Sequence Variant Analysis Using Peptide Mapping By LC–MS/MS
Assessing Genetic Heterogeneity in MAb-Producing Cell Lines

Amy H. Que, Boyan Zhang, Yi Yang, Jennifer Zhang, Gayle Derfus, and Ashraf Amanullah

Monoclonal antibodies are usually expressed in mammalian cell lines and are produced in several variants known as isoforms (1–2). Microheterogeneity can result from posttranslational and enzymatic modifications as well as those caused by processing, alteration, storage, and incorrect translation of the target protein (1, 3). Common sources of heterogeneity include Fc glycosylation, partial carboxypeptidase processing of heavy-chain (HC) C-terminal lysine residues (4), deamination or isomerization (5), Fc methionine oxidation, hinge-region fragmentation (6), aggregation, and sequence variants. Sequence variants are protein isoforms containing unexpected amino acid sequences. They are classified as “product-related impurities.” The presence of unexpected amino acids may pose concerns regarding bioactivity, stability, and immunogenicity (1, 3).

Peptide mapping by mass spectrometry (MS) is a valuable tool for characterizing sequence variants, including single amino acid substitutions in protein variants. These modifications can introduce changes in molecular mass and fragmentation patterns in MS/MS spectra between a precursor (expected) peptide and the modified variant. Consequently, the specific amino acid modification can be identified and localized. To date, most published works on sequence variant characterization involve studies of model proteins such as hemoglobin (Hb) and κ-casein in academic research (9–11). In most cases, sequence variants were isolated and enriched to a high level for the convenience of characterization.

Wade et al. published review articles about detection, characterization, and structural analysis of hemoglobin sequence variants using advanced analytical methods, including electrophoretic and chromatographic methods, MS, and DNA analysis (12, 13). Ahrer et al. reviewed protein characterization, including sequence variants, using chromatographic and electrophoretic techniques (14). In general, however, detection, characterization, and understanding of sequence variants have not been widely studied (or at least not published) in the industrial field of protein therapeutics (7, 14–17).

Detecting and characterizing unanticipated sequence variants, especially those present in low quantities, remains a significant challenge. At Genentech, we have developed sensitive and robust analytical methodologies to detect sequence variants at both the DNA and protein levels. Modern molecular biology methods such as quantitative polymerase chain reaction (QPCR) help rapidly identify single nucleotide polymorphism (SNP) and detect DNA sequence variation (15). A novel approach combining peptide mapping using LC–MS/MS with Mascot error tolerant search (ETS) has been developed and reported for detecting protein sequence variants with amino acid substitutions in low abundance (16–17).

It is vital to select a clone that is suitable for early stage process
development efforts leading to clinical manufacturing. Here we describe the development of a peptide mapping method using LC–MS/MS combined with Mascot ETS to evaluate variants containing amino acid substitutions for a monoclonal antibody. Visual comparison of the wild and variant-type tryptic map UV profiles did not detect the presence of sequence variants. However, when LC–MS/MS data were submitted for the Mascot ETS, a low level of sequence-variant sequence variants. However, when LC–MS/MS data were submitted for the Mascot ETS, a low level of sequence-variant S52G HC was identified and determined to be 0.2% at the peptide level in a lead clone.

**Materials and Methods**

**CHO Cell Lines:** A humanized IgG1 monoclonal antibody X (referred as MAb X in this work) was produced using cell lines derived from Chinese hamster ovary (CHO) DUKX B11 dihydrofolate reductase–negative (DHFR–) host cell lines, derived from Chinese hamster ovary (CHO) DUKX B11 (referred as MAb X in this work) was produced using cell lines (www.chemie.de). Initial culture medium supplemented with protein hydrolysate, adding a nutrient feed on day 3. Glucose concentration was maintained above 3 g/L by glucose addition as needed.

Upon termination of the culture, the cell culture fluid was centrifuged for 10 minutes at 7,000 rpm, then sterile-filtered through a 0.2-μm polyethersulfone filter unit from Pall (www. pall.com). Then clarified culture fluid was loaded on a Prosep-vA high-capacity protein A column from Millipore (www.millipore.com) that had been equilibrated with 25 mM Tris, 25 mM NaCl, and 5 mM EDTA at pH 7.1. The column was washed with three column volumes of equilibration buffer, followed by three column volumes of 0.4 M potassium phosphate at pH 7.0, and then three column volumes of equilibration buffer. The antibody was eluted with 0.1 M acetic acid at pH 2.9 and neutralized to pH 5.0 by the addition of 1.5 M Tris base.

**Trypsin Digestion:** S-carboxy-methylation was conducted before digestion with trypsin: 1 mg of purified protein (e.g., 50 μL at 20 mg/mL) was mixed with 20 μL of 1 M dithiothreitol from Sigma (www.sigmaaldrich.com) and 950 μL denaturing buffer (pH 8.6) containing 6 M guanidine hydrochloride, 360 mM Tris, and 2 mM EDTA, followed by an hour’s incubation at 37 °C. Then 50 μL of 1 M iodoacetic acid from Sigma, freshly prepared in 1 M NaOH, was added to the sample. After a 15-min sample incubation at room temperature in the dark, the alkylation reaction was quenched by addition of 10 μL dithiothreitol. Reduced and S-carboxymethylated samples were exchanged using PD-10 columns containing Sephadex G-25 medium from GE Healthcare Bio-Sciences.
AB (www.gelifesciences.com) into a pH 8.2 buffer containing 25 mM Tris and 2 mM CaCl₂. MS-grade Trypsin from Promega (www.promega.com) was added at a ratio of ~1:50 (w/w) enzyme to protein. Digestion proceeded for five hours at 37 °C, then the reaction was quenched by addition of an aliquot of 10% (v/v) trifluoroacetic acid (TFA).

**LC–MS/MS Analysis:** We analyzed the protein digests using LC–MS/MS on a ThermoFinnigan LTQ instrument equipped with a standard electrospray ionization (ESI) source from Thermo Fisher Scientific (www.thermofisher.com). The system interfaced with an Agilent 1200 HPLC unit with an in-line UV detector from Agilent (www.agilent.com).

In data-dependent scan experiments, the instrument was set to conduct 11 scan events including a single MS scan followed by five repeated cycles of a zoom scan, then one MS/MS scan on the five most intense ions in the MS scan. We analyzed the data from our tandem MS experiments with ThermoFinnigan Xcalibur software (www.thermofinnigan.com). Separation was carried out on a 5-µm, 300 Å, 250 × 2-mm Jupiter C18 column from Phenomenex (www.phenomenex.com) at a flow rate of 0.25 mL/min, with mobile phases containing 0.1% TFA in H₂O (solvent A) and 0.09% TFA in 90% MeCN (solvent B). Column temperature was set at 55 °C. A mixture of peptide sample corresponding to 5 µg of protein was loaded onto the column before injection.

**Mascot Error Tolerant Search:** Mascot by Matrix Science Inc. (www.matrixscience.com) is a search engine that uses MS data to identify proteins from primary sequence databases (19–21). MS/MS ion search of an LC–MS/MS sample corresponding to 5 µg of protein was loaded onto the column before injection. Mascot was used to search against the entire in-house database consisting of ~150 entries of molecular sequences. Our second pass was a manual ETS that provides a comprehensive search for a selected protein hit, providing a wide range of possible variations, including posttranslational modifications (PTM) and amino acid substitutions. The “Mascot Search Parameters” box lists parameters for both standard and ETS.

**RESULTS AND DISCUSSION**

In selecting a clone for early phase clinical production, we typically analyze several clones by comparing their peptide maps to present for the presence of sequence variants in each clone. Our rationale is that a change in the primary structure of a protein produced from a clone could be detected by a change in the peptide mapping profile (e.g., the appearance of new peaks). We compared tryptic peptide maps generated using LC-UV methods of protein-A–purified MAb X from four clones (A, B, C, and D) detected at 214 nm (Figure 1). No visible difference was observed among the four clones, suggesting that no sequence variants were present at high-enough abundance in any of the clones. If there were any variants, then they may have coeluted with another wild-type peptide (or could be present at low abundance). The tryptic peptide map covers ~98% of the molecule’s theoretical amino acid sequence.

To confirm whether sequence variants were present in low quantities, we acquired LC–MS/MS data and submitted them to Mascot with manual ETS for our lead and back-up clones. Mascot ETS generated a list of possible modifications and amino acid substitutions. For the lead clone, one sequence variant with a single amino acid substitution located in the heavy chain was tentatively proposed. The proposed peptide variant was tryptic peptide HC44–67 with a glycine (G) replacing a serine (S) at position 9.

**MASCOT SEARCH PARAMETERS**

<table>
<thead>
<tr>
<th>Mascot Search Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>The search parameters for both standard and ETS are described below:Carboxymethyl was selected as the fixed modification. Variable modifications included oxidation on methionine, deamination on asparagine and glutamine, des-C-terminal lysine, and glycosylation (Go) on asparagine. Trypsin was the specified enzyme. Maximum miscleavage site was set to 1. Precursor peptide charge states were set at 2 and 3. Mass tolerance was set to 800 ppm for both MS and MS/MS.</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2** shows the mass spectra of zoom scans for expected peptides and identified variant peptides. We determined the mass of the expected peptide to be m/z 1199.74 with a charge state z = 2 (Figure 2a), corresponding to m/z 2398.48 of MH⁺, which matches well with the theoretical MH⁺ m/z 2398.13 (Table 1). Similarly, we determined the proposed variant peptide to be m/z 1184.62 with a charge state z = 2 (Figure 2b), corresponding to m/z 2368.24 of MH⁺, which matches well with the theoretical MH⁺ m/z 2368.12 (Table 1). The measured mass change (~30.24 Da) from S9→G closely matches the theoretical mass change (~30.01 Da), indicating the presence of a variant peptide S52G HC (Table 1).

**Figure 3** shows the MS/MS spectra of the expected (Figure 3a) and variant type (Figure 3b) peptides. Both cover a series of Y ions that can be used to confirm the sequence of an expected wild-type and variant peptide. The mass difference between the Y₁₂ and Y₁₅ ions deduces the presence of S at position 9 in the wild peptide (Figure 3a). Y ions in the MS/MS spectrum of the variant peptide
peptide (Figure 3a) were the same as those in the wild peptide (Figure 3a) from Y₄ to Y₁₅. After the Y₁₅ ion, a series of Y–30 ions was also measured (e.g., Y₁₆–30, Y₁₇–30, Y₁₉–30, Y₂₀–30), indicating that the amino acid at position 9 was substituted by one whose mass was 30 Da less, which matches G. So the mass difference between the Y₁₅ and Y₁₆–30 ions deduces the presence of G at position 9 in the variant peptide (Figure 3a). That allowed confirmation that the amino acid serine at position 9 was substituted by glycine, S₉→G using MS and MS/MS spectra.

We compared the extracted ion current of the variant peptide with that of corresponding wild peptide to determine the level of sequence variant at the peptide level in the peak area (Figure 4), which we estimated to be 0.2%. To ensure that such low-level detection of a sequence variant is indeed real and not a false positive, retention-time change should also match the peptide hydrophobicity change between the variant and expected peptides. The detection limit of LC–MS/MS tryptic peptide mapping is peptide dependent. Our peptide map was reproducible and robust. With more than three sample preparations and two injections per sample preparation, we could reproducibly detect the variant peptide containing S52G HC at the ~0.2% level.

In addition, a team studied reproducibility using two antibody mixtures in a modeling study (spiking rhuMAb A into rhuMAb B), reproducibly detecting multiple variant peptides at the 0.5% level (16–17). Alteration in the primary structure of a protein can result from changes at the nucleic acid or protein level (26, 27). Sequence variants may be attributed to genomic DNA mutation (7, 14–15) due to increased mutation rates associated with transfection of DNA into mammalian cells (27) and misincorporation at the protein level due to mistranslation. Harris et al. reported a Y376Q variant in the heavy chain of rhuMAb HER2 produced in CHO cells from a vector carrying the dihydrofolate reductase (DHFR) gene (7). Dorai et al. reported that ~10% of a recombinant peptide–antibody fusion protein expressed by CHO cells contained a Phe to Leu sequence variant (15). Those two antibody variants were verified by analysis using the polymerase chain reaction (PCR), which revealed genetic variations at the DNA level in some subclones of stable parent CHO expression cell lines (15, 16).

Translational errors can result from either misacetylation (mischarging) of the cognate tRNA by the specific tRNA synthetase or codon–anticodon mismatch or through ribosome misreading (28). Yu et al. reported low-level amino acid misincorporation resulting from codon–specific mistranslation in recombinant MAbs as therapeutics expressed in CHO cells (28). And Guo et al. investigated the impact of selection and amplification reagent using MTX on genomic DNA mutation rates of the hypoxanthine-guanine phosphoribosyltransferase (HGPR or HPRT) gene using a 6-thioguanine assay under various concentrations of MTX selection in CHO cells (29). Their results showed that HGPR mutation rate increased as MTX increased during stable cell line development. They reported low levels of two sequence variants observed in two stable cell lines expressing different MAbs, showing that one sequence variant was due to genomic nucleotide sequence change whereas another was probably due to translational misincorporation.

To ensure the safety, efficacy, and consistency of protein therapeutics as required by regulatory authorities, extensive analytical assessment and characterization of recombinant proteins are conducted in process development and later during clinical and commercial manufacturing.
Selection of a clone for protein production is based on high productivity, acceptable metabolite profiles, and acceptable product quality attributes. One of the latter includes assessment of the genetic homogeneity of the cell line used for production of clinical materials so that risks associated with sequence variants can be assessed and mitigated. The LC–MS/MS method reported here is a valuable tool for making this assessment, providing low-level variant sensitivity. It can be used to detect an exact amino acid substitution as well as its location.

References


Corresponding author Amy H. Que is a scientist, Gayle Derfus is an engineer II, and Ashraf Amanullah is director of Oceanside pharma technical development in US biologics for Genentech, 1 Antibody Way, Oceanside, CA 92056; 1-760-231-3061, fax 1-760-231-2465; que.amy@gene.com. Boyan Zhang and Jennifer Zhang are scientists, and Yi Yang is a senior research associate in protein analytical chemistry for pharma technical development in US biologics at Genentech, 1 DNA Way, South San Francisco, CA 94080. 