

Identification of a single base-pair mutation of TAA (Stop codon) → GAA (Glu) that causes light chain extension in a CHO cell derived IgG1

Taylor Zhang,^{1,*} Yungfu Huang,¹ Scott Chamberlain,¹ Tony Romeo,¹ Judith Zhu-Shimoni,¹ Daniel Hewitt,¹ Mary Zhu,¹ Viswanatham Katta,¹ Brad Mauger² and Yung-Hsiang Kao¹

¹Protein Analytical Chemistry; Genentech; South San Francisco, CA USA; ²Early Stage Cell Culture Development; Genentech; South San Francisco, CA USA

Keywords: stop codon, sequence variants, mutation, detection, IgG1

Abbreviations: CHO, Chinese hamster ovary; LC-MS/MS, liquid chromatography with tandem mass spectrometry detection; ETS, error-tolerant search; mAb, monoclonal antibody; DTT, dithiothreitol; TFA, trifluoroacetic acid; ESI-TOF, Electrospray-ionization Time-of-Flight; MALDI, Matrix-assisted laser desorption/ionization

We describe here the identification of a stop codon TAA (Stop) → GAA (Glu) = Stop221E mutation on the light chain of a recombinant IgG1 antibody expressed in a Chinese hamster ovary (CHO) cell line. The extended light chain variants, which were caused by translation beyond the mutated stop codon to the next alternative in-frame stop codon, were observed by mass spectra analysis. The abnormal peptide peaks present in tryptic and chymotryptic LC-MS peptide mapping were confirmed by N-terminal sequencing as C-terminal light chain extension peptides. Furthermore, LC-MS/MS of Glu-C peptide mapping confirmed the stop221E mutation, which is consistent with a single base-pair mutation in TAA (stop codon) to GAA (Glu). The light chain variants were approximately 13.6% of wild type light chain as estimated by RP-HPLC analysis. DNA sequencing techniques determined a single base pair stop codon mutation, instead of a stop codon read-through, as the cause of this light chain extension. To our knowledge, the stop codon mutation has not been reported for IgGs expressed in CHO cells. These results demonstrate orthogonal techniques should be implemented to characterize recombinant proteins and select appropriate cell lines for production of therapeutic proteins because modifications could occur at unexpected locations.

Introduction

Sequence variants, defined as protein variants with unintended amino acid incorporation, are potential critical quality attributes for monoclonal antibodies (mAbs) produced for therapeutic use.¹ Clone screening to select cell lines free of sequence variants is a general practice to support cell culture development. As a result of the development of more sensitive and systematic methods, several sequence variants have been described.²⁻⁴ The process used to detect sequence variants is based on multiple peptide mappings with complementary enzymes, which generates peptides suitable for MS/MS analysis and generally covers 100% of the protein sequence. The peptides are then separated by RP-HPLC coupled with an accurate, high-resolution mass spectrometer. The acquired mass spectra data are subjected to an automated software search, such as Mascot Error Tolerant Search (ETS), against a known protein sequence database. This technique is very sensitive, as sequence variants with serine to asparagine changes due to codon specific mistranslation have been detected at levels as low as 0.1–0.2%.³ Furthermore, the

approach can reproducibly detect 0.5% variants as shown by co-mixing two mAbs.²

Sequence variant identification benefits from an array of analytical methods because the protein variants could be detected as abnormal peaks by other methods. Recently, an S441R mutation was identified in another mAb at 0.5% as a result of an investigation of a new peak in the charge assay that was not revealed in the initial ETS.⁵ In fact, the ETS methodology relies heavily on the correct protein database and might miss a mutation related to stop codons. Even with an ETS performed on a translated DNA database, some spectra are difficult to interpret by software or manually due to limited fragmentation of the peptides. In such a scenario, traditional orthogonal techniques such as N-terminal sequencing will provide invaluable information to identify the variants. Sequence variants can be generated by DNA/RNA mutation or amino acid misincorporation and pinpointing the exact mechanism will enable the design of strategies to avoid sequence variants. Understanding the cause of the variants calls for analyses using nucleic acid based technologies such as PCR or DNA sequencing to complement protein related techniques.^{6,7}

*Correspondence to: Taylor Zhang; Email: yonghuaz@gene.com
Submitted: 07/26/12; Accepted: 09/13/12
<http://dx.doi.org/10.4161/mabs.22232>

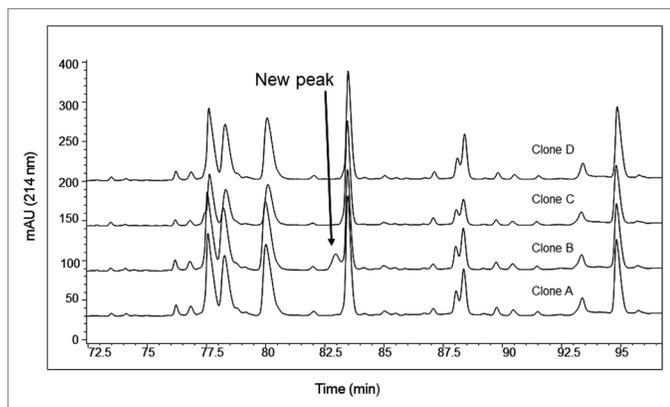


Figure 1. Overlay of tryptic peptides maps of IgG1 derived from four clones. The new peak at 83 min in clone B is absence from other clones.

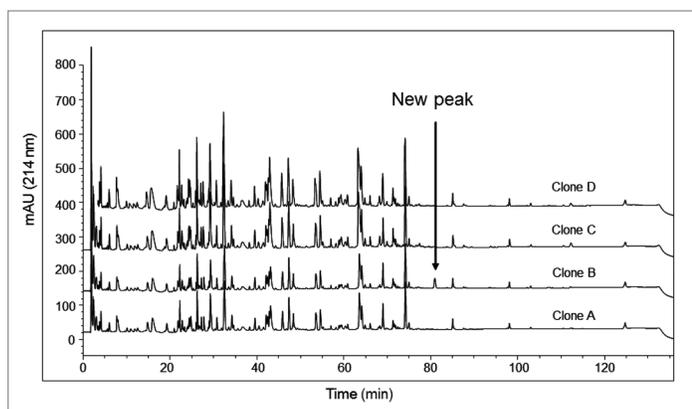


Figure 2. Overlay of chymotryptic maps of IgG1 derived from four clones. The new peak in 80 min is visible in clone B and is absent from other clones.

Sequence variants can occur at different amino acids at random locations.⁸ The stop codons (UAA, UGA and UAG) are important in protein translation because they are recognized by the release factors and signal the termination of the translation.⁹ With read-through, a process in which a stop codon is not recognized, the protein synthesis will stop at the next in-frame stop codon in the expression vector and generate a protein with an extension.¹⁰ Such mistranslation, often referred to as nonsense read-through, has been reported for some *E. coli* derived proteins. For example, Lu et al. observed a UGA stop codon mistranslation as tryptophan in recombinant platelet-derived growth factor expressed in *E. coli*.¹¹ Meng also reported a UGA stop codon read-through during production of human neurotrophin.¹² Avoiding use of UGA as a stop codon is recommended because of the high occurrence of read-through in *E. coli*.¹¹ Recombinant murine interleukin-6 expressed in *E. coli* has contained a C-terminal pentapeptide extension, where the stop codon was mistranslated to glutamine.¹³ The misreading of stop codons in eukaryotes has been reviewed,¹⁴ and the error rate due to stop codon read-through has been estimated at 0.3% in yeast.¹⁵

Interestingly, it was reported that 0.2% of mutations among missense and nonsense mutations in the Human Gene Mutation Database (HGMD) are related to stop codons.¹⁶ For example, a point mutation at the stop codon of BRI gene caused a longer open reading frame and generated an extended protein.¹⁷ Protein extension has also been associated with stop codon mutation for apolipoprotein AII (apoAII), which has a 21-residue peptide extension on the carboxyl terminus.¹⁸ Variants with C-terminal extensions for antibodies expressed in CHO cells have not been reported, yet such modifications should be as likely as mutations at any other site. A stop codon-related mutation will generate a C-terminal extension in the protein and needs to be well characterized for biopharmaceutical development because it can affect the function of the expressed proteins.

Here, we report the observation of IgG1 variants with light chain extensions, in addition to the expected 220 amino-acid light chain. The variants were only expressed in one (clone B) of the four clones evaluated during clone screening. With a combination of different enzymatic peptide maps and LC-MS and LC-MS/MS techniques, N-terminal sequencing, RP-HPLC and nucleic acid based technologies, we confirmed that such variants were generated because of a single base-pair mutation of TAA (Stop codon) to GAA (Glu), enabling us to select appropriate clones for clinical therapeutic process development.

Results

Tryptic and chymotryptic peptide mapping revealed extra peptides for clone B. The tryptic peptide mapping chromatographic profiles (Fig. 1) for the four clones are consistent, except there is a new peak eluting at 83 min for clone B. The molecular weight for the new tryptic peptide is 2101.96 Da and it is not an expected tryptic peptide mass according to the antibody sequence. The MS/MS for the peptide had limited daughter ions (data not shown), from which it was difficult to derive the parent peptide sequence. Similarly, the chymotryptic mapping chromatograms (Fig. 2) also revealed a new peak in the clone B antibody digest eluting around 80 min. The mass for the extra chymotryptic peptide is 3546.58 Da, also not an expected chymotryptic peptide mass. Both of the new peaks account for more than 0.1% of the total peak area in their respective maps, indicating the IgG1 produced from clone B has an unknown variant. Mascot search did not provide the identities of these two new peptides. Again manual interpretation of the MS/MS was not successful in identifying the peptide because the peptides are relatively big and fragments present in MS/MS are limited. MS/MS sequencing of these peptides by MALDI yielded no additional information to electro spray ionization MS/MS. Overall, the sequence coverage from these two peptide maps was 100% and there was no other sequence variant found for these four clones.

N-terminal sequencing of the new tryptic and chymotryptic peptides and Glu-C enzymatic peptide mapping confirmed the light chain extension. The two new peptides were fraction collected and subject to Edman N-terminal sequencing. The new peptide collected from the tryptic map was identified as a peptide with the N-terminal as TVAPTECSEAWP (data not shown).

The first eight amino acids are the correct light chain C-terminal sequence indicating protein translation beyond the stop codon. The chymotryptic peptide has the sequence S(C)QV(T)(H)(E)...where parenthesis denotes a tentative sequence call from the N-terminal sequencing data. The results indicate that the extra chymotryptic peptide is also C-terminal peptide (SCQVTH EGS TVE KTV APT ECS) related, though the sequence does not read far enough to show the extension because of the limited sample used. By inspecting the vector sequence near the light chain stop codon, the predicted protein would have the additional sequence XAWPP...continuing until the next in-frame stop codon, 54 nucleotides later. The X here denotes the unknown amino acid resulting from stop codon mutation or mistranslated amino acid. The N-terminal sequence indicates X is glutamic acid, which is consistent with a single nucleotide mutation from stop codon TAA to GAA as shown in **Figure 3**. The identities of the two new peptides can therefore be deduced as C-terminal light chain peptide extensions. Compared with the correct C-terminal sequence for tryptic peptide TVA PTE CS (M+H: 865.36 Da) and chymotryptic peptide SCQ VTH EGS TVE KTV APT ECS (M+H: 2308.99 Da) and the new tryptic peak (M+H: 2102.96 Da) and chymotryptic peak (M+H: 3546.58Da) have the correct C-terminal light chain sequence plus the same extension of 1237.59Da, consistent with the calculated mass of the C-terminal extension EAW PPW PNL F. The theoretical tryptic peptide with protein translation beyond the stop codon through the next stop codon is TVA PTE CSE AWP PWP NLF IAA YNG YK. This peptide eluted at around 91 min because the peptide is more hydrophobic (data not shown).

To further confirm the stop codon mutation to glutamic acid indicated by N-terminal sequencing, we digested protein product from clone B and from a control clone using the enzyme Glu-C. The Glu-C mapping shows a new peak at 30 min for clone B (**Fig. 4**) compared with the control. The MS data for the related C-terminal peptides are also captured in **Table 1**. The new peak at 135 min is related to the extension peptide and was marked in the chromatogram. The corresponding MS/MS (**Fig. 5**) shows the new peptide at 30 min has the y ion series shifted 129 Da compared with the wild type at 29 min; the results are consistent with glutamic acid as the amino acid in the place of stop codon. From the UV peak areas of the wild type peptide KTV APT TEC S and mutant peptide KTV APT TEC SE, we estimate 15% of the C-terminal peptide has the extension, assuming both peptides have the same UV absorptivity.

Knowing that there were light chain extension variants in clone B, we updated the light chain with 18 amino acid extended sequence in the protein database and performed Mascot search again using the peptide mapping data described above. The Mascot software confirmed three peptides with the extended sequence were present in clone B and were absent in the control clone. These peptides are SCQ VTH EGS TVE KTV APT ECS EAW PPW PNL for the chymotryptic map, KTVAPTECSE for the Glu-C map and IAA YNG YK for the tryptic map. The high Mascot score for these peptides (64, 79 and 37 respectively) confirmed these assignments.

```

ACCGTCGCCCGACAGAGTGTCTGAAGCTTGGCCGCCATGGCCC
T V A P T E C S E A W P P W P
AACTTGTTTATTGCAGCTTATAATGGTTACAAATAA
N L F I A A Y N G Y K

```

Figure 3. The DNA sequence near the stop codon and corresponding translated amino acid sequence.

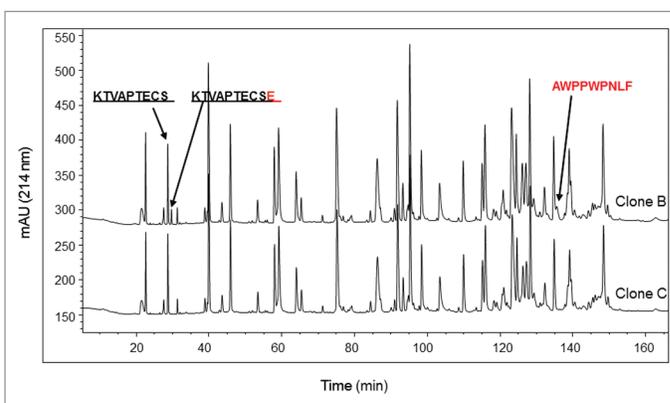


Figure 4. Comparison of Glu-C peptide mapping for clone B and clone C, as a control. The new peak shown at 20 min was the C-terminal light chain peptide with an extra glutamic acid, marked in red. Another extension peak around 130 min was also identified.

Table 1. Peptides observed in Glu-C mapping indicate the clone B has glutamic acid at the position of stop codon and the peptides sequences belong to the vector

Sequence of C-terminal Peptides in Glu-C mapping	KTVAPTECS	KTVAPTECSE	AWPPWPNLF
Theoretical MW (M+H)	993.45	1122.50	1127.57
Observed MW in control (M+H)	993.45	Not observed	Not observed
Observed MW in clone B (M+H)	993.45	1122.50	1127.57

Mass spectrometry and RP-HPLC of the mutated variants at the protein level. The mAb was reduced by DTT and analyzed by LC-MS. Deconvolution of the mass data showed the presence of three major light chain species, the expected light chain mass of 23502 Da (1–220 AA, **Fig. 6A**) and additional masses of 24740 Da and 25493 Da, (**Fig. 6B**). The 25493 Da species is in agreement with a 17 amino acid extension (1–237 AA) and the 24740 Da species is in agreement with ten amino acid extension (1–230 AA) expected based on translation of the vector DNA sequence with the stop codon mutation to glutamic acid.

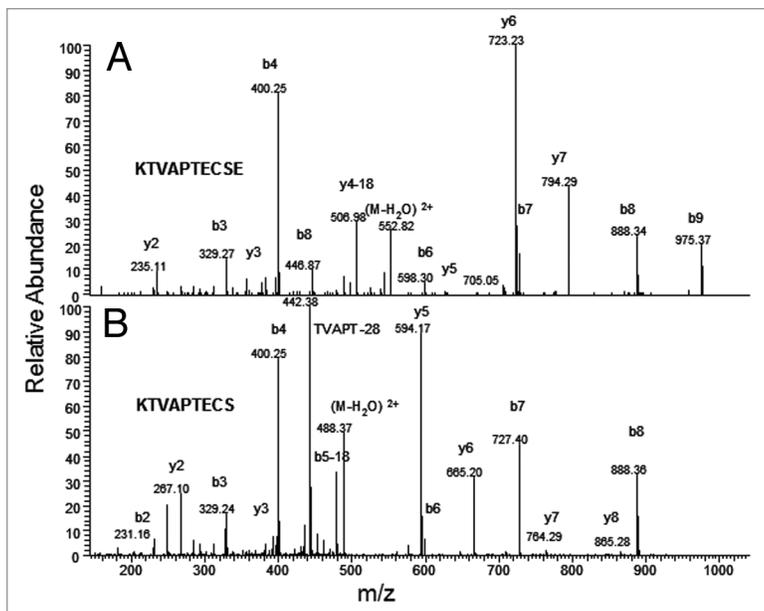


Figure 5. Tandem mass spectra of $(M+2H) 2+$ precursor ions at (A) m/z of 561.75 for peptide KTVAPTECSE; (B) m/z of 497.23 for KTVAPTECS. Fragments ions support the conclusion that the light chain stop codon changed to glutamic acid.

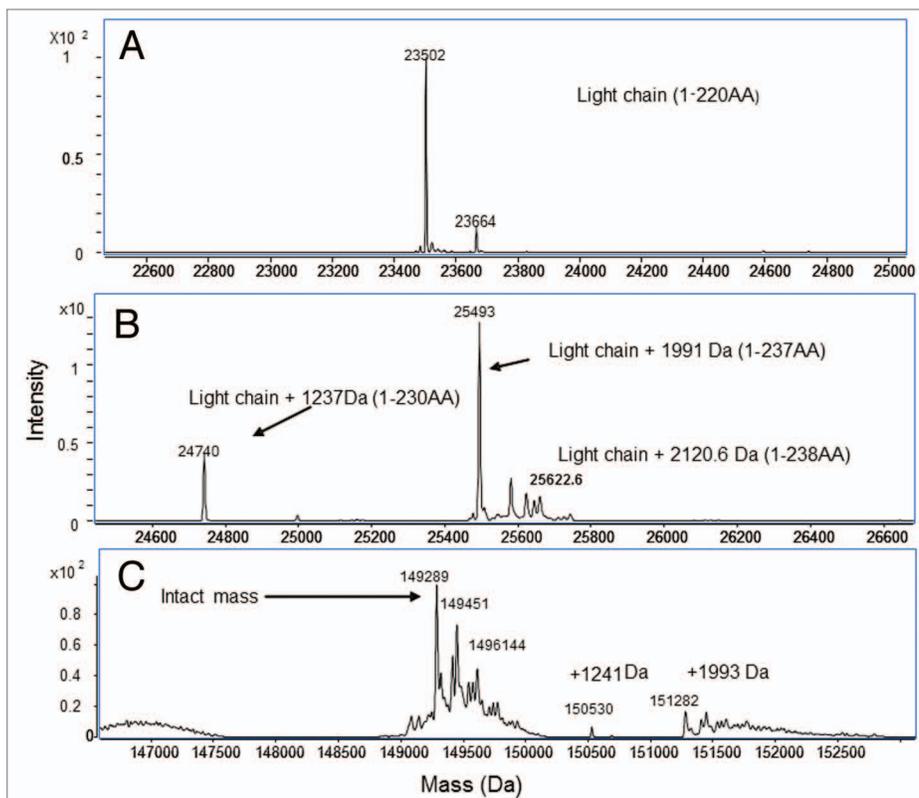


Figure 6. Deconvoluted mass spectra of the reduced antibody light chain (A and B) and the intact antibody (C). The theoretical light chain (1–220AA) is 23502 Da (A) and other light chain species are also observed (B), which is consistent with light chain extension.

The minor mass of the 25622 Da (1–238 AA) species is the result of residual C-terminal lysine on the modified light chain not being completely removed.¹⁹ The exact cause of light chain ending with phenylalanine (1–230 AA) is unclear, one explanation being CHO cell protease activity.²⁰

Mass analysis of the intact mAbs (Fig. 6C) shows the minor species as 151282Da compared with the major G0 species of 149289 Da and related glycan variants (149451 G0/G1, 149613 Da (G1/G1 or G2) and the C-terminal lysine variants (149415, 149577Da). The difference (1993Da) is in agreement with that of the extension generated by a stop codon mutation to glutamic acid, then translation through to the next in-frame stop codon, assuming the terminal lysine is cleaved from the light chain extension. In Figure 5C, the species of 150,530 Da, which is consistent with intact mAbs with one extended light chain (1–230AA), is also minor.

The IgG1s produced from clone B and control clone were reduced and analyzed by RP-HPLC with UV detection (Fig. 7). Consistent with the peptide mapping results, the new peak in the clone B suggests the mutated light chain is more hydrophobic compared with the normal light chain. The light chain variant peaks comprise about 13.6% of the total light chain by UV absorption, which is consistent with the results from the Glu-C map.

DNA sequencing confirmed the single base-pair mutation in TAA (stop codon)-GAA (Glu). cDNA generated from clone B and the control clone were submitted to DNA sequencing from both ends of the light chain to allow for reads flanking the base pair in the stop codon (8485 bp) in both directions. The DNA sequencing electropherogram (Fig. 8A) for clone B showed an additional nucleobase guanine (G) signal that did not surpass the nucleobase thymine (T) signal, but was above baseline noise level, which suggested that a fraction of the mRNA molecules corresponding to the cDNA were mutated from a uracil (U) to a G at the stop codon position in clone B. Thus, the stop codon (TAA) at this location was lost and turned into the codon coding for the amino acid residue glutamic acid (GAA), and protein encoding continued on until the next stop codon. The electropherogram (Fig. 8B) shows that DNA sequence from the control clone has no such mutation. The DNA sequencing results were confirmed with

repeated reactions, and all indicated the mutation in the clone B sequences but not in the control. In addition, this finding is also in agreement with the results from mass spectra data, where a longer coding sequence was obtained from clone B, but not from the control clone.

Discussion

With the wide application of accurate and sensitive mass spectrometry, it is possible to detect very small amounts of peptides with alternative sequences down to the 0.01% level—if the sequence variant is known.³ Nevertheless, peptide mapping with UV detection is still a simple yet powerful method to detect differences in mAbs produced from different clones when the mutation rate is 1 to 5% or above.² By UV peptide mapping with complementary enzymes, we detected the extra peaks from the mutated clone in both maps. We could not identify the peaks because of limited fragmentation due to proline-directed fragmentation,²¹ but the mutation was at a level much higher than 1%, which allowed us to collect the unknown peaks from tryptic and chymotryptic mapping. The N-terminal sequencing of these collected peptides confirmed the two peptides as the C-terminal light chain extension peptides. Furthermore, the Glu-C map confirmed the stop codon was converted to glutamic acid. The overall approach emphasized N-terminal sequencing as a traditional but critical tool for deciphering the mutation, while the ETS approach was limited to good quality MS/MS data and the correct database. Using ETS, the MS/MS was searched against the custom in-house database generated from ~150 protein sequences.⁴ When the mutated peptide sequence is not part of the database, ETS becomes ineffective. After the light chain extension sequence was introduced into the database, Mascot software identified the peptides with good confidence. The finding of a stop codon mutation in our study demonstrates the benefit of establishing the Mascot database with DNA sequence along with the vector sequence. This will help detect a stop codon mutation or N-terminal sequence variants caused by an alternative signal peptidase sequence.²²

Without DNA sequencing data, it would be impossible to differentiate a stop codon mutation in clone B from a translation error because stop codon read-through could also produce proteins with extensions. It is unclear whether the mutant in clone B originated from a contaminant clone or arose as clone B grew. Theoretically, it is possible that the variant in clone B could result from a transcription error because mRNA was extracted, and cDNA generated and amplified was used for the DNA

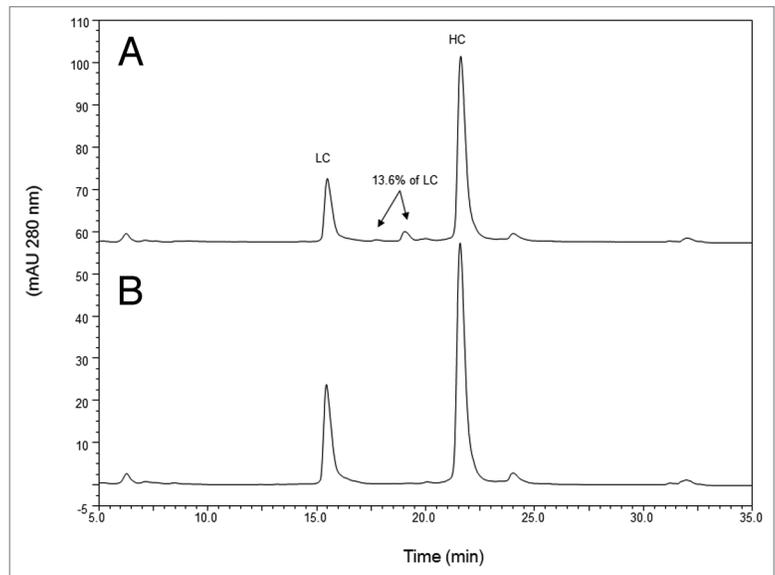


Figure 7. RP-HPLC for the reduced clone B (A) and the control clone (B) showing the mutated light chain is present at around 13.6% of the total light chain.

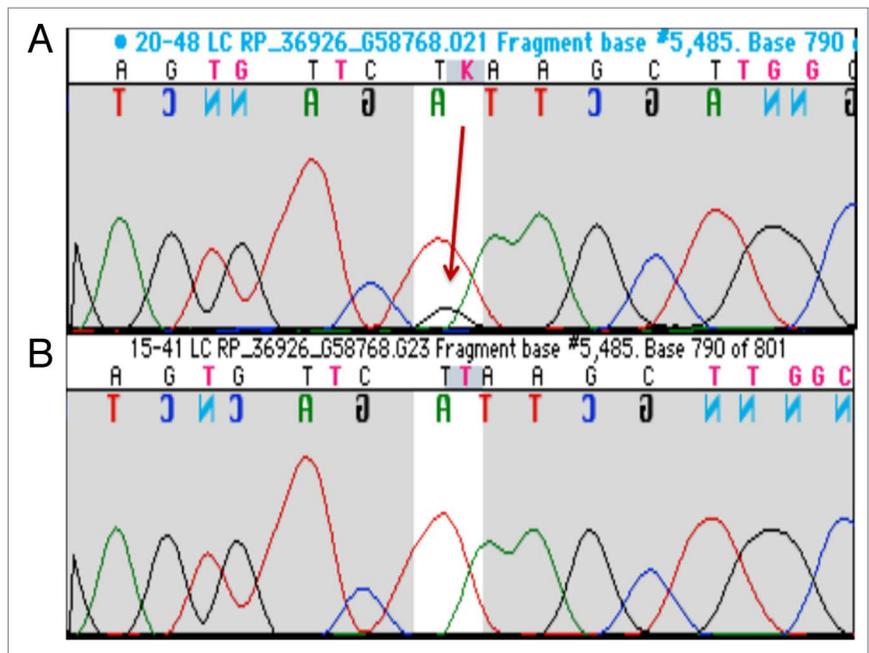


Figure 8. Detection of stop codon (TAA) change to glutamic acid codon (GAA) in mRNA encoding the light chain by DNA sequencing. (A) Small amount of mutated G (shaded in the trace) is detected along with the wild type nucleotide T. (B) The control has no mutation at the highlighted position, only the expected nucleotide T.

sequencing; however, transcription errors from DNA occur with low frequency.²³ Because the mutation variants were 15% of the total protein by Glu-C peptide mapping analysis, and the implied nucleic acid mutation by DNA sequencing was at the similar level, it was reasonable to expect that the mutation could be detected and confirmed by Sanger sequencing technology,

which has a well-documented limit of detection (15–20%).^{24,25} Based on the cDNA sequencing data, we selected other clones for further development instead of developing a feeding strategy to eliminate the variants, as demonstrated in another study.²⁶

To our knowledge, a stop codon mutation on IgG produced by CHO cells has not been reported. This mutation caused an extension of the protein chain. It is also interesting to consider whether codons coding for other amino acids could be mutated to stop codons, which would generate a truncated protein and might be mis-identified as a fragment. In fact, such a mutation has been widely studied in human genetics; incomplete non-functional proteins resulting from such mutations have been associated with the human diseases.²⁷ We should be cautious about the occurrence of variants resulting from nonsense to sense mutations and focus on the possibility of truncated variants resulting from sense to nonsense mutations because eight triplet codons coding for several amino acids can be converted to the amber stop codon by a single base mutation.²⁸ These truncated proteins might be removed by Protein A purification; nevertheless, they are statistically more likely than the protein extension observed here. Unlike the typical C-terminal variants such as lysine,¹⁹ proline amidation,²⁹ or N-terminal signal sequence variants,³⁰ the light chain mutation variant reported here are even longer, up to 18 additional amino acids. This mutation might increase the risk of immunogenicity because the additional sequence is not part of light chain constant region. Variants with a light chain C-terminal extension might have similar bioactivity to the native protein since the binding is controlled through the variable region in the Fab and the mechanism of action for the intended drug is to bind the antigen, but we did not study the biological activity of the variant. Although clone B had higher titer than the other clones, we eliminated it from further development because of the potential immunogenicity risk. We should note that our observations are valid for stable transfection, but not for transient transfection. In stable transfection, the DNA encoding the protein of interest is introduced into the genome itself by transfection, and then the clones are subjected to selection and amplification using selective compounds such as methotrexate (MTX), which is a known mutagen.⁶ In contrast, transient transfection has the DNA introduced at the beginning of the cell culture run and this DNA is not incorporated into the genome, and the above selection process does not apply.

In conclusion, we identified by multiple analytical techniques an unusual Stop221E mutation in the light chain for an IgG1 produced in CHO cell culture. Our finding emphasizes the importance of orthogonal analytical techniques to support biopharmaceutical development and to characterize unintended modifications in therapeutic proteins.

Materials and Methods

Materials. The IgG1 mAbs were manufactured at Genentech. The antibodies were expressed in four different CHO cell clones listed as A-D and purified by protein A affinity chromatography. Trypsin was purchased from Promega. Chymotrypsin was

purchased from Roche Applied Science. Trifluoroacetic acid was purchased from Pierce. HPLC-grade water and acetonitrile were purchased from Burdick and Jackson (Honeywell). Dithiothreitol (DTT) and iodoacetic acid (IAA) were purchased from Sigma.

RCM peptide mapping with Trypsin, Chymotrypsin, Glu-C digestion. Antibody samples were denatured, reduced and alkylated by mixing 50 μ L of mAb sample (20 mg/mL) with 10 μ L DTT (1 M in Milli-Q water) and 950 μ L of the reduction and S-carboxymethylation (RCM) buffer (6 M guanidine hydrochloride, 360 mM Tris and 2 mM EDTA, pH 8.6), and then mixed with 14 μ L IAA (3 M in 1 N NaOH). The reduced and carboxymethylated samples were digested with different enzymes following the typical procedure.²

The online LC/ESI-MS/MS analyses of the tryptic digests were performed using an Agilent Capillary 1200 HPLC system coupled with a Thermo Fisher LTQ Orbitrap mass spectrometer equipped with an electrospray ionization source. The antibody digest was injected onto a Waters C4 column with column temperature set at 55°C. The mobile phases and gradient were similar to the published procedure.²

N-terminal sequencing. Peptides for N-terminal sequencing were partially dried by vacuum centrifugation. Aliquots were loaded onto Polybrene™ -precycled glass fiber filters and loaded into the reaction cartridge of an Applied Biosystems Procise™ 494HT sequencer. Sequencing was performed using liquid TFA cleavage and PTH-amino acids were identified by comparison with standards using 269 nm absorbance.

Nano chip LC/ESI-TOF MS. Mass analyses of antibodies were performed on an Agilent 6210 electrospray-ionization time-of-flight (ESI-TOF) mass spectrometer (Agilent). Agilent Nano-HPLC Chip (equivalent to Zorbax 300SB-C8 analytical column of 43 mm length \times 75 μ m internal diameter, 5 μ m particle size) was used for desalting and separation using a linear gradient from 20% B to 90% B in 6 min. The solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in acetonitrile. Spectra consisting of multiple charged protein ions were deconvoluted using the Agilent MassHunter.

RP-HPLC for reduced antibody. RP-HPLC analysis was performed on an Agilent 1200 HPLC system (Santa Clara, CA, USA) equipped with a binary gradient pump, autosampler, temperature-controlled column compartment, and a diode array detector. The system included a Pursuit 3 diphenyl reversed phase column (150 \times 4.6 mm, 3 μ m, Varian, Lake Forest, CA, USA) that was run at 75°C. The separation was monitored using absorbance at 280 nm. The mobile phase consisted of a gradient of 0.1% TFA in water (mobile phase A) and 0.1% TFA in ACN (mobile phase B). The 46 min method began with a three-minute isocratic hold at 32% mobile phase B, followed by a 26 min linear gradient to 45% mobile phase B. The column was washed at 95% mobile phase B for six minutes and equilibrated at 32% mobile phase B for 10 min.

DNA sequencing to confirm the mutation. The cell cultures were centrifuged in tubes to pellet with a million cells in each pellet and frozen at -70°C. Samples were then processed using the total nucleic acid extraction kit (Qiagen). The RNA was frozen at -70°C until used for RT-PCR by the Onestep RT-PCR kit

(Qiagen). The primers used in the RT-PCR were made to span introns so that genomic DNA would not be amplified. The primers used bind outside of the light chain coding region and were also used as sequencing primers on the cDNA created. cDNA was checked for quality on the Bioanalyzer 2100 and by Nanodrop. The cDNA was then used as the template in Sanger sequencing using Big Dye® chemistry (ABI). This process was repeated and unique results were compiled in Sequencer (Gene Codes) to look for a difference in location flanking the base 5485.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank Laura Simmons and Thu Nguyen for developing the cell lines, Hongbin Liu and Chris Yu for the helpful discussion of sequence variant analysis and John Stults, Reed Harris and Nathan McKnight for their support. We also appreciate the comments from anonymous reviewers.

References

- Harris RJ, Murnane AA, Utter SL, Wagner KL, Cox ET, Polastri GD, et al. Assessing genetic heterogeneity in production cell lines: detection by peptide mapping of a low level Tyr to Gln sequence variant in a recombinant antibody. *Biotechnology (N Y)* 1993; 11:1293-7; PMID:7764191
- Yang Y, Strahan A, Li C, Shen A, Liu H, Ouyang J, et al. Detecting low level sequence variants in recombinant monoclonal antibodies. *MAbs* 2010; 2:285-98; PMID:20400866; <http://dx.doi.org/10.4161/mabs.2.3.11718>
- Yu XC, Borisov OV, Alvarez M, Michels DA, Wang YJ, Ling V. Identification of codon-specific serine to asparagine mistranslation in recombinant monoclonal antibodies by high-resolution mass spectrometry. *Anal Chem* 2009; 81:9282-90; PMID:19852494; <http://dx.doi.org/10.1021/ac901541h>
- Que A, Zhang B, Yang Y, Zhang J, Derfus G, Amanullah A. Sequence variant analysis using peptide mapping by LC-MS/MS. *Bioprocess Int* 2010; 8:52-60
- Ren D, Zhang J, Pritchett R, Liu H, Kyauk J, Luo J, et al. Detection and identification of a serine to arginine sequence variant in a therapeutic monoclonal antibody. *J Chromatogr B Analyt Technol Biomed Life Sci* 2011; 879:2877-84; PMID:21900054; <http://dx.doi.org/10.1016/j.jchromb.2011.08.015>
- Guo D, Gao A, Michels DA, Feeney L, Eng M, Chan B, et al. Mechanisms of unintended amino acid sequence changes in recombinant monoclonal antibodies expressed in Chinese Hamster Ovary (CHO) cells. *Biotechnol Bioeng* 2010; 107:163-71; PMID:20506532; <http://dx.doi.org/10.1002/bit.22780>
- Wen D, Vecchi MM, Gu S, Su L, Dolnikova J, Huang Y-M, et al. Discovery and investigation of misincorporation of serine at asparagine positions in recombinant proteins expressed in Chinese hamster ovary cells. *J Biol Chem* 2009; 284:32686-94; PMID:19783658; <http://dx.doi.org/10.1074/jbc.M109.059360>
- Santos MAS, Tuite MF. New insights into mRNA decoding--implications for heterologous protein synthesis. *Trends Biotechnol* 1993; 11:500-5; PMID:7764419; [http://dx.doi.org/10.1016/0167-7799\(93\)90028-8](http://dx.doi.org/10.1016/0167-7799(93)90028-8)
- Bertram G, Innes S, Minella O, Richardson JP, Stansfield I. Endless possibilities: translation termination and stop codon recognition. *Microbiology* 2001; 147:255-69; PMID:11158343
- Namy O, Rousset J-P, Atkins JF, Gesteland RF. Specification of standard amino acids by stop codon-reading: expansion of decoding rules enriches gene expression. *Springer New York*:79-100
- Lu KV, Rohde MF, Thomason AR, Kenney WC, Lu HS. Mistranslation of a TGA termination codon as tryptophan in recombinant platelet-derived growth factor expressed in *Escherichia coli*. *Biochem J* 1995; 309:411-7; PMID:7626004
- Meng SY, Hui JO, Haniu M, Tsai LB. Analysis of translational termination of recombinant human methionyl-neurotrophin 3 in *Escherichia coli*. *Biochem Biophys Res Commun* 1995; 211:40-8; PMID:7779107; <http://dx.doi.org/10.1006/bbrc.1995.1775>
- Hammacher A, Reid GE, Moritz RL, Simpson RJ. Identification and characterization of recombinant murine interleukin-6 with a C-terminal pentapeptide extension using capillary reversed phase HPLC-MS and edman degradation. *Biomed Chromatogr* 1997; 11:337-42; PMID:9413611; [http://dx.doi.org/10.1002/\(SICI\)1099-0801\(199711\)11:6<337::AID-BMC687>3.0.CO;2-E](http://dx.doi.org/10.1002/(SICI)1099-0801(199711)11:6<337::AID-BMC687>3.0.CO;2-E)
- Cassan M, Rousset J-P. UAG readthrough in mammalian cells: effect of upstream and downstream stop codon contexts reveal different signals. *BMC Mol Biol* 2001; 2:3; PMID:11242562; <http://dx.doi.org/10.1186/1471-2199-2-3>
- Namy O, Hatin I, Rousset JP. Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO Rep* 2001; 2:787-93; PMID:11520858; <http://dx.doi.org/10.1093/embo-reports/kve176>
- Hamby SE, Thomas NS, Cooper DN, Chuzhanova N. A meta-analysis of single base-pair substitutions in translational termination codons ('nonstop' mutations) that cause human inherited disease. *Hum Genomics* 2011; 5:241-64; PMID:21712188; <http://dx.doi.org/10.1186/1479-7364-5-4-241>
- Vidal R, Frangione B, Rostagno A, Mead S, Révész T, Plant G, et al. A stop-codon mutation in the BRI gene associated with familial British dementia. *Nature* 1999; 399:776-81; PMID:10391242; <http://dx.doi.org/10.1038/21637>
- Benson MD, Liepnieks JJ, Yazaki M, Yamashita T, Hamidi Asl K, Guenther B, et al. A new human hereditary amyloidosis: the result of a stop-codon mutation in the apolipoprotein AII gene. *Genomics* 2001; 72:272-7; PMID:11401442; <http://dx.doi.org/10.1006/geno.2000.6499>
- Harris RJ. Processing of C-terminal lysine and arginine residues of proteins isolated from mammalian cell culture. *J Chromatogr A* 1995; 705:129-34; PMID:7620566; [http://dx.doi.org/10.1016/0021-9673\(94\)01255-D](http://dx.doi.org/10.1016/0021-9673(94)01255-D)
- Gao SX, Zhang Y, Stansberry-Perkins K, Buko A, Bai S, Nguyen V, et al. Fragmentation of a highly purified monoclonal antibody attributed to residual CHO cell protease activity. *Biotechnol Bioeng* 2011; 108:977-82; PMID:21404269; <http://dx.doi.org/10.1002/bit.22982>
- Breci LA, Tabb DL, Yates JR 3rd, Wysocki VH. Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Anal Chem* 2003; 75:1963-71; PMID:12720328; <http://dx.doi.org/10.1021/ac2026359i>
- Lyubarskaya Y, Houde D, Woodard J, Murphy D, Mhatre R. Analysis of recombinant monoclonal antibody isoforms by electrospray ionization mass spectrometry as a strategy for streamlining characterization of recombinant monoclonal antibody charge heterogeneity. *Anal Biochem* 2006; 348:24-39; PMID:16289440; <http://dx.doi.org/10.1016/j.ab.2005.10.003>
- Rosenberger RF, Foskett G. An estimate of the frequency of in vivo transcriptional errors at a nonsense codon in *Escherichia coli*. *Mol Gen Genet* 1981; 183:561-3; PMID:7038382; <http://dx.doi.org/10.1007/BF00268784>
- Vogelstein B, Kinzler KW. Digital PCR. *Proc Natl Acad Sci U S A* 1999; 96:9236-41; PMID:10430926; <http://dx.doi.org/10.1073/pnas.96.16.9236>
- Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, et al. Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *J Mol Diagn* 2010; 12:425-32; PMID:20431034; <http://dx.doi.org/10.2353/jmoldx.2010.090188>
- Khetan A, Huang YM, Dolnikova J, Pederson NE, Wen D, Yusuf-Makagiansar H, et al. Control of misincorporation of serine for asparagine during antibody production using CHO cells. *Biotechnol Bioeng* 2010; 107:116-23; PMID:20506364; <http://dx.doi.org/10.1002/bit.22771>
- Atkinson J, Martin R. Mutations to nonsense codons in human genetic disease: implications for gene therapy by nonsense suppressor tRNAs. *Nucleic Acids Res* 1994; 22:1327-34; PMID:8190621; <http://dx.doi.org/10.1093/nar/22.8.1327>
- Belin D. Why are suppressors of amber mutations so frequent among *Escherichia coli* K12 strains? A plausible explanation for a long-lasting puzzle. *Genetics* 2003; 165:455-6; PMID:14573460
- Johnson KA, Paisley-Flango K, Tangarone BS, Porter TJ, Rouse JC. Cation exchange-HPLC and mass spectrometry reveal C-terminal amidation of an IgG1 heavy chain. *Anal Biochem* 2007; 360:75-83; PMID:17113563; <http://dx.doi.org/10.1016/j.ab.2006.10.012>
- Kotia RB, Raghani AR. Analysis of monoclonal antibody product heterogeneity resulting from alternate cleavage sites of signal peptide. *Anal Biochem* 2010; 399:190-5; PMID:20074542; <http://dx.doi.org/10.1016/j.ab.2010.01.008>