Proteomics Analysis for Amino Acid Misincorporation Detection: Mini Review

Joana Tavares, Filipe Assis-Santos and Manuel AS Santos*
Department of Medical Sciences, BiMED- Institute of Biomedicine, University of Aveiro, 3910-193 Aveiro, Portugal

Abstract

Protein biosynthesis is a highly accurate biological process essential for life. Amino acid misincorporation errors (mistranslation) normally occur at low levels, but can increase sharply upon amino acid starvation, exposure to drugs, oxidative stress and other physiological perturbations. These processes disrupt protein function and are normally regarded as being deleterious, however, recent work has shown that they can also be regulated to produce advantageous phenotypes in both prokaryotes and eukaryotes. The biology of such unexpected adaptive mistranslation is poorly understood due to technical difficulties in the identification and quantification of amino acid misincorporations. In this mini-review, we describe proteome scale methodologies involving the use of mass-spectrometry and bioinformatics tools to directly detect and quantify mistranslation events and also indirect functional methods that permit sensitive, flexible and low-cost analysis of site specific amino acid variation.

Keywords: Amino acid misincorporation; Direct tools; Indirect tools; Mistranslation; Mass spectrometry; Reporter proteins

Introduction

The very high energetic and functional costs of protein synthesis require almost instantaneous tight regulation of its rate to guarantee cellular homeostasis during developmental programs, nutritional stress or other physiological perturbations. Fidelity of protein synthesis is equally important as it ensures functional activity and minimizes costly refolding or degradation of aberrant proteins. However, high synthesis rates impose constraints on accuracy that lead to an average error rate of 1 amino acid in every 1000 to 10000, implying that at least 15% of proteins have an average of 1 misincorporated amino acid [1-4]. Protein quality control systems, which in eukaryotes involve the ubiquitin-proteasome system (UPS), the unfolded protein response (UPR) and autophagy [5-7], ensure that levels of protein mistranslation are compatible with life. Mistranslation is produced by erroneous mischarging of transfer RNA (error rate 10^-4) and tRNA-mRNA mispairing in the ribosome (error in vitro 10^-1) [8,9]. However, these errors do not occur at the same level for all amino acids or with similar frequencies at different codon sites; chemically similar amino acids are more often mischarged on tRNAs, and near cognate codons (codons that differ in one base only) lead to higher error rates than noncognate codons. For example, the small difference of a single methyl group between Ile and Val leads to isoleucyl-tRNA synthetase (IleRS) mischarging of tRNA^Ile^ with Val at a rate of ~1% [10]. In E. coli and mammalian cells, Asn and His starvation results in Lys and Gln misincorporation, respectively [11], as low levels of charged tRNA^Asn^ and tRNA^His^ allow charged tRNA^Ile^ and tRNA^Val^ to read codons that differ in the third base [12,13].

Most surprisingly, several microorganisms use protein mistranslation to functionally diversify their proteome and expand adaptation capacity – adaptive mistranslation [14]. For instance, Mycoplasma spp have frequent point mutations and deletions in the editing site of ThRs, LeuRS and PheRS [15,16] and poorly discriminate mischarged non-cognate tRNAs [17]. AsnRS and GinRS are absent in Mycobacterium species, instead they produce charged tRNAs through natural Glu-tRNA^Glu^ and Asp-tRNA^Asp^ misacylation, followed by amino acid transamidation of the tRNA by the protein complex GatCAB [18]. Mutations in GatCAB subunits found in natural isolates of M. tuberculosis or specific physiological conditions increase Gln-to-Glu and Asn-to-Asp mistranslaction [19]. Similarly, the fungal pathogen Candida albicans mistranslates naturally at a very high level by misincorporating both Ser (97%) and Leu (3%) at CUG codons [20]. Moreover, the MetRS loses its natural charging specificity through phosphorylation (humans), temperature dependent structural shift (archaeon Aeropyrum pernix) and removal of a succinyllysine modification (E. coli), it also charges various noncognate tRNAs, leading to Met misincorporation into proteins at multiple codon sites [21-23]. These counterintuitive selective advantages produced by mistranslation [24,25] have direct phenotypic consequences. Indeed, in Mycobacterium tuberculosis PBE increase resistance to rifampicin, in Mycoplasma spp mistranslation contributes to evasion of the host's immune system by increasing cell surface variability (antigenic variation) and in C. albicans mistranslation increases survival in the mouse gut and accelerates evolution of resistance to fluconazole. In humans and other organisms Met misincorporation, which can be reversibly oxidized and reduced, protects proteins against oxidative stress with minimal disruption to their structure [20,25-28]. The above cases of regulated mistranslation contrast sharply with random amino acid misincorporation events induced by stress or by damage to the protein synthesis machinery. The latter normally disrupts protein structure, increases protein degradation and activates the proteotoxic stress network leading to cell death. It can also be a serious problem in the production of therapeutics. Overexpression of recombinant proteins in E. coli, mammalian cells or other heterologous hosts can result in high levels of amino acid misincorporation reducing the therapeutic value of the recombinant therapeutics. For example, monoclonal antibody production in batch cultures of Chinese hamster ovary (CHO) cells often results in Ser-to-Asn, Meta- and ortho-Tyr-to-Phe misincorporation [29]. Similar misincorporations are observed during production of IGF-1, Interleukin 2, human haemagglutinin and other proteins in E. coli [30-32], demonstrating that the identification and quantification of mistranslated proteins is relevant to the biotech
and pharmaceutical industries. Detection and quantification of amino acid misincorporation rates remain a significant challenge since the abundance of error-free proteins is considerably higher than that of aberrant proteins. A variety of mass spectrometry methods (MS) have been developed to directly identify and quantify amino acids misincorporated into proteins, but sensitivity, mass-spectra complexity and a lack of statistical and computational methods to filter rare mistranslated peptides continue to be important challenges that must be overcome. Fluorescent, luminescent, radioactive and other methods based on specific gain of function mutations or detection of radioactive peptides, respectively, have also been developed and are frequently used due to their low cost and relatively high reproducibility and robustness, however these methods only identify misincorporated amino acids in specific proteins or at specific protein sites and largely underestimate global misincorporation rates. We describe below the methods in some detail identifying their advantages and disadvantages.

**Indirect Measurements of Amino Acid Misincorporation Rates**

The indirect methods fall into three groups (Table 1): 1) use of a specific amino acid not observed in a reporter protein; 2) measurement of changes in the isoelectric point of a protein; 3) use of biochemical or genetic reporter systems. These methods detect a specific type of amino acid misincorporation at a time, normally a heterologous recombinant protein expressed in a specific host cell. Low frequency amino acid misincorporations are normally underrepresented or are not detected, and there is also significant error detection bias due to instability of the mistranslated recombinant proteins [33]. Loffeld and Vanderjagt were among the first scientists to detect amino acid misincorporations in proteins by incubating rabbit reticulocytes with radioactive valine and analyzing tryptic fragments of the purified haemoglobin. They estimated a misreading probability per codon of approximately 2-6 x 10^-3 [34]. Other groups have taken advantage of radiolabeled amino acids to detect the incorporation of ^35^S-cysteine into the cysteine-free flagellin of *E. coli* (Figure 1A) [2,35,36] and also at Arg codon sites of an *E. coli* ribosomal protein [2] by growing cells in a medium supplied with a radioactive amino acid that was not present in a protein of interest. Others estimated the error frequency during translation by detecting differences in the isoelectric point (IP) of selected proteins under normal conditions and during amino acid starvation for Asn and Lys [1,12,13]. Such IP differences result in the separation of mistranslated and wild type protein isoforms in two-dimensional (2D) gel electrophoresis (Figure 1B). During Asn starvation, Lys errors at Asn codons altered the isoelectric point of the MS2 coat protein and the measured error frequency reached 5 x 10^-3 [12,13]. Khazaie and collaborators evaluated radioactive phage proteins produced by *E. coli* using 2D gel electrophoresis and were able to detect multiple protein isoforms labeled with misincorporated His and Trp at similar error rates [36].

<table>
<thead>
<tr>
<th>Detection method</th>
<th>Principle</th>
<th>Type of error</th>
<th>Error rate/codon</th>
<th>Organism/Cell</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect Methods</td>
<td>Detection of a specific amino acid not found in the protein</td>
<td>Radioactively labeled Val</td>
<td>Ile → Val</td>
<td>2.0 - 6.0 x 10^-4</td>
<td>Rabbit</td>
</tr>
<tr>
<td>Detection of a specific amino acid not found in the protein</td>
<td>3SS-Cys incorporation in Cys-free flagellin</td>
<td>Arg → Cys</td>
<td>1.0 x 10^-4</td>
<td><em>E. coli</em></td>
<td>[35]</td>
</tr>
<tr>
<td>Detection of a specific amino acid not found in the protein</td>
<td>Radioactively labeled Cys and interference with protease digestion at Arg residues</td>
<td>Arg → Cys, Trp → Cys</td>
<td>10^-3, 3.0 - 4.0 x 10^-3</td>
<td><em>E. coli</em></td>
<td>[2]</td>
</tr>
<tr>
<td>Detection of a specific amino acid not found in the protein</td>
<td>Detection of radioactive His, Lys, Trp-free phage proteins</td>
<td>Lys, Leu, Trp misincorporation</td>
<td>3.0 - 4.0 x 10^-4</td>
<td><em>E. coli</em></td>
<td>[36]</td>
</tr>
<tr>
<td>Electrophoretic heterogeneity in the protein</td>
<td>Change of the isoelectric point pattern</td>
<td>Asn → Lys</td>
<td>5.0 x 10^-5</td>
<td><em>E. coli</em></td>
<td>[13]</td>
</tr>
<tr>
<td>Electrophoretic heterogeneity in the protein</td>
<td>Change of the isoelectric point pattern</td>
<td>Several misincorporations</td>
<td>4.0 x 10^-4</td>
<td><em>E. coli</em></td>
<td>[1]</td>
</tr>
<tr>
<td>Reporter system</td>
<td>Restoration of β-lactamase activity</td>
<td>Gly → Ser</td>
<td>10^-2</td>
<td><em>E. coli</em></td>
<td>[37]</td>
</tr>
<tr>
<td>Reporter system</td>
<td>Restoration of CAT activity</td>
<td>Tyr → His</td>
<td>0.5 x 10^-4</td>
<td><em>S. cerevisiae</em></td>
<td>[38]</td>
</tr>
<tr>
<td>Reporter system</td>
<td>Restoration of F-luc activity</td>
<td>Lys misincorporation</td>
<td>2.0 x 10^-4 - 3.6 x 10^-2</td>
<td><em>E. coli</em></td>
<td>[8]</td>
</tr>
<tr>
<td>Reporter system</td>
<td>Restoration of R-luc activity</td>
<td>Gin → Glu, Asn → Asp</td>
<td>0.20%, 0.80%</td>
<td><em>M. segmentis</em></td>
<td>[25]</td>
</tr>
<tr>
<td>Reporter system</td>
<td>Restoration of GFP fluorescence</td>
<td>Ser/Leu ambiguous codon</td>
<td>0 - 98%</td>
<td><em>C. albicans</em></td>
<td>[20]</td>
</tr>
<tr>
<td>Reporter system</td>
<td>Restoration of GFP fluorescence</td>
<td>Pro → Ala</td>
<td>~6%</td>
<td><em>S. cerevisiae</em></td>
<td>[53]</td>
</tr>
<tr>
<td>Reporter system</td>
<td>Restoration of mCherry fluorescence</td>
<td>Glu → Met</td>
<td>ND</td>
<td>Human cells</td>
<td>[42]</td>
</tr>
<tr>
<td>Direct Methods</td>
<td>Reporter protein</td>
<td>MRM mode MS/MS</td>
<td>Leu → Ser</td>
<td>1 - 2 %</td>
<td><em>S. cerevisiae</em></td>
</tr>
<tr>
<td>Reporter protein</td>
<td>MS/MS of modified side-chains</td>
<td>Asn → Asp</td>
<td>~10%</td>
<td><em>E. coli</em></td>
<td>[45]</td>
</tr>
<tr>
<td>Reporter protein</td>
<td>MS/MS and modified database</td>
<td>Ser → Asn</td>
<td>0.01 - 0.2%</td>
<td><em>E. coli</em>; CHO and NS0 cell lines</td>
<td>[44]</td>
</tr>
<tr>
<td>Reporter protein</td>
<td>MS/MS and precurson ion exclusion approach</td>
<td>Several misincorporations</td>
<td>0.0010%</td>
<td><em>E. coli</em>; mammalian cells</td>
<td>[47]</td>
</tr>
<tr>
<td>Reporter protein</td>
<td>MRM mode MS/MS</td>
<td>Phe → Tyr</td>
<td>0.02%</td>
<td><em>S. cerevisiae</em></td>
<td>[52]</td>
</tr>
<tr>
<td>Reporter protein</td>
<td>MS/MS and data processing with SPIDER algorithm</td>
<td>Pro → Ala</td>
<td>ND</td>
<td><em>S. cerevisiae</em></td>
<td>[53]</td>
</tr>
<tr>
<td>Detection of single amino acid misincorporations</td>
<td>Shotgun MS analysis and dynamic SILAC approach</td>
<td>Asp → Glu, Glu → Asp</td>
<td>~1 x 10^-3, ~0.5 x 10^-4</td>
<td>Mice</td>
<td>[54]</td>
</tr>
<tr>
<td>Detection of single amino acid misincorporations</td>
<td>Systematic shotgun MS analysis and Super-SILAC approach</td>
<td>Leu → Nva</td>
<td>~0.14%</td>
<td><em>E. coli</em></td>
<td>[55]</td>
</tr>
</tbody>
</table>

*Table 1: Methods for detection of mistranslation.*
Gain of function reporter systems are also widely used to determine amino acid misincorporation levels at specific codons. The most commonly used recombinant reporters are GFP, luciferases and β-gal that have loss of function mutations at critical functional residues and enzymatic activity can only be restored through misincorporation of the wild type amino acid (Figure 1C). Recovery of function can be quantified by measuring fluorescent, luminescent or chromogenic signals of those proteins providing an indirect measure of error rates, i.e., misincorporation frequency can be estimated as the ratio of mutant versus wild type enzymatic activity.

Toth and collaborators were the first to use a beta-lactamase gene with a mutation in a AGC (Ser) codon change to GGA/GGC – (Gly), which inactivates beta-lactamase activity. Expression of this gene in mutant E. coli with compromised translational fidelity restored beta-lactamase activity providing an indirect measure of Ser misincorporation at the mutant Gly sites [37]. Similar strategies using mutant chloramphenicol acetyl transferase (CATIII) were used to determine mistranslation levels in S. cerevisiae since His misincorporation at Tyr codon sites restores CATIII activity [38]. These reporter systems are sensitive but have the disadvantage of only measuring one type of misincorporation at a time and underestimate amino acid misincorporation rates. Kramer and Farabaugh used a dual reporter system consisting of a fusion of a mutant form of the firefly luciferase (F-luc) and Renilla luciferase (R-luc), whose luminescence can be measured independently. These two proteins are expressed as a single polypeptide, but their activities are independent, therefore, any difference in the activity of F-luc relative to R-luc reflects changes in protein activity. The F-luc has a series of inactivating mutations at a codon (LS529) site that encodes an essential lysine residue, permitting monitoring gain of function activity of F-luc caused by Lys misincorporation at the various mutant codons [8]. This approach was used and adapted by other groups and became widely used by the scientific community [25,39]. Javid and co-workers inactivated R-luc, instead of F-luc, by mutating the active site to detect and quantify Glu and Asp misincorporations at Glm and Asn codons in mycobacteria [25]. The above approaches have the caveat of requiring the addition of substrate or cofactors to cell lysates to measure enzymatic activity. To circumvent this limitation, a fluorescent reporter assay based on the expression of recombinant GFP capable of reporting amino acid misincorporation at a specific site was developed [20,40,41]. Bezerra et al. constructed a codon-optimized yEGFP gene, in which the UUG Leu codon at position 201 was mutated to an atypical CUG Ser codon in C. albicans. Leu misincorporation at the CUG position resulted in a gain of function of the fluorescent protein that was therefore able to report accurate levels of Leu misincorporation in that pathogenic fungus [20]. Recently, a dual fluorescent reporter (EGFP-mCherry fusion protein) was developed for the detection and quantification of Met mistranslation in mammalian cells. This reporter is a mCherry gene containing a mutation at position 72 (Met72) fused to a wild-type EGFP gene, which is used to determine the expression level of the reporter [42]. The advantage of these dual reporter systems relative to single protein reports is that they overcome biases in the determination of enzyme activity levels caused by variation in protein expression. The other advantage is that fluorescence can be detected by microscopy, flow cytometry and other methods allowing for a broader range of applications.

Direct Measurements of Amino Acid Misincorporation Rates

Advances in tandem mass spectrometry (MS/MS) has allowed researchers to directly measure mistranslation. Here, we differentiate MS/MS analysis of reporter proteins and MS/MS analysis of a total proteome (Table 1 and Figure 2).

MS/MS of reporter proteins

MS/MS analysis of reporter proteins purified to near homogeneity reduces the complexity of the MS spectra search space, increases the probability of detecting mistranslated peptides and the reproducibility...
of the measurements. Deep quantitative MS/MS analysis of the reporter peptides against an appropriate database can detect amino-acid substitutions and locate the specific sites where they occur along the primary structure. This highly sensitive approach permits identification and quantification of a wide variety of amino acid misincorporations, but does not produce a comprehensive view of mistranslation rates because single proteins rarely contain the 61 sense codons of the genetic code; codon context effects affect mistranslation rate and each protein only contains a small number of possible contexts of each codon [43,44]. In 2007, Silva and collaborators constructed a reporter system to detect Ser misincorporation at Leu CUG codon sites in C. albicans. The target protein was expressed, purified by affinity chromatography, digested and the peptides were separated chromatographically before mass spectrometric analyses in multiple reaction monitoring (MRM) mode [43]. Another group quantified Asp misincorporation in recombinant Dihydrofolate Reductase (DHFR) that was digested and methylated before MS/MS analysis. The level of amino acid misincorporation was calculated as the ratio of absolute intensities of the mistranslated peptide to the WT peptide peaks [45]. Several mistranslation errors were identified using this approach in different organisms and cell lines, namely CHO cells, which are frequently used for recombinant expression of target proteins [44,46-49]. This strategy was also used to demonstrate that the tyrosyl-tRNA synthetase is highly error prone [50,51] and that the editing site of some ami noacyl-tRNA synthetases controls the sensitivity of amino acid stress responses [52].

Hoffman and collaborators combined direct and indirect methods to detect Ala substitutions in a reporter protein for MS/MS analysis (Two Tel2-Interacting protein – Tti2) and in EGFP reporter protein. Comparison of the MS/MS and EGFP reporter data showed similar levels of Ala misincorporation at Pro codons, indicating that both methods are reliable and can be used alone or in combination [53].

Proteome scale MS/MS analysis

Single amino acid substitutions in the whole proteome of an organism or cell line can be detected using a systematic shotgun mass spectrometry-based proteomics approach. In this type of approach, complex protein samples are digested with a protease, usually trypsin, and digested peptides are fractionated on HPLC columns, prior to tandem mass spectrometry analysis. Identification of proteins is achieved by searching the MS/MS spectra against a protein database, in which spectra derived from peptide fragmentation is compared with theoretical spectra generated from in silico digestion of the protein database. Recently, quantification of single amino acid misincorporations was achieved using SILAC methods, which use stable isotopes to label amino acids in cell culture. Tuorto et al. identified high frequency of Asp to Glu substitutions and vice-versa in the proteome of mutant mice bone marrow using dynamic SILAC techniques [54]. Dynamic SILAC is a technique that allows detection of variances in de novo protein synthesis since labeled amino acids are present in the growth medium for a short period of time. On the other hand, Cveticic et al. detected norvaline misincorporation in the E. coli proteome using Super-SILAC labeling [55]. In this work, labeled proteins from different samples were mixed and used as a spike-in standard, allowing a more accurate quantification of unlabeled samples [56]. Along with the SILAC approaches, a spectral counting strategy has been also used to quantify mistranslation [55]. This approach combines the number of identified MS/MS spectra from the same protein throughout multiple LC-MS runs to assess relative protein abundance. A handful of researchers have been able to identify a small number of misincorporations from large MS spectra databases using prior knowledge of the misincorporated amino acid and codon sites, demonstrating that identification and quantification of misincorporated amino acids by MS/MS is feasible. However, the high percentage of false positives observed in those data sets highlights the need to optimize the methodologies used in sample preparation, MS/MS and bioinformatics. A successful proteomics study should be well designed, selecting the sample preparation approach that best fits the experiment to improve detection of low abundance proteins and peptides [57]. Indeed, the major challenge in detecting mistranslated peptides is their low abundance against a background of wild-type peptides. Additionally, amino acid misincorporations detected by MS/MS can be misinterpreted as post-translational modifications, such as phosphorylation, glycosylation, nitration and acylation, reducing the reliability and reproducibility of this approach [47]. Bioinformatic tools are also critical in this field. Some researchers use algorithms that search against a database containing all possible modified peptides of their target proteins [45]. However, these databases provide a biased and limited view of the mistranslated peptides, forcing spectra to be identified as peptides present in the manipulated database [58]. To overcome these limitations, one could use error-tolerance searches, such as de novo approaches, at the amino acid sequence level since de novo sequencing identifies peptide sequences by analyzing MS/MS spectra. Other algorithms combine de novo sequencing with database search of the candidate peptides looking for homology [59]. This kind of strategy could help in the detection of misincorporations at a global proteome scale, decreasing the number of false positives.

Conclusion

Several approaches are now available to identify and quantify protein biosynthesis errors, particularly those alloying MS sensitivity approaches and reporter protein systems. Despite the technical difficulties in the identification and quantification of mistranslation errors at the proteome scale, recent works show that errors are possible to quantify and it is likely that future improvements in mass spectrometry and bioinformatics will allow us to tackle the relevance of protein synthesis errors for aging and human disease in a robust and comprehensive manner. Such improvements should also allow the development of methods and host cells essential for the production of high quality recombinant therapeutics.

Acknowledgments

This work was supported by the Portuguese Foundation for Science and Technology through COMPETE2020 and FEDER funds (project FCT-PTDC/BEX-BCM/2121/2014) and Integrated Projects FEDER/CENTRO-45-2015-01.

References


