium variability residues are more likely to deplete binding affinity, supporting two mechanisms expected to effect the system; i.e. substitutions in high variability regions are likely to be accommodated without affecting binding, and substitutions in low variability regions are likely to disrupt protein folding causing only a filtered "non-disrupting" set to be observable in the mature protein. This model is supported by observations that substitutions are more likely to be observed in high variability regions, indicating that substitutions in low variability regions are indeed more likely to prevent folding and undergo degradation, demonstrating selection in these substitutions.

Thirdly, due to this effect of protein stability on allowing detection of translational errors, as further discussed in *4.3.1. Stability and Production of Translational Mutants*, tRNA abundance was not related to the likelihood of translational error substitutions being detected in any apparent manner. While it is quite likely that substitutions with a favourable tRNA balance were more abundant than those with an unfavourable balance, this was not detectable through the techniques used in this project.

Fourthly, in the context of detectable translational errors within the mature protein, the only observed effect of the tRNA ratio appeared to be in disruption of binding affinity, whereby the tRNA ratio acts as an indirect indicator of amino acid rarity: A codon corresponding to a lower abundance tRNA represents a rare amino acid within the cell which has not been replaced through sequence optimisation, replacement of which is more damaging to protein activity.

4.2.7. Computational Analysis of Observed Translational Errors: Reliability and Limitations

While computational analysis must considered critically, the major findings here, that translational errors in protein synthesis only affect the mature protein through the amino acid substitution itself, and that stability plays a key role in allowing observation of translational errors in mature proteins, are both supported by experimental data (see *4.3.3. Analysis of Translation Mutants Initially Identified Through MS: Conclusions*).

Further findings, such as the complex relationship between the evolutionary variability of substituted residues and the effect on protein function, and the non-casual correlation between tRNA substitution ratios and the effect on protein function, represent more minor details of the system. These relationships would benefit from examination of additional

samples, preferably of non-structurally related proteins, and perhaps in alternate cell types, in order to confirm both their reliability and generality.

4.3. Analysis of Translational Errors Initially Identified Through MS

4.3.1. Stability and Production of Translational Mutants

In order to examine why only certain positions showed detectable translational error substitutions, the mutants Y15M (non-observed) and Y172M (observed as translational error) were generated and tested alongside further observed mutants; V156E, A180Q, and P210T.

Initial computational analysis suggested Y15M to be a highly unfavourable mutant, found within a high conservation area, and seen to potentially introduce a helix dipole within the tertiary structure; indicating worse stability than all other mutants tested, as discussed in Table 3.2.

This prediction was tested in Figure 3.14. As mentioned previously, determining abundance by western blot does not provide a direct measure of stability, but also takes into account factors such as production, turnover, and heat shock protein response [7]. This measure of cellular abundance is more relevant to biological function than pure stability measurements, and is an indicator of the actual viability of mutants within growing cells. It may be noted, however, that western blotting is generally matched by coomassie blue staining, indicating that antibody binding does not greatly affect results. All observed mutants are around tenfold less abundant than the WT, in keeping with previous literature stating that random substitutions will negatively affect protein stability [38][107][108]. Notably, this abundance is lower than that observed for the Min-GST construct, previously seen in Figure 3.1, suggesting that random amino acid substitution is more detrimental than poor codon optimisation, in keeping with previous literature claiming amino acid choice and codon optimisation to represent competing systems of similar importance for misfolding avoidance [108].

The non-observed mutant, meanwhile, shows a significantly lower abundance than the other mutants in Figure 3.14 with Y15M at around half the abundance of Y172M. Given the already low abundances of the GST mutants (~1/10 of WT), this represents a very low abundance (~1/20 of WT). This, combined with the computational prediction, is a strong indicator that Y15M was not observed due to the lower stability compared to other translational error mutants, and suggests a major role for stability in translational error protein maturation.

4.3.2. Activity and Normality of Translational Error Mutants

Substitutions produced through translational error and DNA mutation should result in identical effects upon the mature protein. Given that binding affinities for translational errors were gathered in the original MS experiment, it was possible to compare these to binding affinities for the produced DNA mutants, as well as examining the effect of individual substitutions in more detail.

The binding affinities of GST-mutants, indicated through the binding of glutathione magnetic beads, are seen in Figure 1.15.1 to vary strongly and independently of the abundances seen previously in Figure 3.14.1. It is notable that the apparently random mutations result in both higher and lower binding affinities than the wild type, however it must be remembered that binding affinity is not a measure of enzyme activity, and must be maintained at the correct balance for efficiency to be achieved. Here, mutants Y15M, V156E, and P210T show lower binding affinities suggesting alteration of the active site and a reduced ability to interact with the substrate. Mutants Y172M and A180Q show higher binding affinities, perhaps due to subtle alterations of active site dynamics leading to stronger substrate binding. As the balance between substrate binding and transition state binding is highly important for catalysis, it is likely that this alteration would negatively affect GST activity [150][151]. As such, all tested translational errors are seen to negatively impact activity.

The non-observed mutant, Y15M does not appear remarkable in Figure 3.15.1, indicating that alteration of activity has no effect on the likelihood of observing a translational error.

Comparison of fold changes observed in the MS experiment for translational errors, and the gel based experiment for DNA mutants, should not expect perfect correlation due to experimental differences between the two procedures. It was seen in Figure 3.15.2 that the general effect, of either increasing or decreasing binding relative to the WT, was maintained between the two experiments, indicating both sources of substitution to be functionally identical. As such, the introduction of a translational error effects the developing protein through the substitution only.

4.3.3. Analysis of Translation Mutants Initially Identified Through MS: Conclusions

Three main observations may be taken from these experiments.

Firstly they show the non-observed substitution, Y15M, to be produced at lower abundance *in vivo* than comparable observed substitutions, indicating that differences in stability and turnover determine which translational errors are detectable.

Secondly they demonstrate substitutions produced through both translational error and DNA mutation to behave similarly, showing these two processes to produce functionally identical proteins.

Thirdly they confirm that binding affinity and protein stability are separately affected by translational errors, as would be expected for these two separate parameters.

4.3.4. Analysis of Translation Mutants Initially Identified Through MS: Reliability and Limitations

These conclusions are consistent with computational analysis (see *4.2.6. Computational Analysis of Observed Translational Errors: Conclusions*), give clear results, and are consistent with theory; suggesting a strong degree of reliability. It may be noted, however, that only one protein model has been used (GST), only five total mutants were produced, and only one non-observed translational mutant was compared. Future work could focus upon the

creation of a greater range of mutants in order to increase significance, allow more reliable comparison with computational data, and give more confidence to the differences between observed and non-observed translational mutants. Beyond this, work upon a protein structurally unrelated to GST would also be useful in proving the universal relevance of this data.

4.4. Overall Conclusions

This study has examined in detail the ability of western blotting techniques to determine the influence of various pathways upon translational error regulation, using heavily optimised and non-optimised versions of the GST gene in *S. cerevisiae*. Through examination of product levels, possible higher MW derivatives, drug treatments, and knockout mutants, no clear influences were found. In light of these results, either none of the pathways examined have a significant role in translational errors, or the western blotting technique is not capable of detecting alterations.

Alternate techniques were proposed for examining possible mechanisms of translational error regulation. These techniques would examine the activity of the produced protein products, either *in vitro* or *in vivo*, both with relative advantages and disadvantages.

Mass spectrometry data for translational errors and their relative effects on binding affinities was analysed through both follow up experimentation and computational analysis. These complimentary approaches allowed conclusions to be drawn with a confidence which would not have otherwise been possible, based on the limited data sets available.

Firstly, the importance of protein folding in the observation of translational errors was highlighted, as shown in Figure 4.1. This is evidenced in computational analysis firstly by the tendency for substitutions to be observed upon high variability residues, indicating a role for stability in error observation, and secondly by the lack of relationship between the chances of a substitution being observed and the tRNA abundances involved, showing a disconnect between error rates at the translational stage and their observation in the mature protein. Experimental analysis based upon recreating translational errors through DNA mutation supports this, revealing lower abundance of a non-observed substitution in comparison to observed equivalents.

This indicates folding to play an important role in translational error regulation, whereby errors occurring despite proofreading systems often form non-viable protein mutants, which are readily degraded *in vivo*, and thus not detected. This is proposed as a last line error correction system, due to the high energetic costs involved for the cell; however previous literature shows that protease deficiency reduces cells' ability to grow in presence of translational errors [7], suggesting a more general importance of this system.

This process is often overlooked, as approaches commonly used to examine translational error often rely on simulations [21][107][108], or specific reporter systems. For example, the luciferase system has been widely used for error analysis, whereby specific errors either allow luciferase production [152][153], or restore active site functionality through missense

suppression [18][26][154][155][156]. Similarly, techniques have even been developed using MS in order to examine a single residue for misincorporation upon a stable elastin based scaffold [157]. These studies, while useful, only provide information on an error's occurrence at the translational level, and do not take into account the selective processes found at the folding stage, as seen in Figure 4.1.

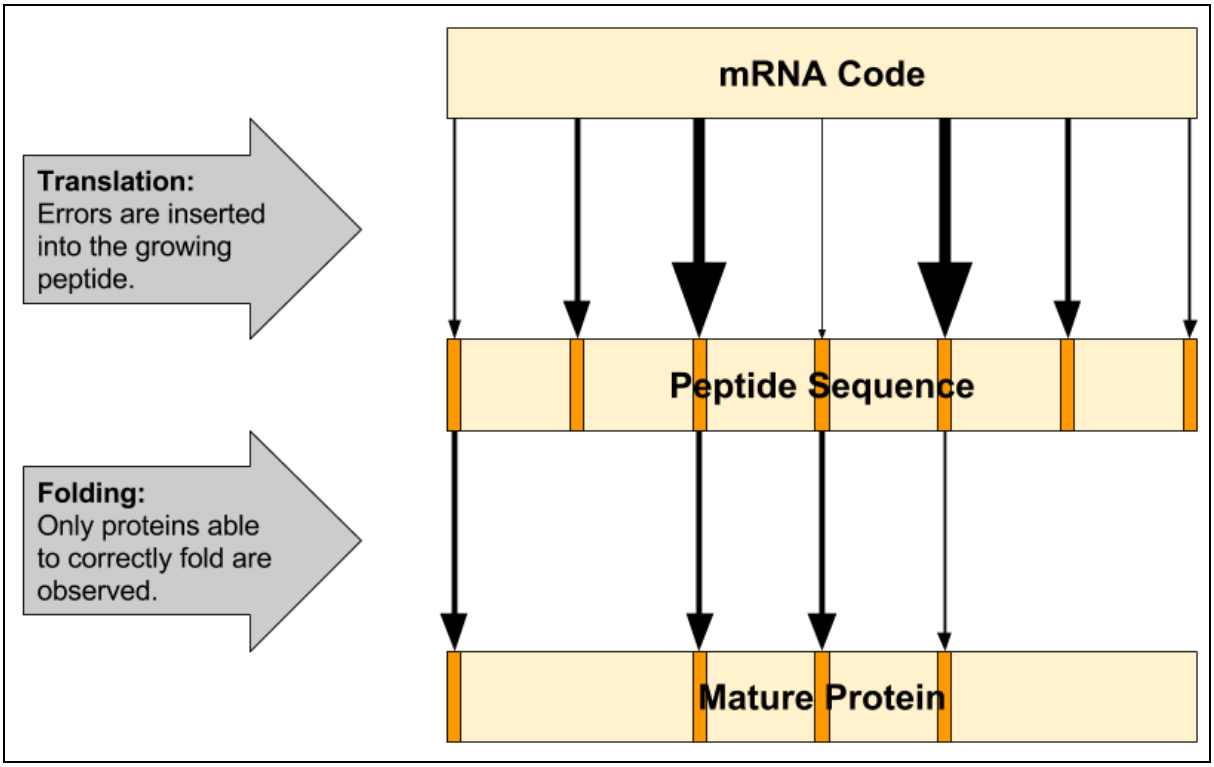


Figure 4.1: Translational errors through the production of a protein. Translational errors are produced in translation as incorrect amino acids are inserted into the growing peptide chain. Many of these errors create unstable proteins which fail to fold and thus face degradation. The resultant level of translational error to be found within mature proteins is therefore dependent upon both rate of occurrence and stability. ***(Orange = Errors. Arrow size indicates abundance. Quantities given are for illustrative purpose only.)***

Secondly, it is seen that protein production is otherwise unaffected by the introduction of a substitution error, and the resultant protein differs from the WT form only through amino acid change. This was evidenced computationally through a clear relationship between substitution positions and their effect upon binding affinity in MS analysis, indicating local effects only, and experimentally through the ability to recreate the affinities of translational error proteins through the production of matching DNA mutants.

Finally, the high level of agreement between these two approaches validates the use of this particular MS assay, and associated computational analysis, for the further study of translational error systems, both in this study and beyond.

5. References

1 Crick, F. (1970) Central dogma of molecular biology. Nature **227**, 561–3.

2 Ling, J., Reynolds, N. and Ibba, M. (2009) Aminoacyl-tRNA Synthesis and Translational Quality Control. Annu. Rev. Microbiol. **63**, 61–78.

3 Jakubowski, H. Quality control in tRNA charging. (2012) Wiley Interdiscip. Rev. RNA **3**, 295–310.

4 Mascarenhas, A. P., An, S., Rosen, A. E., Martinis, S. A. and Musier-Forsyth, K. (2009) Fidelity Mechanisms of the Aminoacyl-tRNA Synthetases. In Protein Engineering, 155–203.

5 Rodnina, M. V and Wintermeyer, W. (2009) Recent mechanistic insights into eukaryotic ribosomes. Curr. Opin. Cell Biol. **21**, 435–443.

6 Fluitt, A., Pienaar, E. and Viljoen, H. (2007) Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. Comput. Biol. Chem. **31**, 335–346.

7 Ruan, B., Palioura, S., Sabina, J., Marvin-Guy, L., Kochhar, S., Larossa, R. A. and Söll, D. (2008) Quality control despite mistranslation caused by an ambiguous genetic code. Proc. Natl. Acad. Sci. U. S. A. **105**, 16502–7.

8 Murugan, A., Huse, D. A. and Leibler, S. (2012) Speed, dissipation, and error in kinetic proofreading. Proc. Natl. Acad. Sci. U. S. A. **109**, 12034–9.

9 Hill, D. E. and Struhl, K. (1986) A rapid method for determining tRNA charging levels in vivo: analysis of yeast mutants defective in the general control of amino acid biosynthesis. Nucleic Acids Res. **14**, 10045–51.

10 Varlamova, E. G., Gol'tiaev, M. V, Novocelov, S. V, Novoselov, V. I. and Fecenko, E. E. (2013) Biosyinthesis and mechanism of selenocysteine incorporation into synthesized proteins. Mol. Biol. (Mosk). **47**, 558–67.

11 Krzycki, J. A. (2013) The path of lysine to pyrrolysine. Curr. Opin. Chem. Biol. **17**, 619–25.

12 Goodenbour, J. M. and Pan, T. (2006) Diversity of tRNA genes in eukaryotes. Nucleic Acids Res. **34**, 6137–46.

13 Nangle, L. A., Motta, C. M. and Schimmel, P. (2006) Global Effects of Mistranslation from an Editing Defect in Mammalian Cells. Chem. Biol. **13**, 1091–1100.

14 Lee, J. W., Beebe, K., Nangle, L. A., Jang, J., Longo-Guess, C. M., Cook, S. A., Davisson, M. T., Sundberg, J. P., Schimmel, P. and Ackerman, S. L. (2006) Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature **443**, 50–55.

15 Yadavalli, S. S. and Ibba, M. (2013) Selection of tRNA charging quality control mechanisms that increase mistranslation of the genetic code. Nucleic Acids Res. **41**, 1104–12.

16 Bezerra, A. R., Simões, J., Lee, W., Rung, J., Weil, T., Gut, I. G., Gut, M., Bayés, M., Rizzetto, L., Cavalieri, D., et al. (2013) Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. Proc. Natl. Acad. Sci. U. S. A. **110**, 11079–84.

17 Wiltrout, E., Goodenbour, J. M., Fréchin, M. and Pan, T. (2012) Misacylation of tRNA with methionine in Saccharomyces cerevisiae. Nucleic Acids Res. **40**, 10494–506.

18 Kramer, E. B. and Farabaugh, P. J. (2007) The frequency of translational misreading errors in E. coli is largely determined by tRNA competition. RNA **13**, 87–96.

19 Farabaugh, P. J., Björk, G. R., Atkins, J., Weiss, R., Thompson, S., Gesteland, R., Barak, Z., Lindsley, D., Gallant, J., Belcourt, M., et al. (1999) How translational accuracy influences reading frame maintenance. EMBO J. **18**, 1427–34.

20 Tinoco, I., Kim, H.-K., Yan, S. and Yan, S. (2013) Frameshifting dynamics. Biopolymers **99**, 1147–66.

21 Wu, G. and Yan, S. (2005) Determination of mutation trend in proteins by means of translation probability between RNA codes and mutated amino acids. Biochem. Biophys. Res. Commun. **337**, 692–700.

22 Varenne, S., Buc, J., Lloubes, R. and Lazdunski, C. (1984) Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Mol. Biol. **180**, 549–76.

23 Heyd, A. and Drew, D. A. (2003) A mathematical model for elongation of a peptide chain. Bull. Math. Biol. **65**, 1095–1109.

24 Tarrant, D. and von der Haar, T. (2014) Synonymous codons, ribosome speed, and eukaryotic gene expression regulation. Cell. Mol. Life Sci. **71**, 4195–4206.

25 Zouridis, H. and Hatzimanikatis, V. (2008) Effects of Codon Distributions and tRNA Competition on Protein Translation. Biophys. J. **95**, 1018–1033.

26 Plant, E. P., Nguyen, P., Russ, J. R., Pittman, Y. R., Nguyen, T., Quesinberry, J. T., Kinzy, T. G. and Dinman, J. D. (2007) Differentiating between Near- and Non-Cognate Codons in Saccharomyces cerevisiae. PLoS One (Lewin, A., ed.) **2**, e517.

27 Satpati, P., Sund, J. and Åqvist, J. (2014) Structure-Based Energetics of mRNA Decoding on the Ribosome. Biochemistry **53**, 1714–1722.

28 Menninger, J. R. (1983) Computer simulation of ribosome editing. J. Mol. Biol. **171**, 383–399.

29 Menninger, J. R., Coleman, R. A. and Tsai, L.-N. (1994) Erythromycin, lincosamides, peptidyl-tRNA dissociation, and ribosome editing. Mol. Gen. Genet. MGG **243**, 225–233.

30 Belcourt, M. F. and Farabaugh, P. J. (1990) Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. Cell **62**, 339–52.

31 Ikemura, T. (1982) Correlation between the abundance of yeast transfer RNAs and the

occurrence of the respective codons in protein genes: differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer R. J. Mol. Biol. **158**, 573–597.

32 Kanaya, S., Yamada, Y., Kudo, Y. and Ikemura, T. (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene **238**, 143–155.

33 Pooja K., S., Stanislav G., K., Fred S., D., John H., M., Daniel A., S. and Paul M., M. (2015) 2μ plasmid in Saccharomyces species and in Saccharomyces cerevisiae. FEMS Yeast Res. **15**.

34 Dittmar, K. A., Goodenbour, J. M. and Pan, T. (2006) Tissue-specific differences in human transfer RNA expression. PLoS Genet. **2**, 2107–2115.

35 Novoa, E. M., Pavon-Eternod, M., Pan, T., Ribas de Pouplana, L., Agris, P. F., Vendeix, F. A., Graham, W. D., Akashi, H., Akashi, H., Bennetzen, J. L., et al. (2012) A role for tRNA modifications in genome structure and codon usage. Cell **149**, 202–13.

36 Duret, L. and Mouchiroud, D. (1999) Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **96**, 4482–7.

37 Grosjean, H., de Crécy-Lagard, V. and Marck, C. (2010) Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes. FEBS Lett. **584**, 252–264.

38 Homma, K., Noguchi, T. and Fukuchi, S. (2016) Codon usage is less optimized in eukaryotic gene segments encoding intrinsically disordered regions than in those encoding structural domains. Nucleic Acids Res. **44**, gkw899.

39 McDonald, M. J., Chou, C.-H., Swamy, K. B. S., Huang, H.-D. and Leu, J.-Y. (2015) The evolutionary dynamics of tRNA-gene copy number and codon-use in E. coli. BMC Evol. Biol. **15**, 163.

40 Carbone, A., Zinovyev, A. and Képès, F. (2003) Codon adaptation index as a measure of dominating codon bias. Bioinformatics **19**, 2005–15.

41 Kudla, G., Murray, A. W., Tollervey, D. and Plotkin, J. B. (2009) Coding-sequence determinants of gene expression in Escherichia coli. Science **324**, 255–8.

42 Tuller, T., Girshovich, Y., Sella, Y., Kreimer, A., Freilich, S., Kupiec, M., Gophna, U. and Ruppin, E. (2011) Association between translation efficiency and horizontal gene transfer within microbial communities. Nucleic Acids Res. **39**, 4743–55.

43 Retchless, A. C. and Lawrence, J. G. (2012) Ecological adaptation in bacteria: speciation driven by codon selection. Mol. Biol. Evol. **29**, 3669–83.

44 Withers, M., Wernisch, L. and dos Reis, M. (2006) Archaeology and evolution of transfer RNA genes in the Escherichia coli genome. RNA **12**, 933–42.

45 Bloom-Ackermann, Z., Navon, S., Gingold, H., Towers, R., Pilpel, Y., Dahan, O.,

Kozak, M., Jackson, R., Hellen, C., Pestova, T., et al. (2014) A Comprehensive tRNA Deletion Library Unravels the Genetic Architecture of the tRNA Pool. PLoS Genet. (Copenhaver, G. P., ed.) **10**, e1004084.

46 Dong, H., Nilsson, L. and Kurland, C. G. (1996) Co-variation of tRNA Abundance and Codon Usage inEscherichia coliat Different Growth Rates. J. Mol. Biol. **260**, 649–663.

47 Candelas, G. C., Arroyo, G., Carrasco, C. and Dompenciel, R. (1990) Spider silkglands contain a tissue-specific alanine tRNA that accumulates in vitro in response to the stimulus for silk protein synthesis. Dev. Biol. **140**, 215-20.

48 Cintrón, I., Capó, L., Plazaola, A., Arroyo, G. and Candelas, G. C. (1999) A spider tRNAAla requires a far upstream sequence element for expression. Gene **231**, 195–201.

49 Sullivan, H. S., Young, L. S., White, C. N. and Sprague, K. U. (1994) Silk gland-specific tRNA(Ala) genes interact more weakly than constitutive tRNA(Ala) genes with silkworm TFIIIB and polymerase III fractions. Mol. Cell. Biol. **14**, 1806–14.

50 Underwood, D. C., Knickerbocker, H., Gardner, G., Condliffe, D. P. and Sprague, K. U. (1988) Silk gland-specific tRNA(Ala) genes are tightly clustered in the silkworm genome. Mol. Cell. Biol. **8**, 5504–12.

51 Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M. and Pagano, A. (2007) The expanding RNA polymerase III transcriptome. Trends Genet. **23**, 614–622.

52 Taneja, R., Gopalkrishnan, R. and Gopinathan, K. P. (1992) Regulation of glycine tRNA gene expression in the posterior silk glands of the silkworm Bombyx mori. Proc. Natl. Acad. Sci. U. S. A. **89**, 1070–4.

53 Wichtowska, D., Turowski, T. W. and Boguta, M. (2013) An interplay between transcription, processing, and degradation determines tRNA levels in yeast. Wiley Interdiscip. Rev. RNA **4**, 709-22.

54 Dreher, T. W., Uhlenbeck, O. C. and Browning, K. S. (1999) Quantitative assessment of EF-1alpha.GTP binding to aminoacyl-tRNAs, aminoacyl-viral RNA, and tRNA shows close correspondence to the RNA binding properties of EF-Tu. J. Biol. Chem. **274**, 666–72.

55 Parker, R. (2012) RNA Degradation in Saccharomyces cerevisae. Genetics **191**, 671-702.

56 Thompson, D. M., Lu, C., Green, P. J. and Parker, R. (2008) tRNA cleavage is a conserved response to oxidative stress in eukaryotes. RNA **14**, 2095–103.

57 Yamasaki, S., Ivanov, P., Hu, G. and Anderson, P. (2009) Angiogenin cleaves tRNA and promotes stress-induced translational repression. J. Cell Biol. **185**, 35-42.

58 Ivanov, P., Emara, M. M., Villen, J., Gygi, S. P. and Anderson, P. (2011) Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. Mol. Cell **43**, 613–623.

59 Chu, D., Barnes, D. J. and von der Haar, T. (2011) The role of tRNA and ribosome competition in coupling the expression of different mRNAs in Saccharomyces cerevisiae. Nucleic Acids Res. **39**, 6705–14.

60 Chan, C. T. Y., Pang, Y. L. J., Deng, W., Babu, I. R., Dyavaiah, M., Begley, T. J., Dedon, P. C., Rozenski, J., Crain, P. F., McCloskey, J. A., et al. (2012) Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat. Commun. **3**, 937.

61 Yarian, C., Townsend, H., Czestkowski, W., Sochacka, E., Malkiewicz, A. J., Guenther, R., Miskiewicz, A. and Agris, P. F. (2002) Accurate translation of the genetic code depends on tRNA modified nucleosides. J. Biol. Chem. **277**, 16391–5.

62 Urbonavicius, J., Qian, Q., Durand, J. M., Hagervall, T. G. and Björk, G. R. (2001) Improvement of reading frame maintenance is a common function for several tRNA modifications. EMBO J. **20**, 4863–73.

63 Tats, A., Tenson, T., Remm, M., Irwin, B., Heck, J., Hatfield, G., Murgola, E., Pagel, F., Hijazi, K., Bossi, L., et al. (2008) Preferred and avoided codon pairs in three domains of life. BMC Genomics **9**, 463.

64 Shpaer, E. G. (1986) Constraints on codon context in Escherichia coli genes their possible role in modulating the efficiency of translation. J. Mol. Biol. **188**, 555–564.

65 Murgola, E. J., Pagel, F. T. and Hijazi, K. A. (1984) Codon context effects in missense suppression. J. Mol. Biol. **175**, 19–27.

66 Precup, J., Ulrich, A. K., Roopnarine, O. and Parker, J. (1989) Context specific misreading of phenylalanine codons. MGG Mol. Gen. Genet. **218**, 397–401.

67 Carrier, M. J. and Buckingham, R. H. (1984) An effect of codon context on the mistranslation of UGU codons in vitro. J. Mol. Biol. **175**, 29–38.

68 Buchan, J. R., Aucott, L. S. and Stansfield, I. (2006) tRNA properties help shape codon pair preferences in open reading frames. Nucleic Acids Res. **34**, 1015–27.

69 Roche, E. D., Sauer, R. T., Bonekamp, F., Jensen, K., Breyer, R., Sauer, R., Calderone, T., Stevens, R., Oas, T., Chen, G., et al. (1999) SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. EMBO J. **18**, 4579–89.

70 Wagner, L. A., Weiss, R. B., Driscoll, R., Dunn, D. S. and Gesteland, R. F. (1990) Transcriptional slippage occurs during elongation at runs of adenine or thymine in Escherichia coli. Nucleic Acids Res. **18**, 3529–35.

71 Baranov, P. V, Hammer, A. W., Zhou, J., Gesteland, R. F. and Atkins, J. F. (2005) Transcriptional slippage in bacteria: distribution in sequenced genomes and utilization in IS element gene expression. Genome Biol. **6**, R25.

72 Gurvich, O. L., Baranov, P. V, Zhou, J., Hammer, A. W., Gesteland, R. F. and Atkins, J. F. (2003) Sequences that direct significant levels of frameshifting are frequent in coding regions of Escherichia coli. EMBO J. **22**, 5941–50.

73 Irwin, B., Heck, J. D. and Hatfield, G. W. (1995) Codon Pair Utilization Biases Influence Translational Elongation Step Times. J. Biol. Chem. **270**, 22801–22806.

74 Jakubowski, H. (2006) Energy Cost of Translational Proofreading in Vivo. Ann. N. Y. Acad. Sci. **745**, 4–20.

75 Korencic, D., Ahel, I., Schelert, J., Sacher, M., Ruan, B., Stathopoulos, C., Blum, P., Ibba, M. and Söll, D. (2004) A freestanding proofreading domain is required for protein synthesis quality control in Archaea. Proc. Natl. Acad. Sci. U. S. A. **101**, 10260–5.

76 Soll, D. (2004) Aminoacyl-tRNA tRNA formation: An essential function in protein synthesis and its quality control. Nucleic Acids Symp. Ser. **48**, 283–284.

77 So, B. R., An, S., Kumar, S., Das, M., Turner, D. A., Hadad, C. M. and Musier-Forsyth, K. (2011) Substrate-mediated fidelity mechanism ensures accurate decoding of proline codons. J. Biol. Chem. **286**, 31810–20.

78 Martinis, S. A. and Boniecki, M. T. (2010) The balance between pre- and post-transfer editing in tRNA synthetases. FEBS Lett. **584**, 455–459.

79 Nangle, L. A., De Crecy Lagard, V., Doring, V. and Schimmel, P. (2002) Genetic code ambiguity. Cell viability related to the severity of editing defects in mutant tRNA synthetases. J. Biol. Chem. **277**, 45729–33.

80 Bacher, J. M. and Schimmel, P. (2007) An editing-defective aminoacyl-tRNA synthetase is mutagenic in aging bacteria via the SOS response. Proc. Natl. Acad. Sci. U. S. A. **104**, 1907–12.

81 WATSON, J. D. and CRICK, F. H. (1953) Genetical implications of the structure of deoxyribonucleic acid. Nature **171**, 964–7.

82 Bebenek, K., Pedersen, L. C. and Kunkel, T. A. (2011) Replication infidelity via a mismatch with Watson-Crick geometry. Proc. Natl. Acad. Sci. U. S. A. **108**, 1862–7.

83 Wang, W., Hellinga, H. W. and Beese, L. S. (2011) Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. Proc. Natl. Acad. Sci. U. S. A. **108**, 17644–8.

84 Singh, V., Fedeles, B. I. and Essigmann, J. M. (2015) Role of tautomerism in RNA biochemistry. RNA **21**, 1–13.

85 Satpati, P. and Åqvist, J. (2014) Why base tautomerization does not cause errors in mRNA decoding on the ribosome. Nucleic Acids Res. **42**, 12876–84.

86 Farabaugh, P. J. (2000) Translational frameshifting: implications for the mechanism of translational frame maintenance. Prog. Nucleic Acid Res. Mol. Biol. **64**, 131–70.

87 Tsuchihashi, Z. and Kornberg, A. (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. Proc. Natl. Acad. Sci. U. S. A. **87**, 2516–20.

88 Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J. and Varmus, H. E. (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature **331**, 280–3.

89 Brierley, I., Digard, P. and Inglis, S. C. (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell **57**, 537–47.

90 Diwan, G. D. and Agashe, D. (2016) The Frequency of Internal Shine-Dalgarno-like Motifs in Prokaryotes. Genome Biol. Evol. **8**, 1722–33.

91 Li, G.-W., Oh, E. and Weissman, J. S. (2012) The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature **484**, 538–541.

92 Harger, J. W., Meskauskas, A. and Dinman, J. D. (2002) An "integrated model"of programmed ribosomal frameshifting. Trends Biochem. Sci. **27**, 448–454.

93 Plant, E. P., Jacobs, K. L. M., Harger, J. W., Meskauskas, A., Jacobs, J. L., Baxter, J. L., Petrov, A. N. and Dinman, J. D. (2003) The 9-Å solution: How mRNA pseudoknots promote efficient programmed- 1 ribosomal frameshifting. Rna **9**, 168–174.

94 Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. C. and Brierley, I. (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. Nature **441**, 244–7.

95 Léger, M., Dulude, D., Steinberg, S. V and Brakier-Gingras, L. (2007) The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral programmed -1 ribosomal frameshift. Nucleic Acids Res. **35**, 5581–92.

96 Craigen, W. J. and Caskey, C. T. (1986) Expression of peptide chain release factor 2 requires high-efficiency frameshift. Nature **322**, 273–5.

97 Morris, D. K. and Lundblad, V. (1997) Programmed translational frameshifting in a gene required for yeast telomere replication. Curr. Biol. **7**, 969–76.

98 Matsufuji, S., Matsufuji, T., Wills, N. M., Gesteland, R. F. and Atkins, J. F. (1996) Reading two bases twice: mammalian antizyme frameshifting in yeast. EMBO J. **15**, 1360–70.

99 Kawakami, K., Pande, S., Faiola, B., Moore, D. P., Boeke, J. D., Farabaugh, P. J., Strathern, J. N., Nakamura, Y. and Garfinkel, D. J. (1993) A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in Saccharomyces cerevisiae. Genetics **135**, 309–20.

100 Shoemaker, C. J., Eyler, D. E. and Green, R. (2010) Dom34:Hbs1 Promotes Subunit Dissociation and Peptidyl-tRNA Drop-Off to Initiate No-Go Decay. Science. **330(6002)**, 369–372.

101 Shao, S., Brown, A., Santhanam, B. and Hegde, R. S. (2015) Structure and Assembly Pathway of the Ribosome Quality Control Complex. Mol. Cell **57**, 433–444.

102 Johansson, M., Zhang, J. and Ehrenberg, M. (2012) Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. Proc. Natl. Acad. Sci. U. S. A. **109**, 131–6.

103 Ribas de Pouplana, L., Santos, M. A. S., Zhu, J.-H., Farabaugh, P. J. and Javid, B. (2014) Protein mistranslation: friend or foe? Trends Biochem. Sci. **39**, 355–362.

104 Byrgazov, K., Vesper, O. and Moll, I. (2013) Ribosome heterogeneity: another level of complexity in bacterial translation regulation. Curr. Opin. Microbiol. **16**, 133–139.

105 Sauert, M., Temmel, H. and Moll, I. (2015) Heterogeneity of the translational machinery: Variations on a common theme. Biochimie **114**, 39–47.

106 Zaher, H. S. and Green, R. (2010) Hyperaccurate and error-prone ribosomes exploit distinct mechanisms during tRNA selection. Mol. Cell **39**, 110–20.

107 Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O. and Arnold, F. H. (2005) Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. U. S. A. **102**, 14338–43.

108 Yang, J.-R., Zhuang, S.-M. and Zhang, J. (2010) Impact of translational error-induced and error-free misfolding on the rate of protein evolution. Mol. Syst. Biol. **6**, 421.

109 Guo, H. H., Choe, J. and Loeb, L. A. (2004) Protein tolerance to random amino acid change. Proc. Natl. Acad. Sci. **101**, 9205–9210.

110 Bloom, J. D., Silberg, J. J., Wilke, C. O., Drummond, D. A., Adami, C. and Arnold, F. H. (2005) Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. **102**, 606–611.

111 Trudeau, D. L., Kaltenbach, M. and Tawfik, D. S. (2016) On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins. Mol. Biol. Evol. **33**, 2633–2641.

112 Shah, P. and Gilchrist, M. A. (2010) Effect of correlated tRNA abundances on translation errors and evolution of codon usage bias. PLoS Genet. **6**, e1001128.

113 Kochupurakkal, B. S., Iglehart, J. D., Wallace, J., Salehi-Ashtiani, K. and Hao, T. (2013) Nourseothricin N-Acetyl Transferase: A Positive Selection Marker for Mammalian Cells. PLoS One (Almeida-Porada, G., ed.) **8**, e68509.

114 Baker Brachmann, C., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P. and Boeke, J. D. (1998) Designer deletion strains derived fromSaccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast **14**, 115–132.

115 Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature **418**, 387–391.

116 Liu, C., Apodaca, J., Davis, L. and Rao, H. (2007) Proteasome inhibition in wild-type yeast Saccharomyces cerevisiae cells. Biotechniques **42(2)**, 158.

117 von der Haar, T. (2007) Optimized Protein Extraction for Quantitative Proteomics of Yeasts. PLoS One (Preiss, T., ed.) **2**, e1078.

118 Abcam. Anti-Ubiquitin antibody (ab19247). Available at: http://www.abcam.com/ubiquitin-antibody-ab19247.html [Acessed 26.11.2017].

119 Thermo Fisher Scientific. Restore Western Blot Stripping Buffer. Available at: https://www.thermofisher.com/order/catalog/product/21059 [Acessed 26.11.2017].

120 Thermo Fisher Scientific. SYPRO Ruby Protein Gel Stain. Available at: https://www.thermofisher.com/order/catalog/product/S12000 [Acessed 26.11.2017].

121 Sigma-Aldrich. cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail Protease Inhibitor Cocktail Tablets provided in a glass vial, Tablets provided in a glass vial. Available at: https://www.sigmaaldrich.com/catalog/product/roche/11836170001?lang=en&region=US [Acessed 26.11.2017].

122 GE Healthcare Life Sciences. Chelating Sepharose Fast Flow, 50 ml. Available at: http://www.gelifesciences.com/webapp/wcs/stores/servlet/catalog/en/GELifeSciences/products/AlternativeProductStructure_17344/17057501 [Acessed 26.11.2017].

123 Thermo Fisher Scientific. Slide-A-Lyzer Dialysis Cassettes, 10K MWCO, 3 mL. Available at: https://www.thermofisher.com/order/catalog/product/66381 [Acessed 26.11.2017].

124 Pall Corporation. Microsep™ Advance Centrifugal Devices - Plasmid Purification. Available at: https://shop.pall.com/us/en/laboratory/dna-rna-purification/plasmid-purification/microsep-advance-centrifugal-devices-zidgri78l64 [Acessed 26.11.2017].

125 Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. Nat. Methods **9**, 671–675.

126 Solomon, R. W. (2009) Free and open source software for the manipulation of digital images. AJR. Am. J. Roentgenol. **192**, W330-4.

127 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res. **28**, 235–242.

128 DeLano WL. (2002) The PyMOL Molecular Graphics System. Available at: https://pymol.org/ [Acessed 26.11.2017].

129 Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. **44**, W344–W350.

130 Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T. and Ben-Tal, N. (2013) ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. Isr. J. Chem. **53**, 199–206.

131 Ashkenazy, H., Erez, E., Martz, E., Pupko, T. and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. **38**, W529-33.

132 Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. and Ben-Tal, N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res. **33**, W299-302.

133 Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. and Ben-Tal, N. (2003) ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. Bioinformatics **19**, 163–164.

134 Eddy, S. R. S. (2004) Where did the BLOSUM62 alignment score matrix come from? **22**, 1035–1036.

135 Bulmer, M. (1988) Are codon usage patterns in unicellular organisms determined by selection-mutation balance? J. Evol. Biol. **1**, 15–26.

136 Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2**, 13–34.

137 Brandman, O., Stewart-Ornstein, J., Wong, D., Larson, A., Williams, C. C., Li, G.-W., Zhou, S., King, D., Shen, P. S., Weibezahn, J., et al. (2012) A Ribosome-Bound Quality Control Complex Triggers Degradation of Nascent Peptides and Signals Translation Stress. Cell **151**, 1042–1054.

138 Guydosh, N. R. and Green, R. (2014) Dom34 Rescues Ribosomes in 3′ Untranslated Regions. Cell **156**, 950–962.

139 Min, E. E., Roy, B., Amrani, N., He, F. and Jacobson, A. (2013) Yeast Upf1 CH domain interacts with Rps26 of the 40S ribosomal subunit. RNA **19**, 1105–1115.

140 Parr, C. L., Keates, R. A. B., Bryksa, B. C., Ogawa, M. and Yada, R. Y. (2007) The structure and function ofSaccharomyces cerevisiae proteinase A. Yeast **24**, 467–480.

141 Radhakrishnan, A., Chen, Y.-H., Martin, S., Alhusaini, N., Green, R. and Coller, J. (2016) The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. Cell **167**, 122–132.e9.

142 Lee, D. H. and Goldberg, A. L. (1996) Selective Inhibitors of the Proteasome-dependent and Vacuolar Pathways of Protein Degradation in Saccharomyces cerevisiae. J. Biol. Chem. **271**, 27280–27284.

143 Pickart, C. M. (2001) Mechanisms Underlying Ubiquitination. Annu. Rev. Biochem. **70**, 503–533.

144 Kursula, I., Heape, A. . and Kursula, P. Non-fusion GST from S. japonicum in complex with glutathione. To be Publ.

145 Atkinssg, W. M., Wangll, R. W., Bird+, A. W., Newtonll, D. J. and Lull, A. Y. H. (1993) The Catalytic Mechanism of Glutathione S-Transferase (GST). J. Biol. Chem. **268**, 19188–19191.

146 Javid, B., Sorrentino, F., Toosky, M., Zheng, W., Pinkham, J. T., Jain, N., Pan, M., Deighan, P. and Rubin, E. J. (2014) Mycobacterial mistranslation is necessary and sufficient for rifampicin phenotypic resistance. Proc. Natl. Acad. Sci. U. S. A. **111**, 1132–7.

147 Jones, T. E., Alexander, R. W. and Pan, T. (2011) Misacylation of specific nonmethionyl tRNAs by a bacterial methionyl-tRNA synthetase. Proc. Natl. Acad. Sci. U. S. A. **108**, 6933–8.

148 Netzer, N., Goodenbour, J. M., David, A., Dittmar, K. A., Jones, R. B., Schneider, J. R., Boone, D., Eves, E. M., Rosner, M. R., Gibbs, J. S., et al. (2009) Innate immune and chemically triggered oxidative stress modifies translational fidelity. Nature **462**, 522–526.

149 Jungreis, I., Lin, M. F., Spokony, R., Chan, C. S., Negre, N., Victorsen, A., White, K. P. and Kellis, M. (2011) Evidence of abundant stop codon readthrough in Drosophila

and other metazoa. Genome Res. **21**, 2096–113.

150 Kirby, A. J. (1996) Enzyme Mechanisms, Models, and Mimics. Angew. Chemie Int. Ed. English **35**, 706–724.

151 Amyes, T. L. and Richard, J. P. (2013) Specificity in transition state binding: the Pauling model revisited. Biochemistry **52**, 2021–35.

152 Stahl, G., Salem, S. N. Ben, Chen, L., Zhao, B. and Farabaugh, P. J. (2004) Translational accuracy during exponential, postdiauxic, and stationary growth phases in Saccharomyces cerevisiae. Eukaryot. Cell **3**, 331–8.

153 Stahl, G., Ben Salem, S., Li, Z., McCarty, G., Raman, A., Shah, M. and Farabaugh, P. J. (2001) Programmed +1 translational frameshifting in the yeast Saccharomyces cerevisiae results from disruption of translational error correction. Cold Spring Harb. Symp. Quant. Biol. **66**, 249–58.

154 Azpurua, J., Ke, Z., Chen, I. X., Zhang, Q., Ermolenko, D. N., Zhang, Z. D., Gorbunova, V. and Seluanov, A. (2013) Naked mole-rat has increased translational fidelity compared with the mouse, as well as a unique 28S ribosomal RNA cleavage. Proc. Natl. Acad. Sci. U. S. A. **110**, 17350-17355.

155 Kramer, E. B., Vallabhaneni, H., Mayer, L. M. and Farabaugh, P. J. (2010) A comprehensive analysis of translational missense errors in the yeast Saccharomyces cerevisiae. RNA **16**, 1797–808.

156 Béatrice C. Ortego, Jeremiah J. Whittenton, Hui Li, Shiao-Chun Tu, and and Richard C. Willson. (2007) In Vivo Translational Inaccuracy in Escherichia coli: Missense Reporting Using Extremely Low Activity Mutants of Vibrio harveyi Luciferase. Biochemistry **46**, 13864–13873.

157 Mohler, K., Aerni, H.-R., Gassaway, B., Ling, J., Ibba, M. and Rinehart, J. (2017) MS-READ: Quantitative measurement of amino acid incorporation. Biochim. Biophys. Acta - Gen. Subj. **1861(11 Pt B)**, 3081-3088.

158 Wen, D., Vecchi, M. M., Gu, S., Su, L., Dolnikova, J., Huang, Y.-M., Foley, S. F., Garber, E., Pederson, N. and Meier, W. (2009) Discovery and investigation of misincorporation of serine at asparagine positions in recombinant proteins expressed in Chinese hamster ovary cells. J. Biol. Chem. **284**, 32686–94.

159 Randau, L., Calvin, K., Hall, M., Yuan, J., Podar, M., Li, H. and Söll, D. (2005) The heteromeric Nanoarchaeum equitans splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves. Proc. Natl. Acad. Sci. U. S. A. **102**, 17934–9.

160 Soma, A., Onodera, A., Sugahara, J., Kanai, A., Yachie, N., Tomita, M., Kawamura, F. and Sekine, Y. (2007) Permuted tRNA Genes Expressed via a Circular RNA Intermediate in Cyanidioschyzon merolae. Science (80-.). **318**.

161 Harris, V. H., Smith, C. L., Jonathan Cummins, W., Hamilton, A. L., Adams, H., Dickman, M., Hornby, D. P. and Williams, D. M. (2003) The Effect of Tautomeric Constant on the Specificity of Nucleotide Incorporation during DNA Replication: Support for the Rare Tautomer Hypothesis of Substitution Mutagenesis. J. Mol. Biol.

326, 1389–1401.