Glycine to glutamic acid misincorporation observed in a recombinant protein expressed by *Escherichia coli* cells

Yunping Huang,* Brian O'Mara, Matthew Conover, Richard Ludwig, Jinmei Fu, Li Tao,[†] Zheng Jian Li,[†] Siegfried Rieble,[†] Michael J. Grace,[†] and Reb J. Russell[†]

Department of Biologics Product & Process Development, Bristol-Myers Squibb Technical Operations, Princeton, New Jersey 08543

Received 21 December 2011; Accepted 10 February 2012

DOI: 10.1002/pro.2046

Published online 17 February 2012 proteinscience.org

Abstract: A novel amino acid misincorporation, in which the intended glycine (Gly) residues were replaced by a glutamic acid (Glu), was observed in a recombinant protein expressed by *Escherichia coli*. The misincorporation was identified by peptide mapping and liquid chromatography–tandem mass spectrometric analysis on proteolyzed peptides of the protein and verified using the corresponding synthetic peptides containing the misincorporated residues. Analysis of the distribution of the misincorporated residues and their codon usage shows strong correlation between this misincorporation and the use of rarely used codon within the *E. coli* expression system. Results in this study suggest that the usage of the rare codon GGA has resulted in a Glu for Gly misincorporation.

Keywords: glycine to glutamic acid misincorporation; recombinant; *Escherichia coli*; rare codon; GGA

Introduction

The synthesis of a protein starting from genetic code through complete assembly of a polypeptide chain is the culmination of a multistep process involving deoxyribonucleic acid (DNA) replication, messenger ribonucleic acid (mRNA) transcription, and polypeptide synthesis. Conceptually, errors within any stage of gene expression can result in misincorporation of an unintended amino acid into the polypeptide chain, generating a sequence variant of the target protein. Although the overall fidelity of protein synthesis relies on the combined accuracy of all processes involved, each process has its own probability of introducing a variant. Although misincorporation rates have been found to vary under different

expression conditions and between different species,¹ the often-quoted estimates of error rates for the individual processes in typical cells are as follows: ~ 1 in 10^8 for DNA replication by bacteriophages, *Escherichia coli* (*E. coli*), and various eukaryotes,² ~ 1 in 10^5 bases for transcription in *E. coli*,^{3,4} and ~ 1 in 10^4 for codon translation in proteins produced by *E. coli* and mammalian expression systems.^{1,5} Furthermore, the error rates can be significantly higher under stressed conditions, such as presence of mutagenic reagents,⁶ exposure to reactive oxygen species,⁷ amino acid starvation,⁸ and the use of high-yield expression systems.⁹ It is well documented that overexpression can lead to nutritional stress on host cells, which in turn can cause increased misincorporation in recombinant proteins, especially in heterologous systems.^{8,9} Nevertheless, overexpression of recombinant proteins is almost always desirable in the production of recombinant proteins intended for structural and functional analysis in biochemical studies as well as for the manufacture of therapeutic proteins.

The consequence of amino acid misincorporation in a recombinant protein can be varied and may result in altered catalytic constants, specificity, and

[†]The authors are employees of Bristol-Myers Squibb Company and have received stock and other compensation from BMS as part of this employment. This work was funded by Bristol-Myers Squibb Company.

*Correspondence to: Dr. Yunping Huang, Department of Biologics Product & Process Development, Bristol-Myers Squibb Technical Operations, P.O. Box 5400, Princeton, NJ 08543. E-mail: yunping.huang@bms.com

stability.^{10–13} Within proteins produced for therapeutic use, amino acid misincorporation can potentially induce immune response or abnormal receptor–ligand interactions.¹⁴ Classified as a product-related impurity, the misincorporated populations in protein therapeutics are generally difficult to remove during downstream purification and must be controlled and minimized within the upstream fermentation process. They must be assessed to ensure drug safety and efficacy. In addition, characterization and control are also required by the regulatory authorities. Therefore, it is essential to understand the mechanisms underlying the cellular processes that contribute to the generation of sequence variants and control the production process to reduce or eliminate sequence variants whenever possible.

Misincorporation of amino acids occurs in both nature and recombinant proteins. Deleterious mistranslations have been the cause of disease in both humans and mice and can be detrimental to life.¹⁵ On the other hand, misincorporation can play an important role in the evolution of novel proteins by generating phenotypic and genetic diversity.¹⁵ Various types of amino acid misincorporations have been reported where mouse or human proteins were expressed in *E. coli* cells, including: norleucine for Met,¹⁶ norvaline for Leu,¹⁷ Gln for Arg,¹⁸ Lys for Arg,^{19,20} Gln for His,²¹ and Cys for Tyr.²² Most of the sequence variants are at low levels; however, higher levels of misincorporation have also been reported. For example, a Lys for Arg misincorporation produced two functional proteins with different properties at almost a 1:1 ratio.²⁰ Although mammalian expression systems are generally considered to be of higher fidelity, amino acid misincorporations have also been observed in recombinant proteins expressed from Chinese Hamster Ovary cells such as Gln for Tyr,²³ Leu for Phe,²⁴ Ser for Asn,²⁵ and Asn for Ser,²⁶ Arg for Ser,⁶ Thr for Pro,²⁷ Arg for Met,²⁷ Gln for Leu,²⁷ and Gly for Ser.²⁸

Historically, high levels of misincorporation have been detected using electrophoresis-based methodologies such as isoelectric focusing⁹ or two-dimensional gels.⁸ Edman sequencing¹⁷ and ion-exchange chromatography²⁷ have also been used. In many cases, the misincorporated protein would need to be isolated from the normal protein pool, which if possible is typically very difficult. Characterization of low levels of misincorporation remains a challenge because the error-free molecules are typically several orders of magnitude higher in concentration than the misincorporated ones, thus overloading most unbiased detection methods. Because of recent advances in mass spectrometry (MS) including increased sensitivity, selectivity (resolution), and accuracy, coupled with the ability to perform analysis with no prerequisite for isolation of the misincorporated proteins before analysis, MS has become the method of choice for characterizing protein misincor-

poration. Misincorporated proteins can be detected using intact mass analysis of a whole protein if the misincorporation is not isobaric with the native protein. Furthermore, peptide mapping can be performed to confirm the presence of misincorporation and identify the location of the misincorporated amino acid using online liquid chromatography–tandem mass spectrometric (LC–MS/MS) analysis.

Here, we report that Gly to Glu misincorporation can occur under typical recombinant protein production conditions. This phenomenon was observed in a 100-amino acid recombinant, Protein X, expressed in *E. coli*. Misincorporation has been identified at two out of nine Gly sites and was found to be specific to the GGA codon.

Results and Discussion

Identification of Gly to Glu misincorporation within the tryptic peptide T1 of protein X

During LC–MS/MS analysis of tryptic peptides of Protein X, a +72.018 Da variant of the T1 peptide (sequence MGVSDVPR) was observed. The extracted-ion chromatograms (EICs) of the expected and variant T1 peptides (doubly charged ions at m/z 430.719 and 466.728, respectively, ± 10 ppm) illustrate that these two peptides are not chromatographically resolved in the reversed-phase HPLC [22.6 min vs. 22.5 min, Fig. 1(A,B)]. The collision-induced dissociation MS/MS spectra of the expected and variant peptides are compared in Figure 2(A,B). The differences in b_2 and b_3 ions in the two spectra [m/z 189.1 and 288.2 for expected T1 peptide in Fig. 2(A), and m/z 261.1 and 360.2 for the variant peptide in Fig. 2(B)] indicate a change in either the first or the second amino acid. Consistent with this data, the spectra of these two species share the same y series ions up to y_6 and diverge at the y_7 ions [m/z 729.5 for the expected T1 peptide in Fig. 2(A) and m/z 801.5 for the variant peptide in Fig. 2(B)]. These data suggest the difference exists in the second amino acid. The +72.018 Da difference in the parent ions is attributed to an elemental difference of $C_3O_2H_4$ between the two species, which is consistent with a Gly to Glu misincorporation variant. In addition, the y_6 ion (m/z 672.4) and b_2 ion (m/z 261.2) in Figure 2(B) are more abundant than the y_6 ion (m/z 672.5) and b_2 ion (m/z 189.1) in Figure 2(A), which can be attributed to the presence of Glu residue instead of the native Gly residue. This preferential cleavage of the C-terminal bond of acidic residue (glutamic acid) has been reported in a line of studies in the cases of peptides containing acidic residues.^{29–31}

The misincorporation of Gly2Glu in the T1 peptide was further confirmed with the use of the synthetic peptide MEVSDVPR, which was spiked into the tryptic digest of Protein X. The synthetic peptide

Figure 1. Extracted-ion chromatograms of the expected and variant T1 peptides in the digest and the spiked-in synthetic variant T1 peptide (+2 ions, ± 10 ppm). A: Expected T1 peptide MGVS DVPR; B: variant peptide MEVS DVPR; and C: spiked-in synthetic peptide MEVS DVPR.

MEVS DVPR eluted at the same retention time (22.5 min) as the variant peptide observed in the non-spiked sample digest, as shown in Figure 1(C). The variant T1 peptide observed in the nonspiked sam-

ple also showed the same mass spectrum and MS/MS fragmentation pattern as those observed in the corresponding synthetic peptide, with MS/MS shown in Figure 2(B,C), respectively.

Figure 2. MS/MS spectra of the expected and variant T1 peptides in the digest and the spiked-in synthetic variant T1 peptide. A: Expected T1 peptide MGVS DVPR (parent ion m/z 430.7); B: variant peptide MEVS DVPR (parent ion m/z 466.7); and C: spiked-in synthetic peptide MEVS DVPR (parent ion m/z 466.7).

Figure 3. Extracted-ion chromatograms of the expected and variant chymotryptic peptides in the digest (+2 ions, ± 10 ppm). A: Expected chymotryptic peptide GETGGNSPVQEF and B: variant chymotryptic peptide GETEGNSPVQEF.

Absence of the Gly to Glu misincorporation variant of desMet T1 peptide in tryptic digest

It is of interest to note that the corresponding Gly to Glu misincorporated variant of desMet T1 (T1 with N-terminal methionine hydrolyzed, sequence EVSDVPR) was not observed within the EICs even though the native desMet T1 peptide (sequence GVSDVPR) was found to be the more abundant species relative to the native full-length T1 peptide (data not shown). When the digested sample is spiked with the synthetic peptide EVSDVPR, a new peak is observed in the chromatogram, demonstrating that if the variant was present it would be detectable under the conditions of the experiment. This indicates if the variant (EVSDVPR) is present the misincorporation level for this peptide variant is significantly less than its full-length counterpart.

The presence of Gly to Glu misincorporation in full-length T1 peptide and absence of Gly to Glu misincorporation in desMet T1 peptide may be attributed to the relative hydrolysis rates of the leading methionine for the native and variant proteins. Hirel *et al.*³² reported that the excision rate of N-terminal methionine in *E. coli* is governed by the side-chain length of the second amino acid. The extent of methionine cleavage is 97.1% if the second amino acid is Gly and 4.97% if the second amino acid is Glu in the proteins studied in their report. Therefore, when Gly to Glu misincorporation occurs at Gly₂, the hydrolysis of the N-terminal methionine in the variant protein is significantly suppressed compared with that of its native counterpart. As a result, the relative amount of the Gly2Glu variant is

enriched within the full-length protein. It is this enrichment that raised the level of the full-length T1 variant peptide to a level that warranted further investigation, leading to its identification as a Gly to Glu misincorporation.

Observation of Gly to Glu misincorporation in tryptic and chymotryptic digests

A second Gly to Glu misincorporation site was observed within the T4 tryptic peptide (ITY-G₃₉ETG₄₂G₄₃NSPVQEFTVPSR) in Protein X. Although the MS/MS spectrum is available, it is difficult to locate the exact site of Gly to Glu misincorporation because of the size of the T4 tryptic peptide (20 residues) and the presence of three glycines. However, chymotrypsin digestion yielded a shorter peptide (12 residues), in which the Gly to Glu misincorporated variant has also been identified. The EICs (EIC, ± 10 ppm) and corresponding MS/MS spectra of the native (G₃₉ETG₄₂G₄₃NSPVQEF, m/z 611.272, +2 ions) and variant (m/z 647.282, +2 ions) chymotryptic peptide are shown in Figures 3 and 4, respectively. As shown in Figure 4, MS/MS spectra of these two species share the same y series ions up to y₈ (m/z 877.4), which indicates that the Gly to Glu misincorporation is not at the Gly₄₃ residue. In addition, MS/MS spectra of these two species also share the same b₃ ions (m/z 288.2), which suggests that Gly to Glu misincorporation is not at G₃₉ residue, either. Combined, this information indicates that the Gly to Glu misincorporation site is at the Gly₄₂ residue. Also, the peak at m/z 417.2 matches the b₄ ion from the Gly42Glu misincorporated

Figure 4. MS/MS spectra of the expected and variant chymotryptic peptides in the digest. A: Expected chymotryptic peptide GETGGNSPVQEF and B: variant chymotryptic peptide GETEGNSPVQEF.

variant peptide [Fig. 4(B)]. The b_4 and y_8 ions in Figure 4(B) (m/z 417.2 and 877.4) are more abundant than those in Figure 4(A) (m/z 345.1 and 877.4), which can be attributed to the presence of the acidic Glu₄₂ residue instead of the native Gly₄₂ residue.

To further confirm the Gly42Glu misincorporation, synthetic peptides of the three possible Gly to Glu misincorporated variant peptides (EETGGNSPV-

QEF, GETEGNSPVQEF, and GETGENSPVQEF) were spiked into the chymotryptic digest of Protein X, followed by LC-MS/MS analysis. The EICs, MS spectra, and MS/MS spectra of the synthetic peptides were compared with those of the variant peptides from the Protein X digest. Although all three of the synthetic peptides have the same molecular mass (1292.552 Da), only synthetic peptide GETEGNSPVQEF shared the same retention time

Figure 5. MS/MS spectra of the spiked-in synthetic peptides. A: GETGENSPVQEF; B: GETEGNSPVQEF; and C: EETGGNSPVQEF.

as the variant peptide in Protein X digest (29.9 min). The other two variant peptides were observed as new peaks within the spiked-sample chromatograms (30.5 min for GETGENSPVQEF and 29.4 min for EETGGNSPVQEF). Moreover, when the MS/MS spectra were compared [Fig. 5(A–C)], only the MS/MS fragmentation pattern of GETEGNSPVQEF matched that of the variant peptide from Protein X digest [Figs. 4(B) and 5(B)]. As shown in Figure 5, although the three synthetic peptides shared the same y series ions up y₇ (*m/z* 820.4), the y₇ ion for peptide GETGENSPVQEF [in Fig. 5(A)] is more abundant than the y₇ ions for the other two peptides [in Fig. 5(B,C)]; the y₈ ion for peptide GETEGNSPVQEF [*m/z* 877.5 in Fig. 5(B)] is more abundant than the y₈ ions for the other two peptides [*m/z* 949.6 in Fig. 5(A) and *m/z* 877.5 in Fig. 5(C)]; the b₃ ion for peptide EETGGNSPVQEF [*m/z* 360.2 in Fig. 5(C)] is different from the b₃ ions for the other two peptides [*m/z* 288.2 in Fig. 5(A,B)]. These observations are consistent with their respective Gly to Glu misincorporation and the previously reported preferential fragmentations at the C-terminal of an acidic residue of a peptide.^{29–31}

Estimation of the relative levels of Gly to Glu misincorporation and codon correlation

There are nine glycines in Protein X. The levels of Gly to Glu misincorporation were evaluated for each Gly in three lots of Protein X using tryptic peptides, as shown in Table I. Chromatographic separation of the variant peptides from the native forms was not successful on the chromatographic system used. For this reason, relative quantitation using ultraviolet detection could not be performed. Estimation of the percentage of misincorporation was determined by dividing the peak area of the variant peptide peak by the sum of the peak areas for the native and variant peptides based on their EICs. This approach provides only an estimation of the relative percentages of the misincorporated peptide levels because of an uncertainty in the relative levels of peptide ionization efficiency. Verification of relative ionization efficiency was made using available synthetic peptides. The ionization efficiencies of the peptides tested fell within twofold of one another, providing an acceptable level of confidence in the misincorporation level estimates. When estimating the level of misincorporation at Gly₂ residue, the percentage was determined by including both the desMet and full-length peptide peak areas within the calculation. The misincorporation levels at Gly₂ were estimated to be between 2.2 and 9.5% when only full-length variants were considered. However, the levels dropped down to 0.4–1.8% when the calculated estimate included data for both full-length and desMet variants. The misincorporation levels at Gly₄₂ were found to be 0.12–0.15%. Although Gly to Glu misincorporations

Table I. Estimated Amount of Gly to Glu Misincorporation in Protein X

Peptide	Codon	% G to E misincorporation		
		Lot 1	Lot 2	Lot 3
Full-length T1	Gly ₂ (GGA)	9.5	2.2	3.4
desMet T1	Gly ₂ (GGA)	ND ^a	ND	ND
Normalized Gly to Glu misincorporation at Gly ₂		1.8	0.76	0.44
T2	Gly ₂₅ (GGT)	ND	ND	ND
T4	Gly ₃₉ (GGC)	0.14	0.12	0.15
	Gly ₄₂ (GGA)			
	Gly ₄₃ (GGC)			
T6	Gly ₆₃ (GGT)	ND	ND	ND
	Gly ₆₇ (GGT)			
T9	Gly ₉₇ (GGT)	NA ^b	NA	NA
	Gly ₉₉ (GGT)			
Estimated percent of protein with a misincorporated Glu		1.9	0.9	0.6

^a Amounts were estimated from the peak areas in the extracted-ion chromatograms of the normal and misincorporated peptides.

^b T9 is highly hydrophilic, which elutes with buffer front, and therefore cannot be quantified accurately.

are consistently observed in Gly₂ and Gly₄₂ within each of the three protein lots tested, the variant has not been observed at any of the other seven possible sites.

When the codon usage at each glycine site was analyzed (listed in Table I), a correlation between codon usage and Gly to Glu misincorporation is evident. Glycine can be encoded by four different codons (GGT, GGC, GGA, and GGG), which are all frequently used in mammalian systems. Although expressed in *E. coli*, a human gene construct is used in the expression of Protein X: codon GGA was encoded for both Gly₂ and Gly₄₂, whereas codons GGT and GGC were encoded for the remaining glycines within the cDNA construct. However, although the GGT and GGC codons are frequently seen in *E. coli* genes, the GGA and GGG codons are rarely used in this system.³³ It is known that difference in codon usage between prokaryotes and eukaryotes can have a significant impact on heterologous protein expressions, such as low protein expression level, mRNA and plasmid stability, and amino acid misincorporation.³⁴ It has been reported that the rare codons GGA and GGG have caused low expression levels of human tropoelastin in *E. coli*.³³ Results in this study indicate that usage of the rare codon GGA has resulted in a Glu for Gly misincorporation.

Amino acid misincorporation can be attributed to variation or mutation in the DNA sequence, error in mRNA transcription, and mistranslation during protein synthesis. In our study, the observed range of 10^{−2}–10^{−3} in Gly to Glu misincorporation is significantly higher than the published error rates for

mutation, transcription, or translation. Although no DNA mutation was observed during DNA sequencing, if present the level may have been below the level detectable by the method used. Even though it cannot be ruled out completely, the misincorporation being caused by genomic DNA mutation—the probability of DNA mutations at Gly₂ and Gly₄₂ simultaneously—should be very low because of the nature of DNA mutation. Similarly, it is unlikely that the misincorporation is caused by error in mRNA transcription. Amino acid misincorporation during translation may be the result of either mischarging the cognate transfer RNA (tRNA) or codon-anticodon mismatch between the ribosome and tRNA (misreading). If the tRNA was mischarged, it would be expected to give rise to misincorporation at all Gly locations because a common glycyl-tRNA synthetase charges all glycine-coding tRNAs. However, all detectable glycine replacements by glutamic acid are specific to the GGA codon in all of the lots investigated here. Therefore, it is unlikely that the tRNA is being mischarged. It is apparent that the Gly to Glu misincorporation observed in this study is related to rare codon usage by the *E. coli* expression system. It is possible that misreading of the second base—GGA (codon for Gly) with GAA (codon for Glu)—resulted in the observed misincorporation. It has been reported previously that a second base misread of the rare arginine codon AGA resulted in substitution with lysine codon AAA in insulin-like growth factor³⁵ and an HIV-1gp160 fragment.¹⁹ For the time being, more work is needed to fully understand the cause of the Gly to Glu misincorporation observed in this study.

Conclusions

Glycine to glutamic acid misincorporation has been observed in a recombinant human protein expressed in *E. coli* by peptide mapping and LC-MS/MS analysis. Misincorporation has been identified at two out of nine Gly sites and was found to be specific to the GGA codon, a frequent codon for humans but a rare one for *E. coli*, indicating possible correlation between misincorporation and rare codon usage.

Materials and Methods

Materials

Synthetic peptides were obtained from AnaSpec (Fremont, CA). Sequencing grade-modified trypsin was purchased from Promega (Madison, WI). Sequencing grade chymotrypsin was purchased from Roche Applied Science (Penzberg, Germany). All other chemicals were obtained from Sigma (St. Louis, MO) in highest grades commercially available.

Expression and purification

Protein X was expressed in *E. coli* as an inclusion body (IB). After fermentation, the cells were lysed

with the aid of a microfluidizer and the IBs were isolated using a centrifuge followed by a series of detergent and buffer washes. The washed IBs were solubilized and reduced using 4M guanidine hydrochloride, and 3.3 mM tris(2-carboxyethyl)phosphine hydrochloride in acetate buffer at pH 5.0. The protein was then refolded by batch dilution into a 20 mM citrate, pH 2.1, buffer. After the refold was complete, the material was filtered and purified using cation-exchange chromatography and hydrophobic interaction chromatography.

Sample preparations

For trypsin digestion, the Protein X samples (about 1 mg/mL) in acetate buffer were adjusted to pH 7 before trypsin was added at an enzyme to protein ratio of 1:25 (weight/weight). The protein/trypsin mixture was incubated at 37°C for 3 h. The digestion reaction was terminated by adjusting the solution pH to 2 by adding aqueous 2M trifluoroacetic acid (TFA). Chymotrypsin digestion was performed by incubating chymotrypsin and Protein X mixture at an enzyme-to-protein ratio of 1:30 in the digestion buffer (2M urea, 100 mM tris, and 10 mM CaCl₂, pH 7.6). The mixture was incubated at room temperature for 24 h.

For spiking experiments, the synthetic peptides were prepared at 0.1 mg/mL in water and spiked into digests at a 1:100 ratio (v/v).

LC-MS/MS. The tryptic digest and chymotryptic digest were both chromatographically separated using a Waters H Class UPLC before being analyzed by an LTQ Orbitrap mass spectrometer. Approximately 10 µg of digested proteins was injected for LC-MS/MS analysis. A Waters Acquity BEH C18 column (1.7 µm, 2.1 mm × 150 mm) was used for separation (at 40°C). For trypsin digest, a linear gradient of 2–80% mobile phase B over 110 min was used to elute the peptides (mobile phase A: 0.1% acetic acid and 0.005% TFA in water; mobile phase B: 0.1% acetic acid and 0.005% TFA in acetonitrile) at a flow rate of 0.05 mL/min. For chymotrypsin digest, a gradient of 2–80% mobile phase B over 65 min was used to elute the peptides (mobile phase A: 0.1% acetic acid and 0.05% TFA in water; mobile phase B: 0.1% acetic acid and 0.05% TFA in acetonitrile) at a flow rate was 0.1 mL/min.

The LTQ Orbitrap mass spectrometer was operated in the positive ion mode with a spray voltage of 3.5 kV. The heater temperature and capillary temperature were set at 300 and 350°C, respectively. MS acquisition was performed in the Orbitrap with a resolution of 30,000, and the 10 most intense ions in each duty cycle were selected for MS/MS in the LTQ Velos. The dynamic exclusion feature was enabled. Collision energy of 35% was used to obtain fragmentation spectra. Thermo Proteome DiscovererTM 1.0 was used for data analysis.

References

- Reynolds NM, Lazazzera BA, Ibba M (2010) Cellular mechanisms that control mistranslation. *Nat Rev Microbiol* 8:849–856.
- Kunkel TA (2004) DNA replication fidelity. *J Biol Chem* 279:16895–16898.
- Rosenberge RF, Foskett G (1981) An estimate of the frequency of in vivo transcriptional errors at a non-sense codon in *Escherichia coli*. *Mol Gen Genet* 183:561–563.
- Blank A, Gallant JA, Burgess RR, Loeb LA (1986) An RNA polymerase mutant with reduced accuracy of chain elongation. *Biochemistry* 25:5920–5928.
- Kramer EB, Farabaugh PJ (2007) The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* 13:87–96.
- Guo D, Gao A, Michels DA, Feeney L, Eng M, Chan B, Laird MW, Zhang B, Yu XC, Joly J, Snedecor B, Shen A (2010) Mechanisms of unintended amino acid sequence changes in recombinant monoclonal antibodies expressed in Chinese Hamster Ovary (CHO) cells. *Biotechnol Bioeng* 107:163–171.
- Ling J, Söll D (2010) Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site. *Proc Natl Acad Sci USA* 107:4028–4033.
- Parker J, Pollard JW, Friesen JD, Stanners CP (1978) Stuttering: high-level mistranslation in animal and bacterial cells. *Proc Natl Acad Sci USA* 75:1091–1095.
- Scorer CA, Carrier MJ, Rosenberger RF (1991) Amino acid misincorporation during high-level expression of mouse epidermal growth factor in *Escherichia coli*. *Nucleic Acids Res* 19:3511–3516.
- Langridge J (1974) Mutation spectra and the neutrality of mutations. *Aust J Biol Sci* 27:309–319.
- Nene V, Glass RE (1984) Genetic studies on the beta subunit of *Escherichia coli* RNA polymerase. IV. Structure-function correlates. *Mol Gen Genet* 194:166–172.
- Knowles JR (1987) Tinkering with enzymes: what are we learning? *Science* 236:1252–1258.
- Cupples CG, Miller JH (1988) Effects of amino acid substitutions at the active site in *Escherichia coli* beta-galactosidase. *Genetics* 120:637–644.
- Harris RJ (2005) Heterogeneity of recombinant antibodies: linking structure to function. *Dev Biol* 122:117–127.
- Moura GR, Carreto LC, Santos MA (2009) Genetic code ambiguity: an unexpected source of proteome innovation and phenotypic diversity. *Curr Opin Microbiol* 12:631–637.
- Lu HS, Tsai LB, Kenney WC, Lai PH (1988) Identification of unusual replacement of methionine by norleucine in recombinant interleukin-2 produced by *E. coli*. *Biochem Biophys Res Commun* 156:807–813.
- Apostol I, Levine J, Lippincott J, Leach J, Hess E, Glascock CB, Weickert MJ, Blackmore R (1997) Incorporation of norvaline at leucine positions in recombinant human hemoglobin expressed in *Escherichia coli*. *J Biol Chem* 272:28980–28988.
- McNulty DE, Claffee BA, Huddleston MJ, Kane JF (2003) Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Expr Purif* 27:365–374.
- Calderone TL, Stevens RD, Oas TG (1996) High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*. *J Mol Biol* 262:407–412.
- Aguirre B, Costas M, Cabrera N, Mendoza-Hernández G, Helseth DL, Jr, Fernández P, de Gómez-Puyou MT, Pérez-Montfort R, Torres-Larios A, Gómez-Puyou A (2011) A ribosomal misincorporation of Lys for Arg in human triosephosphate isomerase expressed in *Escherichia coli* gives rise to two protein populations. *PLoS ONE* 6:e21035.
- Lu HS, Fausset PR, Sotos LS, Clogston CL, Rohde MF, Stoney KS, Herman AC (1993) Isolation and characterization of three recombinant human granulocyte colony stimulating factor His→Gln isoforms produced in *Escherichia coli*. *Protein Expr Purif* 4:465–472.
- Masuda M, Dohmae N, Nonaka T, Oikawa T, Hisanaga SI, Goedert M, Hasegawa M (2006) Cysteine misincorporation in bacterially expressed human α -synuclein. *FEBS Lett* 580:1775–1779.
- Harris RJ, Murnane AA, Utter SL, Wagner KL, Cox ET, Polastri GD, Helder JC, Sliwkowski MB (1993) Assessing genetic heterogeneity in production cell lines: detection by peptide mapping of a low level Tyr to Gln sequence variant in a recombinant antibody. *Nat Biotechnol* 11:1293–1297.
- Dorai H, Sauerwald T, Campbell A, Kyung YS, Goldstein J, Magill A, Lewis MJ, Tang Q, Jan D, Ganguly S, Moore G (2007) Investigation of product microheterogeneity: a case study in rapid detection of mutation in mammalian production cell lines. *Bioprocess Int* 9:66–72.
- Wen D, Vecchi MM, Gu S, Su L, Dolnikova J, Huang Y, Foley SF, Garber E, Pederson N, 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* 284:32686–32694.
- Yu XC, Borisov OV, Alvarez M, Michels DA, Wang YJ, Ling V (2009) Identification of codon-specific serine to asparagine mistranslation in recombinant monoclonal antibodies by high-resolution mass spectrometry. *Anal Chem* 81:9282–9290.
- Yang Y, Strahan A, Li C, Shen A, Liu H, Ouyang J, Katta V, Francissen K, Zhang B (2010) Detecting low level sequence variants in recombinant monoclonal antibodies. *mAbs* 2:285–298.
- Que AH, Zhang B, Yang Y, Zhang J, Derfus G, Amanuel (2010) A sequence variant analysis using peptide mapping by LC-MS/MS. *Bioprocess Int* 8:52–60.
- Tsaprailis G, Nair H, Somogyi A, Wysocki VH, Zhong W, Futrell JH, Summerfield SG, Gaskell SJ (1999) Influence of secondary structure on the fragmentation of protonated peptides. *J Am Chem Soc* 121:5142–5154.
- Tsaprailis G, Somogyi A, Nikolaev EN, Wysocki VH (2000) Refining the model for selective cleavage at acidic residues in arginine-containing protonated peptides. *Int J Mass Spectrom* 195/196:467–479.
- Gu C, Tsaprailis G, Brei L, Wysocki VH (2000) Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides. *Anal Chem* 72:5804–5813.
- Hirel PH, Schmitter JM, Dessen P, Fayat G, Blanquet S (1989) Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci USA* 86:8247–8251.
- Martin SL, Vrhovski B, Weiss AS (1995) Total synthesis and expression in *Escherichia coli* of a gene encoding human tropoelastin. *Gene* 154:159–166.
- Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. *Curr Opin Biotechnol* 10:411–421.
- Seetharam R, Heeren RA, Wong EY, Braford SR, Klein BK, Aykent S, Kotts CE, Mathis KJ, Bishop BF, Jennings MJ, Smith CE, Siegel NR (1988) Mistranslation in IGF-1 during over-expression of the protein in *Escherichia coli* using a synthetic gene containing low frequency codons. *Biochem Biophys Res Commun* 155:518–523.