Rapid identification of an antibody DNA construct rearrangement sequence variant by mass spectrometry

Rebecca A Scott*, Rich Rogers, Alain Balland, and Lowell J Brady

Analytical Sciences; Amgen Inc.; Seattle, WA USA

Keywords: sequence variant, mass spectrometry, C-terminal extension

During cell line development for an IgG1 antibody candidate (mAb1), a C-terminal extension was identified in 2 product candidate clones expressed in CHO-K1 cell line. The extension was initially observed as the presence of anomalous new peaks in these clones after analysis by cation exchange chromatography (CEX-HPLC) and reduced capillary electrophoresis-sodium dodecyl sulfate (rCE-SDS). Reduced mass analysis of these CHO-K1 clones revealed that a larger than expected mass was present on a sub-population of the heavy chain species, which could not be explained by any known chemical or post-translational modifications. It was suspected that this additional mass on the heavy chain was due to the presence of an additional amino acid sequence. To identify the suspected additional sequence, de novo sequencing in combination with proteomic searching was performed against translated DNA vectors for the heavy chain and light chain. Peptides unique to the clones containing the extension were identified matching short sequences (corresponding to 9 and 35 amino acids, respectively) from 2 non-coding sections of the light chain vector construct. After investigation, this extension was observed to be due to the re-arrangement of the DNA construct, with the addition of amino acids derived from the light chain vector non-translated sequence to the C-terminus of the heavy chain. This observation showed the power of proteomic mass spectrometric techniques to identify an unexpected antibody sequence variant using de novo sequencing combined with database searching, and allowed for rapid identification of the root cause for new peaks in the cation exchange and rCE-SDS assays.

Introduction

During development of recombinant biologic drug products, confirmation of the protein sequence is an important aspect of the characterization package that supports process development and regulatory filings. In recent years, improvements in mass spectrometry (MS)-based characterization assays and associated analysis software has enabled detection of sequence variants, even when present at very low levels. Sequence variants can be classified into 3 general groups: N-terminal variants due to incomplete cleavage of the signal sequence,1 single amino acid changes due to DNA mutations2,3 or protein translation errors4-7 and sequence variants due to DNA re-arrangement and mis-splicing.8 Detection of sequence variants typically employs the use of high-resolution mass spectrometry assays, with peptide map analysis followed by computer algorithm searching of the data (such as programs like Mascot9 or Mass Analyzer10). In the case of sequence variants due to N- or C-terminal extensions or DNA re-arrangement, top-down MS methods such as intact or reduced mass are usually the most appropriate methods of detection.

Of the 3 types of sequence variants, those due to DNA re-arrangement and mis-splicing may be the easiest to detect (when present at appreciable levels) because they are more likely to appear in orthogonal assays such as cation exchange chromatography (CEX), reduced capillary electrophoresis-sodium dodecyl sulfate (rCE-SDS) or size exclusion chromatography (SEC), since they often result in changes to the molecular weight (MW) or net charge of the product. However, the sequence of these species may be relatively harder to determine because DNA splicing errors can include construct or genomic sequence that may not be in frame with the normal antibody (or protein) sequence. In these cases, extensive biochemical characterization, including
isolation of the sequence variant and sequencing by proteolysis combined with N-terminal sequencing, has been the method of choice to determine the new protein sequence.

Recently, Zhang et al. published a report on the observation of a sequence variant of a monoclonal antibody (mAb) due to a mutation in the stop codon of heavy chain (HC), resulting in the C-terminal addition of several amino acids. Their investigation showed the value of combining reduced mass and peptide map data to determine the identity of the sequence variant caused by mutation of the original stop codon. In comparison to the current study, the sequence variant reported by Zhang and co-workers was relatively easy to identify because it resulted from a simple addition of amino acids from the non-translated, in-frame region to the antibody, allowing manual interpretation of the data and relatively easy assignment of the added amino acid sequence.

In the example described below, manual assignment of the data and determination of the aberrant amino acid sequence was not readily possible. Instead, we used an approach based on proteomic database searching combined with information obtained by de novo MS/MS spectra assignment. Utilization of semi-automated methodology resulted in a rapid turn-around of analytical data and confident assignment of a novel sequence variant. Furthermore, the relatively easy identification of the unknown species allowed the clones in question to be quickly discarded and project resources to be concentrated on more promising candidates for clone selection and further development.

Results

Analysis of clones by rCE-SDS and CEX assays

The cell line development for mAb1 included the testing of several pools and clones expressed in CHO-K1. Several pools from individual transfections were analyzed, and those selected for sub-cloning were chosen based on optimal cell-culture performance and an assessment of product quality attributes. In total, forty-eight CHO-K1 clones were analyzed. This initial set of clones was narrowed to a final set of 3 top candidates with optimal product quality and cell-culture performance characteristics. These clones were named CHO-K1 #4, #24, and #34 (summarized in Table 1) and will be referred by clone numbers subsequently in the text.

For each clone, production runs were performed in 2 L bioreactors and both harvest and day of culture samples were purified by protein A affinity chromatography. Additionally, bioreactors were run at both pH 6.9 and pH 7.1 for both clones to test the effect of pH on cell-culture and product quality attributes. A standard panel of product quality assays, including rCE-SDS, CEX, and mass analysis were performed on each clone. Comparison of the results for clones 24 and 34 showed the presence of new peaks in both the CEX and rCE-SDS assays. For clone 34, these new peaks were observed at both pH conditions, whereas for clone 24, only the pH 6.9 condition showed new peaks. For clone 4, no new peaks were detected in either assay at either bioreactor pH condition, and profiles were comparable to the control CHO-K1 pool material (data not shown).

Figure 1 shows the rCE-SDS profiles for the control sample compared with clones 24 and 34 produced at pH 6.9. A new peak was visible as a shoulder to the HC peak in both samples. In the control sample, a very minor peak was observed in this region. For most CHO produced samples, low-levels of non-consensus glycosylation (NCG) have been shown to migrate in this region. To test the possibility that the post-HC peaks observed by rCE-SDS could be due to NCG, samples of clones produced at both bioreactor pH conditions were treated by overnight digestion with PNGaseF following reduction and alkylation, which is known to remove NCG. Analysis by rCE-SDS after treatment showed collapse of the HC and non-glycosylated HC peak, as expected, but no significant change in the post-HC shoulder for the samples where it was present (data not shown). This experiment ruled out the presence of NCG as a source of the new species in these 2 clones.

Figure 2 shows the CEX profiles of clone 24 and 34 produced at pH 6.9 compared with the control. For clone 34, the CEX profile at pH 7.1 was comparable to the pH 6.9 data. For both clone 34 and clone 24 produced at pH 6.9, new peaks were observed as shoulders to existing peaks with later retention times. However, clone 24 produced at pH 7.1 (data not shown) did not show extra peaks and the CEX profile was similar to that obtained for the control sample. The observation of new shoulders in the CEX data, combined with the later migrating peak following HC by rCE-SDS, suggested that the additional species present in these samples was due to a post-translational modification of the antibody of significant MW or due to the presence of a new species with a different sequence than expected. The observation that clone 24 appeared to show a pH dependence for the presence of the new species could not be explained by these results, but warranted further investigation. To further understand the root cause for the anomalous peaks observed in the CEX and rCE-SDS assays, reversed phase (RP)-ultra performance liquid chromatography (UPLC)-time of flight (TOF)-MS of the clones was performed.

Analysis of clones by reduced mass

Both bioreactor conditions of clone 24 and the pH 6.9 condition of clone 34 were analyzed using reduced RP-UPLC-MS with TOF detection. These samples were compared to the parent CHO-K1 pool as a control. Prior to separation by RP-UPLC, samples were treated with PNGaseF to remove the N-linked carbohydrate moiety. Separation of clone 24 produced at pH 6.9 by RP-UPLC showed the presence of a pre-HC shoulder (Fig. 3, top panel) not observed in the control sample. Similar pre-HC peak regions were also observed in the UV traces for clone 34 at pH 6.9 and clone 24 at pH 7.1 (data not shown). No other

Table 1. List of samples tested

| Samples Tested                  |
|--------------------------------|
| Control: CHO K1 pool material, produced at pH 7.0 |
| CHO-K1 clone 24, produced at pH 6.9 or pH 7.1 |
| CHO-K1 clone 34, produced at pH 6.9 or pH 7.1 |
unexpected peaks were observed in the RP-HPLC analysis of the samples.

Extraction of the raw spectra corresponding to the main peak in clone 24 and deconvolution showed the presence of the mass expected for the HC, with a mass of 49390.81 Da (Fig. 3D). The expected mass for mAb1 HC after de-N-glycosylation is 49389.71 Da, a mass error of 1.1 Da or 22 ppm. A comparable mass was also observed for the control sample HC main peak (data not shown). Deconvolution of the spectra corresponding to the pre-peak region for the pH 6.9 conditions of clones 24 and 34 showed that the major mass for the pre-peak was consistent with that observed for the main peak. In addition to this expected species, a new mass was observed for both samples at 53206.44 Da (53206.63 Da for Clone 34, pH 6.9), corresponding to a mass shift of 3815.54 Da relative to the theoretical base mass (Fig. 3A and 3B). Analysis of the pre-peak region of clone 24 produced at pH 7.1 showed that the major species was as expected, as for the other clone samples. However, in addition to the expected species a new mass of 50438.59 Da was also observed, corresponding to a mass shift +1047.79 Da (Fig. 3C).

No appreciable level of the 53206.44 Da species observed for the other conditions was observed for the pH 7.1 clone 24 sample.

As noted above, no other unexpected peaks were observed in the reduced RP-UPLC analysis. Deconvolution of the light chain (LC) peak was performed for each sample, with masses corresponding closely to expected values for all samples (data not shown).

Table 2 summarizes the mass data obtained by RP-HPLC-MS for the HC components of clone 24 produced at pH 6.9 or pH 7.1 (as noted above, clone 34 produced at either pH showed comparable data to the clone 24 pH 6.9 condition). To confirm that the HC mass shifts were the only significant changes to the clones with anomalous rCE-SDS and CEX profiles, we performed intact mass analysis of the PNGaseF treated samples. To limit the work required for this study, we focused on the 2 pH conditions of clone 24, since the reduced mass data for clone 34 suggested that both pH conditions gave similar results as the clone 24 sample produced at pH 6.9.

Figure 4 shows the intact mass profiles for the control sample compared with both conditions of clone 24. For all 3 samples, the major species observed was consistent with the expected

Figure 1. Abnormal rCE-SDS traces vs. normal controls. (A) is the rCE-SDS profile of the control sample, compared to the rCE-SDS profiles of (B) Clone 34 produced at pH 6.9 and (C) Clone 24 produced at pH 6.9. The 2 clone samples have an anomalous peak (indicated by arrows) eluting immediately after the heavy chain, compared to the control.
intact mass for de-N-glycosylated mAb1, 145625.32 Da. For the pH 6.9 bioreactor sample (Fig. 4B), 2 new masses were observed, 149444.10 Da and 153259.72 Da. These species correspond to sequential additions of 3815.61 Da and 3815.62 Da, respectively. This observation is consistent with the mass shift observed for the HC of this sample by reduced RP-HPLC-MS analysis. Observation of these species, and no others, confirmed that the only significant change to clone 24 produced at pH 6.9 was the addition of approximately 3816 Da to the HC. Observation of a more abundant signal for the 149444.10 Da mass suggested that most antibodies with a mass shift had one modified and one normal HC, resulting in a single addition of the 3816 Da species. The low level of the 153259.72 Da species indicated that the population of antibodies in the sample with two 3816 Da additions was low.

A very similar observation was made for clone 24 produced at pH 7.1, except that the mass shifts observed were 1047.76 Da and 1048.68 Da for the 2 new species present. These results were also in agreement with the reduced mass data, and again indicated that most of the antibodies in the sample with a modification had one normal and one HC with an approximately 1048 Da increase in mass. The low level of the 147723.81 Da mass corresponding to the 2 mass shifts of approximately 1048 Da showed the limited amount of antibodies with 2 modified HCs.

To determine the region of the HC where the additional mass was present, reduced limited proteolysis with Lys-C was performed to cleave the HC into Fd and Fc fragments. Samples were de-N-glycosylated and mass data was obtained by RP-HPLC-MS for clone 24 produced at pH 6.9 and 7.1. The masses observed for the LC and Fd regions for these samples were comparable to the control (data not shown).

Fig. 5 shows the limited Lys-C profiles mass profiles for the Fc region of the control sample (panel A) compared with both conditions of clone 24. For both clone 24 samples, a low level mass of 25956 Da was present (compared to the major, expected mass of 25178 Da which was also observed) consistent with a 777.8 Da mass increase for the Fc region of the HC. This mass suggested that the additional mass observed by reduced and intact mass could be localized to the Fc region and was susceptible to partial cleavage by Lys-C. For both the pH 6.9 (panel B) and 7.1 samples (panel C), treatment with limited Lys-C proteolysis reduced the additional mass observed from 3816 Da and 1048 Da, respectively, to approximately 777.8 Da. Since no chemical or post-translational modifications were observed for LC or Fd (data not shown), the most likely explanation was the presence of a lysine residue within a putative additional sequence that could be cleaved under these conditions. The observation that the same mass shift of approximately 777.8 Da was observed for both...
samples suggested that the +1048 Da species observed in the pH 7.1 condition was a truncated version of the +3816 Da species observed in the pH 6.9 condition.

We attempted to match the mass shifts observed by reduced and intact mass analysis to possible sequence extensions resulting from mutation of the HC stop codon and translation of usually non-translated sequence, analogous to the observation from the publication by Zhang et al. No matches were observed, suggesting a more complex explanation for these mass shifts. We also considered the possibility that the mass shifts could be due to an unusual post-translational modification, but the mass values could not be related to any known modification. To further investigate the identity of the C-terminal variant, peptide mapping of CHO-K1 Clone 24 produced under different bioreactor conditions was performed.

**Peptide mapping of clones**

To identify the specific nature of the C-terminal extension observed by reduced and intact mass analysis, tryptic peptide map analysis of the CHO-K1 clone 24 produced at pH 6.9 and 7.1 was performed. Clone 34 was not included in the peptide map analysis because clone 24 at pH 6.9 showed a C-terminal extension of the same mass; exclusion of clone 34 was done to save instrument time and resources. Digested peptides were
separated by UPLC on a C8 column with on-line MS detection using an Orbitrap mass spectrometer. Fig. 6 shows a comparison of the A214 nm UV traces for the clone 24 samples compared with the control. Three new peaks by mass not observed in the control were detected in the pH 6.9 sample, and named #1 through #3. (Peak #1 co-elutes with an expected tryptic peptide.) For the pH 7.1 sample, only one new peak was detected, corresponding to peak #3 in the pH 6.9 sample. MS analysis of these peaks showed the presence of unique masses not observed in the control sample (Table 3). Attempts to assign the newly observed masses to the sequence of mAb1 using Mass Analyzer were not successful. Inspection of the MS/MS data for each peak showed the presence of fragmentation patterns consistent with an amino acid sequence (Fig. 7). However, the most likely possible sequence fragments assigned for peaks #1 and #2 did not match any tryptic peptides from the mAb1 coding sequence or from nearby non-coded regions of the construct sequence.

*De novo* analysis of the MS/MS spectra corresponding to peak #3 produced a plausible sequence consistent with an additional mass of 777.34 Da on the C-terminal peptide of the HC, if the C-terminal lysine residue was removed. The additional mass of 777.34 Da was consistent with the additional mass observed on the Fc region of the HC from the limited Lys-C analysis. This sequence was SLSLSPGMNEXXK. Because the C-terminal 7 amino acid residues (S443-G449) of the expected HC sequence were observed as part of this unknown peptide, the location of the extra mass was deduced to be at the C-terminus of HC. Furthermore, since this peak was common to the pH 6.9 and pH 7.1 clone 24 samples, this provided further evidence that both samples contained a similar C-terminal extension. The additional 2 peaks present in the pH 6.9 sample were hypothesized to be either an additional chemical or post-translational modification or sequence variant, or a further extension of the peptide present in peak #3. Given that the C-terminal residue identified in peak #3 was a lysine residue, cleavage after this residue in the trypsin digestion seemed a likely explanation, and suggested further C-terminal extension as a more likely explanation. The additional sequence identified for peak #3 by the *de novo* MS/MS assignment was inconsistent with the non-translated sequence C-terminal to the HC coding region, even with removal of the codons for C-terminal lysine, the stop codon or both. This observation suggested that a simple mutation near the coding region for the C-terminus of the HC resulting in a read-through error was not responsible for the additional mass observed in these mAb1 CHO-K1 clones.

To further identify the remaining unknown peaks (#1 and #2) in the pH 6.9 sample, and to fully account for the 3815.54 Da extension observed by reduced mass, both clones were digested with Asp-N and Glu-C. These enzymes were chosen to provide potentially overlapping sequence with the peptide identified in peak #3, and the differential digestion specificity was used to identify additional peptides for *de novo* and proteomic database searching (see below). As with the tryptic digests, new peaks were observed in both clone 24 samples that were absent from control samples (data not shown). To enable identification of these unknown peptides, the MS/MS data from the 3 enzymatic digests were used in a search against the entire DNA construct of mAb1 using Proteome

| Species (PNGaseF treated samples) | Theoretical Mass (Da) | Experimental Mass (Da) | Difference (ppm) | Mass Shift |
|-----------------------------------|-----------------------|------------------------|------------------|-----------|
| Heavy Chain (des K)               | 49389.71              | 49390.81               | 22               | n/a (1.1 Da) |
| Heavy Chain (des K) + 1048 Da mass shift (Clone 24, pH 7.1) | n/a                   | 50438.24               | n/a              | 1047.79 Da |
| Heavy Chain (des K) + 3816 Da mass shift (Clone 24, pH 6.9) | n/a                   | 53206.44               | n/a              | 3815.54 Da |

**Figure 4.** Intact mass analysis of control (panel A) and clone 24 pH 6.9 (panel B) and pH 7.1 (panel C). Analysis of clone 24 at pH 6.9 showed the presence of 2 new species corresponding to mass extensions of approximately 3515.6 Da, while the same clone at pH 7.1 showed the presence of mass extensions of approximately 1048 Da.
Discoverer. The observation, discussed above, that the C-terminal extension was not due to a simple change in the DNA coding region near the C-terminus of HC suggested that a more complex or random sequence rearrangement event might be the root cause of the C-terminal sequence extension. Since the source of the possible new sequence was most likely derived from the vector DNA rearrangement (as opposed to genomic DNA incorporation), each of the 3 reading frames for the entire antibody expression vector in both the 3'-5' and 5'-3' direction were translated in silico. Using this strategy, numerous theoretical amino acid sequence fragments of varying length were produced, each terminating in a stop codon depending on the frame and direction of the in silico translation. A total of 3198 sequences from the construct were entered into a Proteome Discoverer database as individual proteins, which included the antibody sequence. The MS/MS data from the trypsin, Glu-C and Asp-N peptide maps of the clone 24 samples were searched against the database using the settings described in the methods section. Search results from Proteome Discoverer database searching are summarized in Table 4. Comparison of the high-confidence sequence hits from this search showed the presence of an amino acid sequence, TSSSDFGNTRQ,
which was common to all 3 enzymatic digestions from the pH 6.9 clone 24 sample. This sequence was found to correspond to Peak 2 in the tryptic peptide map. This sequence fragment originates from a region of the 3’-5’ frame 3 in silico translation of the LC construct. The mAb1 control sample, as expected, did not produce any matches to the translated vector DNA sequence other than the normally translated antibody sequence. A second sequence hit for the pH 6.9 clone 24 sample, SNSAYLK, from the 3’-5’ frame 1 in silico translation of the LC DNA construct, was identified with high confidence in the tryptic digestion. This peptide was found to correlate to Peak 1.

The amino acid sequences identified from de novo and database searching of the mass data were searched manually by comparison to the translated DNA vector sequences for the mAb. Two short sequences from the LC vector were identified and confirmed after re-interrogation of the MS data. When the sequences already identified by de novo and database searching were assembled, the full sequence of the extension on the C-terminus of the pH 6.9 clone 24 sample, corresponding to an addition of 3618 Da, was identified as MNESTSKIRSAYLKTAGFEDTSSSDTFGNTRQ (Fig. 8). Interestingly, the full extension was observed to be derived from 2 frames of the LC construct. The first 18 amino acids of the sequence (MNESTSKIRSAYLKT) were from the 3’-5’ frame 1 of the LC construct, and the last 17 amino acids of the sequence (GFTEDTSSSDTFGNTRQ) were from the 3’-5’ frame 3 of the LC construct. The first 9 amino acids of the full extension were consistent with the +1048 Da extension observed by reduced mass analysis, and the entire sequence was consistent with the +3816 Da extension observed by reduced mass. The length of the extension was observed to correlate with bioreactor conditions, where the higher pH for clone

### Table 3. Mass data for unique species observed in tryptic digests of clone 24 samples. The most likely sequence determined by de novo interpretation of the data is shown.

| Peak | Observed in clone 24, pH 6.9 (3816.86 Da extension) | Observed in clone 24, pH 7.1 (1048.54 Da extension) | Observed monoisotopic mass (Da) | Most plausible de novo sequence |
|------|--------------------------------------------------|--------------------------------------------------|-------------------------------|--------------------------------|
| 1    | Yes                                              | No                                               | 781.42                        | XNSAYLK                        |
| 2    | Yes                                              | No                                               | 2021.88                       | XXKFTEDSSSDTFGNTRX             |
| 3    | Yes                                              | Yes                                              | 1436.70                       | SLSLSPPGMNEXXX                 |

Figure 7. MS/MS fragmentation for new peaks identified in clone 24 at pH 6.9. The most plausible sequence determined by manual de novo analysis of the results are shown for peak #1 (A), peak #2 (B), and peak #3 (C). The most likely y- and b-ion series for the putative sequence shown in each panel is given.
24 resulted in apparent cleavage within the C-terminal extension. This pH dependence was not observed for clone 34.

**Discussion**

A C-terminal extension was observed in 2 CHO-K1 clones and was found to be due to the re-arrangement of the DNA construct. The extension was derived from the LC vector non-translated sequence and was inserted prior to the C-terminal lysine of the HC sequence. No translated variant sequence with the lysine at the C-terminus was found. Two forms of this extension were observed for clone 24 based on the pH condition of the bioreactor; for clone 34, only the longer extension was present. When the longer extension was present in a sample, changes to both rCE-SDS and CEX profiles were observed. However, for clone 24 grown at pH 7.1, only the shorter extension was observed and the rCE-SDS and CEX profiles were not significantly affected.

Given the challenge of identifying a sequence extension derived from non-translated DNA, the combination of reduced and intact mass analysis with proteomic searching described here to quickly identify the species shows the power of this approach. Furthermore, the findings described above demonstrate the value of mass spectrometry testing even at early phases of development. Had the short (9 amino acid residues) version of the C-terminal extension been the only form present, the changes in rCE-SDS and CEX product quality assays would not have provided a signal that further investigation was required and resources would not have been expended on clones unsuitable for further development. Mass analysis provides a valuable check and screening tool to eliminate these kinds of rare sequence variants. This technique could be employed in the future for clone screening experiments when unexpected species are observed and a simple post-translational modification cannot explain the mass shift.

**Materials and Methods**

**Materials**

Trypsin, Asp-N, Glu-C, Lys-C, and dithiothreitol (DTT) were purchased from Roche Applied Science. LC/MS grade solvents, 0.1% trifluoroacetic acid (TFA) in water and 0.1% TFA in acetonitrile (ACN) were bought from Burdick and Jackson (Honeywell). Sodium iodoacetate was purchased from Sigma, and PNGaseF from New England Biolabs. BioSpin-6 columns (Biorad) were used for desalting. CHO-K1 clones were produced at Amgen.

**Reduced CE-SDS and non-consensus glycosylation PNGaseF digestion test**

A standard rCE-SDS method was used for these analyses on a Beckman PD-800 plus instrument. Briefly, this method separates proteins based on differences in their hydrodynamic size under reducing and denaturing conditions. The protein species are bound to SDS, an anionic detergent, and electrokinetically injected into a bare fused silica capillary filled with SDS gel buffer. An electrical voltage is applied across the capillary, under which the SDS coated proteins are separated by their difference in migration in a hydrophilic polymer based solution. Proteins are detected by a photodiode array detector as they pass through a UV detection window. Purity is evaluated by determining the percent corrected peak area of each component. For determination of the identity of the post-HC minor peak region, the sample was reduced and alkylated using the procedure described for peptide map analysis (see below). Reduced and alkylated samples were treated with and without PNGaseF to test whether the post-HC peak showed the presence of N-linked glycans.

---

**Table 4.** Amino acid sequences identified in Proteome Discoverer when searching against the entire DNA vector construct for mAb1. Three enzymatic digestions were performed on clone samples containing additional mass on the heavy chain. A common amino acid sequence, TSSSDTGFNTRQ, was found in all 3 digestions, which were found to correlate with Peak 2 in the tryptic peptide map. An additional amino acid sequence, SNSAYLK, was identified in the tryptic digestion that was found to correlate to Peak 1.

| Enzyme | Peak Retention Time (min) | m/z       | Amino Acid Sequence from Proteome Discoverer Search against mAb1 DNA construct |
|--------|--------------------------|-----------|---------------------------------------------------------------------------------|
| Trypsin| 46.63                    | 2224.95 (2+) | CITGFEDTSSSDTGFNTRQ                                                            |
|        | 29.55                    | 782.40 (1+) | SNSAYLK                                                                        |
| Glu-C  | 26.71                    | 1300.58 (2+) | TSSSDTGFNTRQ                                                                  |
| Asp-N  | 24.94                    | 1415.60 (2+) | DTSSSDTGFNTRQ                                                                 |
|        | 25.02                    | 938.43 (2+) | DTFGNTRQ                                                                      |

**Figure 8.** C-terminal sequence extension observed on mAb1 CHO-K1 clones 24 and 34. The sequence extension is derived from the LC 3’-5’ frame 1 construct region and from the LC 3’-5’ frame 3 construct region. Notations indicate the extension observed for clone 24 produced at pH 7.1 (+3816 Da) and at pH 6.9 (+3816 Da). For clone 34, the +3816 Da extension was observed at both pH 6.9 and 7.1.
Cation exchange chromatography

Charged variants were separated using a YMC BioPro SP-F, 5 μm column (100 x 4.6 mm) held at a temperature of 30°C on a Waters Alliance HPLC system. Separation conditions for 110 μg injections of antibody included an initial gradient hold of 2 minutes at 100% buffer A (20 mM sodium phosphate, pH 6.5), followed by a linear gradient of 0–10% buffer B (Buffer A + 500 mM sodium chloride) in 30 minutes at a flow rate of 0.8 mL/min. Separations were monitored at 220 nm.

RP-HPLC-MS for reduced antibody

Samples were diluted to 2 mg/mL, deglycosylated with PNGaseF at 37°C for 2 hours, and then reduced at 55°C for 30 minutes under denaturing conditions with 100 mM DTT and 6 M guanidine hydrochloride at pH 8.3. Reduced mass analyses of the clones were performed on a Waters Acquity UPLC system connected inline to an Agilent LC/MSD TOF. A Waters Acquity BEH Phenyl column of 2.1 mm internal diameter x 150 mm length, particle size 1.7 μm, was used for separation of the LCs and HCs under linear gradient conditions from 34% B to 40% B in 7.5 minutes. Solvent A was 0.1% TFA in water and solvent B was 0.1% TFA in ACN. Spectra containing charged protein ions were deconvoluted using Agilent Mass Hunter software.

Intact mass analysis of intact antibody samples

Samples were diluted to 2 mg/mL and deglycosylated with PNGaseF at 37°C for 2 hours. Intact mass analyses of samples were performed on an Agilent LC/MSD TOF using an SEC-MS method adapted from ref. 12. A Waters Acquity BEH200 SEC column of 4.6 mm internal diameter x 150 mm length, particle size 1.7 μm, was used for rapid desalting and isotropic elution of the antibody. The elution solvent was 0.1% formic acid in 15% ACN and 85% water. Spectra containing charged protein ions were deconvoluted using Agilent Mass Hunter software.

Limited Lys-C proteolysis

Samples were diluted to 2 mg/mL, deglycosylated with PNGaseF at 37°C for 2 hours, and then digested with Lys-C at 37°C for 80 minutes with endoproteinase ratio of 1:400, based on the procedure described in ref. 13. Samples were subsequently reduced at 55°C under denaturing conditions with 100 mM DTT and 6 M guanidine hydrochloride at pH 8.3. Mass analysis was performed on an Agilent LC/MSD TOF. A Waters Acquity BEH Phenyl column of 2.1 mm internal diameter x 150 mm length, particle size 1.7 μm, was used for separation of the Fc/2, LC and Fd under linear gradient conditions from 30% B to 45% B in 7.5 minutes. Solvent A was 0.1% TFA in water and solvent B was 0.1% TFA in ACN. Spectra containing charged protein ions were deconvoluted using Agilent Mass Hunter software.

Peptide mapping with trypsin, Glu-C, and Asp-N digestions

Samples were reduced and alkylated under denaturing conditions by mixing 10 μL of each sample (10 mg/mL) with 1.8 μL of 0.5 M DTT and 3.4 μL of 0.5 M sodium iodoacetate in denaturing buffer (5.5 M Guanidine HCl, 0.2 M Tris HCl at pH 8.3). Samples were desalted using BioSpin-6 columns, and subsequently digested with either trypsin, Asp-N, or Glu-C. Samples were digested at 37°C with an endoproteinase ratio of 1:10. Trypsin digestions were carried out for 35 minutes, Asp-N digestions were carried out for 4 hours, and Glu-C digestions were carried out overnight (approximately 18 hours). LC/MS/MS analyses of the proteinase digested samples were performed using a Waters Acquity UPLC system coupled with a Thermo Fisher Orbitrap Velos mass spectrometer equipped with an electrospray ionization source. The digested samples were injected onto Waters C8 (trypsin) or C18 (Asp-N and Glu-C) BEH columns, with a column temperature of 60°C. Solvent A was 0.1% TFA in water, and solvent B was 0.1% TFA in ACN. Peptides were eluted using a gradient of 0% B to 50% B in 180 minutes. Data analysis was performed using Thermo Fisher Xcalibur software, in addition to database searching performed in Thermo Fisher Proteome Discoverer.

Creation of the FASTA file for database searches

The DNA vectors, for the HC and LC, were translated (all 3 frames) using the Translate tool created by the Swiss Institute of Bioinformatics (http://web.expasy.org/translate/). Each translated protein sequence (separated by DNA stop codons) was given a unique number. All of the translated protein sequences were formatted as a FASTA database. The translated vector protein sequences were combined with common contaminants (http://maxquant.org/contaminants.zip) and the NCBI CHO proteome. All of the mass spectrometry data was searched with the combined database.

Sequest Search of Mass Spectrometry Data

The MS/MS data were searched against our combined database, with each sequence appearing in both normal and reversed orientation. The search workflows were setup using Proteome Discoverer (version 1.3.0.399). The searches used the Sequest search algorithm. Percolator was used for peptide validation (http://noble.gs.washington.edu/proj/percolator/). The search parameters for the trypptic digestes were configured for semi-trypptic enzyme specificity allowing 2 missed cleavages. The search parameters for the Glu-C and Asp-N digestes were configured for full enzyme specificity allowing for 1 missed cleavage. The precursor tolerance was 10 ppm and the fragment tolerance was 0.8 Da for all enzymatic conditions. The variable modifications included for all of the searches were oxidation (M + 15.995) and deamidation (N, Q + 0.984). Carboxymethylation (C + 58.005) was set as a static modification in all cases. A 1% false-discovery rate (FDR) was applied to peptide-spectrum matches (PSM). The FDR for PSMs were calculated by Percolator and reported as q-values. At the protein level, manual inspection of the MS/MS spectrum was performed for any protein identification supported by one only peptide.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgment

We acknowledge Sean Rumberger for performing the investigation of the potential non-consensus glycosylation by rCE-SDS.
References

1. Meert C, Brady LJ, Guo A, Balland A. Characterization of antibody charge heterogeneity resolved by preparative immobilized pH gradients. Anal Chem 2010; 82:3510-8; PMID:20364842; http://dx.doi.org/10.1021/ac902408r

2. Harris R, Murtane AA, Utter S, Wagner K, Cox E, Polastrini G, Helder JC, Sliwkowski MB. 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 1993; 11:1293-7; PMID:7764191; http://dx.doi.org/10.1038/nbt1193-1293

3. Dorai H, Sauerwald T, Campbell A, Kyung Y, Goldstein J, Magill A, Lewis M, Tang Q, Jan D, Ganguly S. Investigation of product microheterogeneity: a case study in rapid detection of mutation in mammalian production cell lines. Bioprocess International 2007; 5, no. 8:66

4. Huang Y, O’Mara B, Conover M, Ludwig R, Fu J, Tao L, Li ZJ, Riebel S, Grace MJ, Russell RJ. Glycine to glutamic acid misincorporation observed in a recombinant protein expressed by Escherichia coli cells. Protein Sci 2012; 21:625-32; PMID:22362707; http://dx.doi.org/10.1002/pro.2846

5. Ferey L, Carvalhal V, Yu XC, Chan B, Michels DA, Wang YJ, Shen A, Redd J, Duval B, Laird MW. Eliminating tyrosine sequence variants in CHO cells lines producing recombinant monoclonal antibodies. Biotechnol Bioeng 2013; 110, no. 4:1087-97; PMID:23108857; http://dx.doi.org/10.1002/bit.24759

6. Zeck A, Regula JT, Larrailer V, Mauriz B, Popp O, Gopfert U, Wingsheff F, Vollersten UE, Gott IH, Koll H, et al. Low level sequence variant analysis of recombinant proteins: an optimized approach. PLoS One 2012; 7, no. 7:1-10; PMID:22792284; http://dx.doi.org/10.1371/journal.pone.0040328

7. Zhang Z, Shah B, Bondarenko P. GU and certain wobble position mismatches as possible main causes of amino acid misincorporations. Biochemistry 2013; 52, no. 45:8165-76; PMID:24128183; http://dx.doi.org/10.1021/bi401002c

8. Zhang T, Huang Y, Chamberlain S, Romeo T, Zhu Xiong J, Hewitt D, Zhu M, Katta V, Mauger B, Kao Y-H. Identification of a single base-pair mutation of TAA (Stop codon) >GAA (Glu) that causes light chain extension in a CHO cell derived IgG1. mAbs 2012; 4, no. 6:694-700; PMID:22908810; http://dx.doi.org/10.4161/mabs.22232

9. Yang Y, Strahan A, Li G, Shen A, Liu H, Ouyang J, Katta V, Francissen K, Zhang B. Detecting low level sequence variants in recombinant monoclonal antibodies. mAbs 2010; 2, no. 3:285-98; PMID:20400866; http://dx.doi.org/10.4161/mabs.2.3.11718

10. Zhang Z. Large-scale identification and quantification of covalent modifications in therapeutic proteins. Anal Chem 2009; 81, no. 20:8354-64; PMID:19764700; http://dx.doi.org/10.1021/ac901193n

11. Valliere-Douglass J, Kodama P, Mujacic M, Brady L, Wang W, Wallace A, Yan B, Reddy P, Treuheit MJ, Balland A. Asparagine-linked oligosaccharides present on a non-consensus amino acid sequence in the CH1 domain of human antibodies. J Biol Chem 2009; 284, no. 47:32493-506; PMID:19767389; http://dx.doi.org/10.1074/jbc.M109.014803

12. Brady L, Valliere-Douglass J, Martinez T, Balland A. Molecular mass analysis of antibodies by on-line SEC-MS. JASMS 2008; 19:502-9; PMID:18258452; http://dx.doi.org/10.1016/j.jasms.2007.12.006

13. Gadgil H, Bondarenko P, Pipes G, Dillon T, Banks D, Abel J, Kleeman G, Treuheit M. Identification of cysteinylation of a free cysteine in the Fab region of a recombinant monoclonal IgG1 antibody using Lys-C limited proteolysis coupled with LCMS analysis. Anal Biochem 2006; 355, no. 2:165-74; PMID:16828048; http://dx.doi.org/10.1016/j.ab.2006.05.037