



A Spike-Control Approach that Evaluates High Resolution Mass Spectrometry-Based Sequence Variant Analytical Method Performance for Therapeutic Proteins

Jinhui Zhang¹ · Mack Shih¹ · Haoheng Yan^{2,3} · Thomas O'Connor¹ · Chengjie Ji⁴ · Patrick J. Faustino¹

Received: 21 March 2023 / Accepted: 20 April 2023 / Published online: 1 May 2023

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023

Abstract

An amino acid sequence variant (SV) is defined as an unintended amino acid substitution in protein drug products. SVs contribute to product heterogeneity and can potentially impact product quality, safety, immunogenicity, and efficacy. The analysis of biotherapeutics for SVs is important throughout the product life cycle including clone selection, development of nutrient feed strategies, commercial manufacturing process, and post-approval changes to monitor product quality. The proposed analytical procedure for SVs consists of both qualitative (identification of SVs) and quantitative (quantitation of identified SVs) components. The complexities of SV analysis and the variety of current procedures highlight the need for a systematic approach for assessing the capability of these methodologies to reliably identify and quantitate SVs in biotherapeutics. We described here a “spike-control” approach for evaluating SV analytical procedure. The concept was adopted from quality control samples routinely used in analytical procedure validation. One FDA approved monoclonal antibody (mAb) was spiked with accurate amounts of highly homologous mAb to create mAb samples containing low yet accurate levels of “artificial” SVs. Spike-control samples were denatured, reduced, alkylated, digested and then analyzed by high resolution Orbitrap mass spectrometry. *In silico* analysis revealed four single amino acid differences between the two mAbs that could be used to represent SVs in the spike-control samples. All four “artificial” SVs were reliably identified by the current workflow. Analytical range (0.01% to 2%), accuracy and precision of identified SVs have also been evaluated. Overall, spike-control sample(s) helped to demonstrate that the SV analytical procedure (*i.e.*, sample preparation, LC separation, mass spectrometry determinations and bioinformatic software) was fit for purpose and suitable for the identification and quantitation of SVs at a pre-determined threshold.

Keywords ionization efficiency · mass spectrometry · sequence variants · system suitability

Introduction

Recombinant therapeutic proteins are a major class of medicines that continues to grow to treat numerous human diseases. Therapeutic proteins have an inherent degree of

product heterogeneity (or micro-heterogeneity) due to the biosynthetic processes used by living cell organisms to produce them [7]. Heterogeneity in therapeutic proteins may arise in a number of ways including DNA replication error, amino acid misincorporation, and post-translational modification. This heterogeneity may affect the expected clinical performance of final products [4]. Characterization of the product-related variants contributing to product heterogeneity and determination of the relative levels of these variants are important for product quality evaluation of therapeutic proteins and similarity assessment of biosimilars [4, 7].

Among the variants contributing to product heterogeneity, amino acid sequence variants (SVs) caused by unintended amino acid substitution during protein translation are increasingly drawing attention from both the biopharmaceutical industry and regulatory agencies [1, 5, 10, 15, 20]. In natural biological systems, SVs are usually present at a very low level due to the high fidelity of translational machinery and

✉ Jinhui Zhang
Jinhui.Zhang@fda.hhs.gov

¹ Office of Testing and Research, Office of Pharmaceutical Quality, Center for Drug Evaluation and Research, Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA

² Office of Biotechnology Products, Office of Pharmaceutical Quality, Center for Drug Evaluation and Research, Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA

³ Global Regulatory Affairs, Shanghai Henlius Biotech Inc., 9F, Innov Tower, 1801 Hongmei Road, Shanghai, China

⁴ NovaBioAssays, 52 Dragon Court, Woburn, MA 01801, USA

the error correction processes, which evolved to reduce the occurrence of spontaneous errors during protein biosynthesis. For recombinant therapeutic proteins, biomanufacturing systems are developed to achieve maximum yield by selecting high-producing recombinant cell lines followed by intensified bioprocessing in bioreactors [5, 6, 10]. Such strategies can impose higher than normal translational machinery stress for the recombinant cell lines during the expression of therapeutic proteins [12, 16]. If not fully optimized, elevated levels of SVs could be generated in the protein products resulting in a variety of quality issues (e.g. stability, aggregates, etc.). Therefore, identification and quantitation of SVs may be used to guide process development and optimization in addition to being assessed as a product quality attribute (PQA) throughout the lifecycle of the drug product [5, 6, 10, 16].

The analysis of therapeutic proteins for SVs is a two-stage analytical procedure where identification of SVs is followed by quantitation of the identified SVs. Identification of SVs is a qualitative assay by nature, whereas determination of the relative/absolute levels of the identified SVs is a quantitative assay. Generally speaking, fit-for-purpose analytical methods should be designed and evaluated based on product knowledge, especially structure, heterogeneity, and the ranges of variability of different variants [4, 7]. However, amino acid substitutions can theoretically occur anywhere in the primary sequence of a therapeutic protein product with widely varying levels of abundance. The lack of qualitative and quantitative information of the “target” analytes makes the identification and quantitation of SVs an analytical challenge. Bottom-up peptide mapping by LC–MS/MS using high resolution mass spectrometry has become the “go-to” method for SV analysis [11]. The experimental workflow is similar to that for primary sequence confirmation of therapeutic proteins, albeit the subsequent bioinformatic data analysis is done using different algorithms. High resolution mass spectrometry methods along with advanced bioinformatics approaches have been reported to be capable of detecting SVs at levels approaching 0.001% [2]. However, the SV detection limit highly depends on the specific mass spectrometry instruments, data acquisition, and bioinformatic analysis software parameters used. Additionally, differences in sample processing methods such as the digestive enzymes, digestion conditions, and chromatographic separation procedures can also impact the specificity and accuracy and thus the overall performance and outcome of the SV analytical procedure.

A recently published survey of industry practices regarding SV analysis coordinated by the International Consortium for Innovation & Quality in Pharmaceutical Development (IQ Consortium) showed that there were significant differences in SV identification and quantitation among the participating companies [15]. The differences included the protease digestion procedure, peptide charge status selected for MS/

MS, isotopic peptide(s) selected for quantitation, pre-defined threshold, the software used for data interpretation, and the use of orthogonal methods (LC–MS analysis of intact antibodies and subunits, charge-based electrophoretic or chromatographic methods) [15]. Moreover, the survey found that these SV analytical methods did not typically use a system suitability standard that might provide assurance that the LC–MS/MS system is operating at a consistent, sensitive level of detection. The survey indicated that only three out of sixteen laboratories reported using an internal protein or antibody containing a relatively high-level SV (e.g., 0.5%) as a positive control [15].

A commentary from multiple companies was recently published on best practices for SV analysis to de-risk early product development of various therapeutic proteins. The key point in the commentary was that SV analysis “*should incorporate basic assay controls that establish that the specific method is scientifically sound and fit for the purpose of detecting and quantitating SVs*” [11]. The commentary discussed the possibility of using a common protein standard such as NISTmAb antibody, which contains low levels of SVs, as a basic assay control [11]. However, high variation of these low abundant SVs in NISTmAb evidenced by several cross-lab studies indicated it might not provide the best system suitability samples for the quality-control of SV analytical methods [1, 19]. For example, Zhang A *et al.* reported that though three laboratories were able to identify 21–23 low abundance (0.01%–0.1%) SVs in NISTmAb, only 12 SVs were commonly identified by all three laboratories, and the relative standard deviations of relative abundance of only 5 SVs were lower than 30% [19]. This variation appears to be inevitable based on a lack of standardization where minor changes in sample preparation, chromatographic separation and mass spectrometry data acquisition settings can potentially impact method performance.

Here, we proposed a spike-control approach, by including quality control samples routinely used in analytical procedure validation [8], as a system suitability or basic assay control for SV analytical methods. We spiked 0.025%–0.5% of one mAb to another highly homologous (> 95%) mAb to create mAb samples containing low yet accurate levels of “artificial” SVs. Both mAbs have been on US market for more than ten years, so their structures are fully characterized, and product quality is well controlled. Though not conceptually novel, this approach using well controlled system suitability samples can be used to practically demonstrate that sample preparation, chromatography separation, mass spectrometry data acquisition and bioinformatics analysis are needed to be fit for purpose for both qualitative and quantitative SV analysis. This work provides an initial procedural framework that can be further optimized for specific therapeutic protein products, especially mAbs, with procedural flexibility as analytical methods and techniques evolve.

Materials and Methods

Antibody Drug Products and Reagents

Two monoclonal antibody drug products, Herceptin® and Pertuzumab®, which contain Trastuzumab and Pertuzumab, respectively, were acquired from the US market (WEP Clinical, Morrisville, NC). Sequencing grade trypsin, Lys-C and Asp-N were purchased from Promega (Madison, WI). All other chemical reagents and lab supplies used in the protein digestion and LC–MS/MS were from Sigma-Aldrich (St. Louis, MO) unless otherwise specified. Filtered 18 mΩ DI water was supplied in house by a Millipore Milli-Q System (Bedford, MA).

Spike-Control Samples

Spike-control samples were made by spiking Trastuzumab into Pertuzumab to create mAb samples containing low yet accurate levels of “artificial” SVs. The two mAbs have > 95.5% amino acid sequence homology based on NCBI Blast sequence comparison.

Sample Preparation

A typical denaturation, reduction, and enzymatic digestion procedure was used for the sample preparation in the peptide mapping-based SV analysis. Briefly, 280 µg of each spike-control sample in 10 µL was mixed with 60 µL 7.2 M guanidine hydrochloride (pH = 7.5), and 2 µL 250 mM TCEP, and then incubated at 50°C for 30 min. The final concentration of guanidine hydrochloride was 6 M. Alkylation was performed by adding 4 µL 200 mM iodoacetamide in water and incubated at 25°C for 30 min, 4 µL 0.1 M DTT was added to the reaction mixture to quench the remaining iodoacetamide. For tryptic digestion, 13.4 µL alkylated mAb (total 30 µg protein) was diluted with 115 µL 100 mM Tris–HCl, pH 7.5 to reduce the concentration of guanidine and then mixed with 15 µL 0.1 µg/µL trypsin in 100 mM Tris–HCl buffer (pH = 7.5) and incubated for 4 h at 37°C. For Lys-C/Asp-N digestion, 13.4 µL alkylated mAb (total 30 µg protein) was diluted with 10 µL of 100 mM Tris–HCl buffer, pH 7.5, and then mixed with 12.5 µL 0.2 µg/µL Lys-C in 100 mM Tris–HCl buffer (pH = 7.5) and incubated for 2 h at 37°C. Then 78 µL of 100 mM Tris–HCl buffer (pH = 7.5) and 30 µL 0.02 µg/µL µg Asp-N was added to the mixture and incubated for overnight at 37°C. Finally, 4 µL of 20% formic acid was added to each digestion mixture to stop the reaction.

LC–MS/MS Method and Data Acquisition

The digested mAb samples (4 µg in 20 µL) were analyzed using a Waters ACQUITY UPLC system interfaced to a Thermo Scientific Orbitrap Exploris™ 480

mass spectrometer for the SV analysis. A Waters UPLC BEH C18 UPLC Column (300 Å, 1.7 µm particle size, 2.1 mm × 150 mm) was used for the peptide separation. Mobile phase A was 0.1% formic acid in water and mobile phase B was 0.1% formic acid in acetonitrile. Elution was started with isocratic 0.1% mobile phase B for 5 min, mobile phase B was increased to 35% over 75 min (programmed gradient curve type 5), then to 95% over the next 5 min (curve type 5). The column was then washed with 95% mobile phase B for 5 min and re-equilibrated in 0.1% mobile phase B for 20 min. Flow rate was 0.25 mL/min. Between two sample injections, one initial mobile phase blank run was performed to eliminate potential carryovers from the previous sample, which could potentially cause a false positive for the SV identification. The MS data were acquired on Thermo Scientific Orbitrap Exploris™ 480 mass spectrometer. Data acquisition was started with full scan at a resolution of 60 k (at m/z 400), followed by 5 data-dependent MS/MS scans. Detailed MS and MS/MS parameters are listed in Table I.

SV Identification and Quantitation

The acquired MS/MS data were first searched for SV using Sequence Variant Validators, a feature embedded in BYOS® software (Protein Metrics) to identify all potential SVs [13]. Common post-translational modifications along with all the possible amino acid substitutions resulting from single nucleotide change were included in the search. The SV candidates were identified by BYOS® software and were further verified by examining the MS1 mass accuracy, MS/MS fragmentation pattern, isotopic pattern, and the retention time shift of the SV peptide relative to its desired product form. The relative abundance of each SV was defined as the ratio of the peak area of its extracted ion chromatograms (XICs) from MS1 scan to that of the native peptide.

The spike-control samples were prepared by the principal investigator at the FDA and then delivered to an analytical site for analysis. Scientists who processed the samples, conducted the LC–MS/MS analysis and the bioinformatic analysis only knew that the samples were Pertuzumab spiked with unknown levels of another unknown antibody (*i.e.*, they were blinded and did not know there was Trastuzumab in the samples).

Results

SV Identification

mAb samples analyzed in the current study were Pertuzumab containing low, yet accurate levels of “artificial” SVs caused by spiked Trastuzumab. For the proof of concept study, peptides from Pertuzumab were considered as a desired model product (expected amino acid sequence, as defined in ICH Q6B

Table I Thermo Scientific Orbitrap Exploris™ 480 Mass Spectrometer MS and MS/MS Data Acquisition Parameters

MS1 Full Scan Parameters		Data Dependent MS2 Data Acquisition Parameters	
Orbitrap Resolution	60000	Isolation Window (m/z)	2
Scan Range (m/z)	300–2000	Isolation Offset	Off
RF Lens (%)	40	Collision Energy Mode	Fixed
AGC Target	Custom	Collision Energy Type	Normalized
Normalized AGC Target (%)	300	HCD Collision Energy (%)	28
Maximum Injection Time Mode	Custom	Orbitrap Resolution	15000
Maximum Injection Time (ms)	100	TurboTMT	Off
Micro scans	1	Scan Range Mode	Auto
Data Type	Profile	AGC Target	Custom
Polarity	Positive	Normalized AGC Target (%)	300
		Maximum Injection Time Mode	Custom
		Maximum Injection Time (ms)	200

Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products [7]), whereas Trastuzumab peptides that included a single amino acid difference at corresponding locations were considered to be SVs. Amino acid sequence comparison and alignment of the two mAb drugs Trastuzumab and Pertuzumab using the “Basic Local Alignment Search Tool” (BLAST, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) revealed four peptides containing a single amino acid SV in the spike-control samples (Fig. 1). Specifically, 24 K > R is in the light chain (LC), the other three variants 72 V > A, 79L > A and 97A > S are in the heavy chain (HC). The sequence of the four peptides containing SVs and the corresponding peptide from desired products is listed in Table II. These four SVs represent commonly reported SVs from commercially available protein reference standard, marketed therapeutic proteins, as well as SV caused by change of bioprocessing. For example, single amino acid SV K > R has been reported in NISTmAb at 0.02–0.04% [19], whereas V > A and A > S have been reported in antibodies on the US market or are being developed up to

0.15% [1, 19]. In an experimental bioprocessing study, alanine was reported to be the major mis-incorporated amino acid when bioreactor medium for Chinese hamster ovary (CHO) cells was starved for valine (SV V > A) [16].

For SVs identification, 0.025%, 0.1% and 0.5% Trastuzumab was spiked into Pertuzumab (w/w) and the aliquoted samples were digested concurrently by trypsin or Lys-C / Asp-N as described above. LC–MS/MS analysis was followed by bioinformatic analysis using BYOS® software that was able to identify all four single amino acid SVs in Pertuzumab spiked with 0.025%, 0.1% and 0.5% Trastuzumab. The XIC peak area of desired product peptides (from Pertuzumab) and single amino acid SVs (from Trastuzumab), as well as relative abundance of SVs, are listed in Table II. 24 K > R at light chain and 79L > A at heavy chain were identified in both trypsin and Lys-C + Asp-N digested samples, whereas 97A > S was only identified in trypsin digested sample and 72 V > A was only identified in Lys-C + Asp-N digested sample. This result highlighted the importance of using multiple protease

Fig. 1 Sequence alignment of the light chain (LC, (a)) and heavy chain (HC, (b)) of Trastuzumab (Herceptin) and Pertuzumab (Perjeta). Peptides with amino acid sequence differences were highlighted in red (digestion by trypsin) and blue (digestion by Lys-C/Asp-N) boxes.

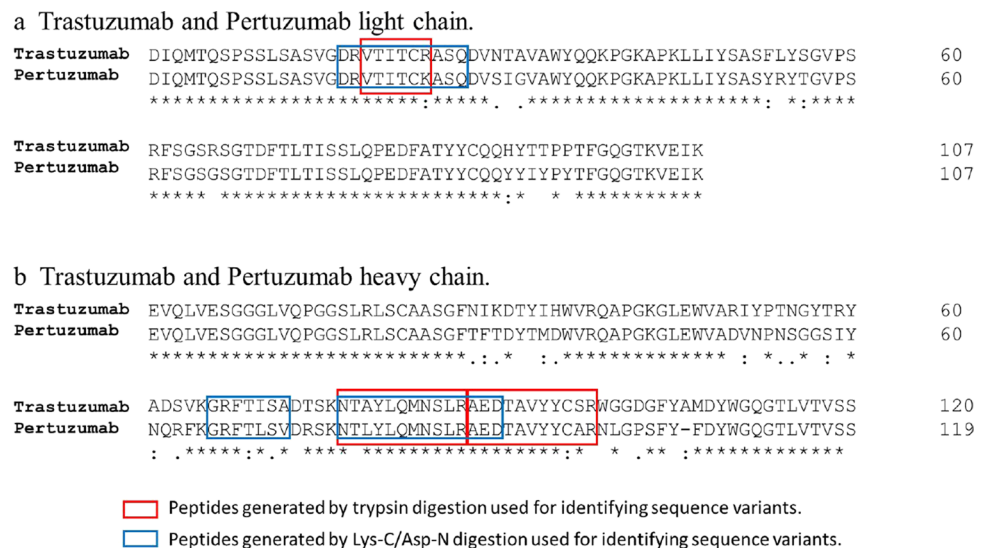


Table II Four SVs Identified in 0.1% and 0.5% Spiked Samples and their Relative Abundance

Sequence	Position	Location	Relative abundance (%)			XIC peak area		
			0.025% spiked	0.1% spiked	0.5% spiked	0.025% spiked	0.1% spiked	0.5% spiked
Trypsin Digestion								
AEDTAVYYCAR	97A > S	HC*				1.75E+09	1.80E+09	1.80E+09
AEDTAVYYCSR			0.02	0.08	0.4	3.50E+05	1.50E+06	7.20+E06
NLYLQMNLSLR	79L > A***	HC				9.70E+08	9.60E+08	9.40E+08
NTAYLQMNLSLR			0.03	0.18	1.05	2.90E+05	1.70E+06	1.00E+07
VTITCK***	24 K > R****	LC**				7.90E+08	8.00E+08	7.70E+08
VTITCR			0.06	0.26	1.53	4.70E+05	2.10E+06	1.20E+07
Lys-C/Asp-N Digestion								
DRVITITCK	24 K > R	LC				6.90E+09	6.30E+09	7.60E+09
DRVITITCRASQ			0.02	0.08	0.48	1.40E+06	5.00E+06	3.70E+07
GRFTLSV	72 V > A	HC				3.10E+09	3.40E+09	2.90E+09
GRFTLSA			0.03	0.15	1.19	9.20E+05	5.00E+06	3.50E+07
NLYLQMNLSLRAE	79L > A	HC				7.10E+09	7.10E+09	7.50E+09
NTAYLQMNLSLRAE			0.02	0.07	0.42	1.40E+06	5.00E+06	3.20E+07

*Heavy chain

**Light Chain

***This SV was identified by both Trypsin and Lys-C/Asp-N Digestion

****This SV was identified by both Trypsin and Lys-C/Asp-N Digestion

digestion procedures to maximize the possibility of identifying the potential SVs. Figure 2a and b illustrate the identification of LC 24 K > R and HC 79L > A SVs by MS/MS analysis for samples in trypsin and Lys-C + Asp-N digested peptides.

SV Quantitation: Relative Abundance

The extracted ion chromatogram (XIC peak area of desired product peptide from Pertuzumab) was very consistent among 0.025%, 0.1% and 0.5% Trastuzumab spiked samples from both trypsin and Lys-C/Asp-N digestion. As can be seen from Table II, in each digestion group, the XIC peak area of SV peptides (from Trastuzumab) and their relative abundance were roughly four-fold higher in 0.1% spiked samples compared to those in 0.025% spiked samples, and five-fold higher in 0.5% spiked samples compared to those in 0.1% spiked samples, which suggests the current workflow is suitable for relative quantitation of all four SVs without corresponding standards and internal standards. However, for the two SVs identified in both trypsin and Lys-C/Asp-N digested samples, their relative abundance calculated by the XIC peak area of SV peptide (from Trastuzumab) divided by desired product peptide (from Pertuzumab) were different between the two digestion groups. For example, relative abundance of light chain 24 K > R and heavy chain 79L > A were 0.08% and 0.07% in 0.1% spiked samples digested by Lys-C/Asp-N. However, the relative abundance of these two SVs were 0.26% and 0.18% in trypsin digested sample, respectively, which were higher than those in Lys-C/Asp-N digested samples. It is well known

that amino acid composition affects the ionization efficiency of peptides, which in turn has impact on the absolute signal intensity [9]. Alanine and polar amino acids serine and glutamine have been reported to have lower ionization efficiency [9], thus these three additional amino acids in the light chain 24 K > R SV peptide DRVITITCRASQ might contribute to a lower relative abundance of this SV when using Lys-C/Asp-N digestion (Table II). Similarly, an industrial research group also reported a significant difference in ionization efficiency between a SV peptide and corresponding peptide from the desired product [14]. The pattern suggested that it would be wise to conservatively interpretate relative abundance of SVs calculated by mass spectrometry signal intensity of peptides without the use of appropriate standards and internal standards.

SV Quantitation: Precision and Accuracy and Linearity

To demonstrate the current workflow's precision and accuracy for quantitative analysis of the identified SVs and the establishment of a corresponding analytical linearity range, eight aliquots of Pertuzumab were spiked with 0.01%, 0.025%, 0.05%, 0.1%, 0.25%, 0.5%, 1% and 2% of Trastuzumab, then denatured, reduced, alkylated and digested with trypsin as described above. These eight samples defined as standards were analyzed as calibrators. Quality control samples defined as standards were prepared similarly with Pertuzumab spiked with 0.025% 0.1% and 1% Trastuzumab. All calibrators and quality control standard samples were analyzed within same

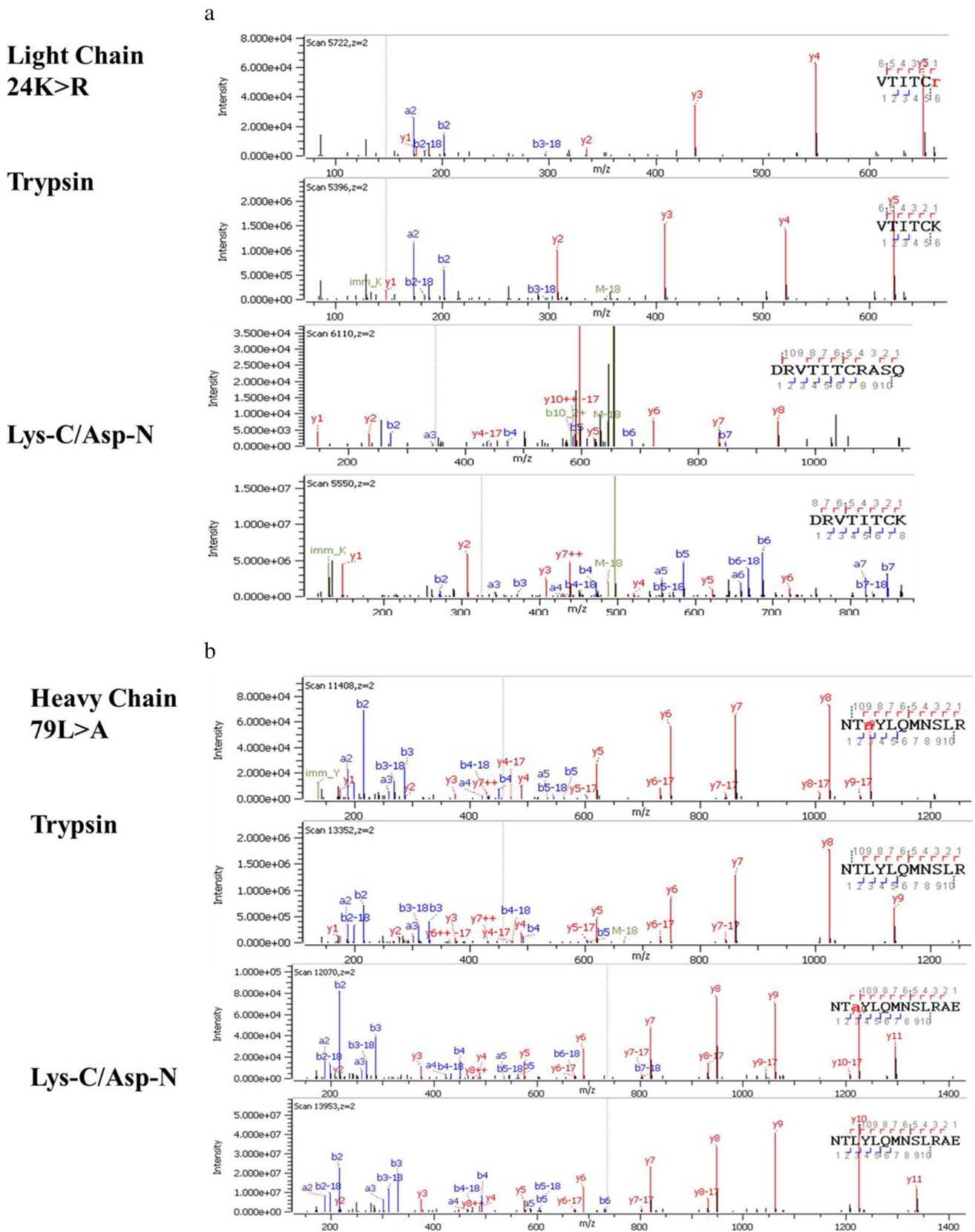


Fig. 2 Representative sequence variants identified by LC–MS/MS in Pertuzumab spiked with 0.025% Trastuzumab: **(a)** identification of single amino acid substitution 24 K > R on light chain by MS/MS spectrum in trypsin and Lys-C/Asp-N digested peptides; **(b)** identification of single amino acid substitution 79L > A on heavy chain by MS/MS spectrum in trypsin and Lys-C + Asp-N digested peptides.

LC–MS run as described above. The XIC peak areas of three SVs identified from trypsin digested peptides, LC 24 K>R (VTITCR), HC 79L>A (NTAYLQMNSLR) and HC 97A>S (AEDTAVYYCSR), were plotted against spiking levels (w/w) to demonstrate analytical range and linearity. As can be seen from Fig. 3, the calibration curves were linear for all three SVs from 0.01% to 2%. The analytical range covers the relative abundance of most SVs reported in the literature [1, 16, 19]. The analytical range was established by evaluating QC standard samples over the range bracketed by standards.

Accuracy and precision at three levels, 0.025%, 0.1% and 1% for all three SVs were determined based on the calculated relative abundance of the corresponding SVs, LC 24 K>R, HC 79L>A and HC 97A>S, in the quality control samples. The accuracy values were within the guidance specification range of 85–115% of the nominal concentration, the relative standard deviation (RSD) for each concentration level of each SV did not exceed 15% (Table III).

Discussion

SVs are a broad type of process related impurity, and their levels could be significantly elevated as a result of depletion of critical amino acid nutrients [6, 16]. The presence of elevated SVs contributes to the heterogeneity of the final

drug products and can also impact the drug efficacy and safety if the amino acid substitution is located in the biological binding regions or introduces a non-human amino acid sequence. Therefore, a fit-for-purpose analytical procedure should be developed to comprehensively characterize SVs in the final drug product and demonstrate comparable levels of SVs, if existing, in biosimilars [4]. As stated by the FDA in “Development of Therapeutic Protein Biosimilars: Comparative Analytical Assessment and Other Quality-Related Considerations Guidance for Industry [4]”, these comparative methods “do not necessarily need to be validated,” but “should be scientifically sound, fit for their intended use, and provide results that are reproducible and reliable.” Product knowledge including structure and heterogeneity of the desired product along with known and potential impurities are crucial for the selection and evaluation of the analytical methods [4].

However, unlike other analytical procedures for structural characterization, identification and quantitation of SVs in recombinant protein drug products, is truly a journey of exploring “unknown unknowns.” The theoretical number of SVs caused by unintended amino acid substitution may be high. Though experimental research of systematic amino acid starvation and the review of the scientific literature has indicated that certain single amino acid substitution such as

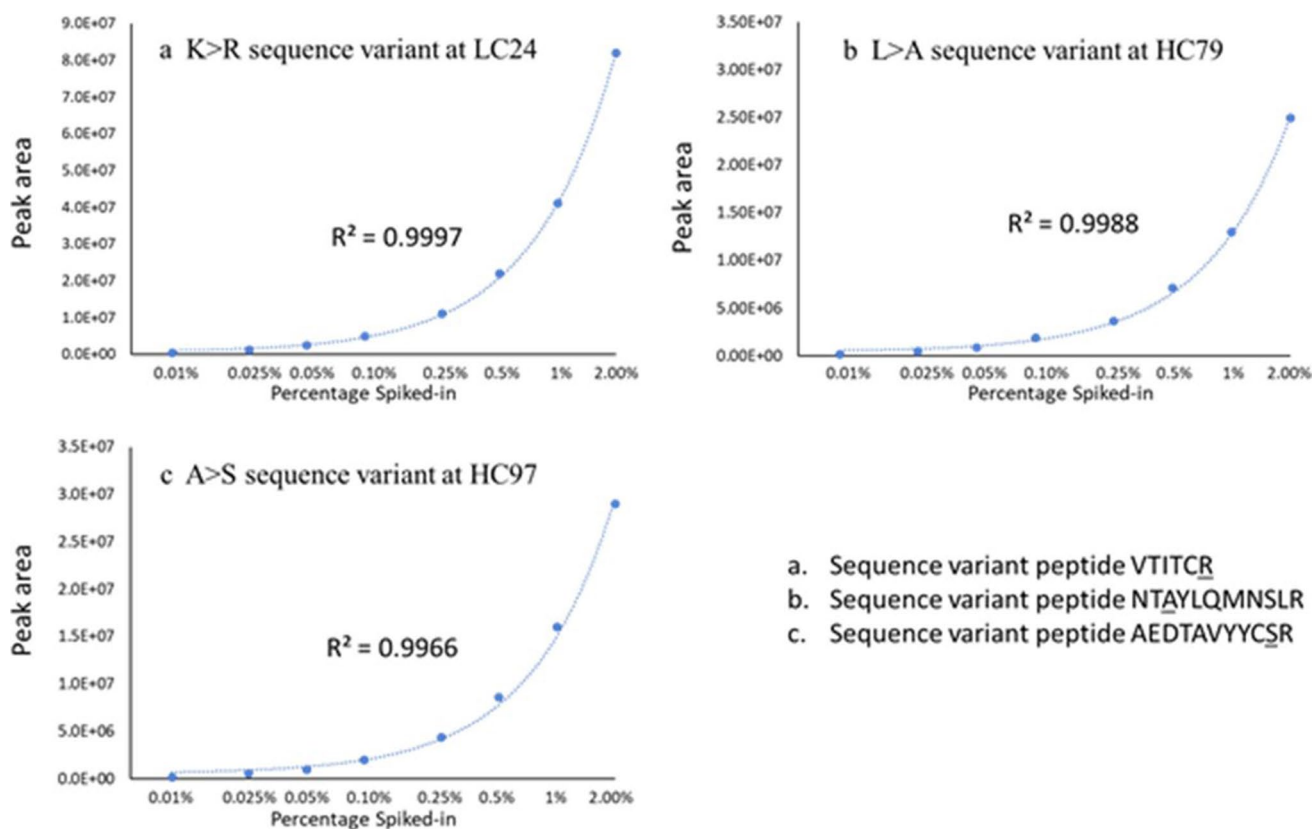


Fig. 3 Agreement of linearity between percentage spiked and XIC area was observed in all three sequence variants (K>R in LC, L>A and A>S in HC).

Table III Intra-day Accuracy and Precision of Three Selected SVs

Sequence	Position	Location	Accuracy (%)			Precision (% R.S.D.)		
			0.025% spiked	0.1% spiked	1% spiked	0.025% spiked	0.1% spiked	1% spiked
AEDTAVYYCAR	97A>S	HC*						
AEDTAVYYC S R			94.10	105.49	103.73	5.98	4.58	1.69
N T LYLQMNSLR	79L>A	HC						
N A YLQMNSLR			98.37	95.60	101.47	6.42	2.81	1.97
VTITCK***	24 K>R	LC**						
VTITC R			102.89	99.58	98.07	6.37	4.96	3.43

*Heavy chain

**Light Chain

C>Y, Y>F and S>N could be more prevalent in mAbs [16, 19], prediction of SV “hot spots” in a given recombinant protein is not currently possible. Many biopharmaceutical companies optimized their LC–MS/MS methods for SV analysis by using newer versions of mass spectrometry instruments, extending LC separation of peptides, and applying multiple enzymes digestion [10, 14, 20]. These enhanced characterization practices are also reflected in recent BLA submissions to the Agency (proprietary information). A certain degree of method qualification, or at least appropriate basic assay controls, should be performed to demonstrate that performance characteristics and detection parameters, such as sensitivity and resolution, are optimized and therefore sample results are reliable and accurate data can be processed and interpreted [3].

Yang Y *et al.* and Brady LJ *et al.* described similar spike-control approaches for evaluating mass spectrometry-based sequence analysis as we are reporting here [2, 17]. In the first study, 0.5–5.0% of a recombinant humanized anti-HER2 mAb (rhUmAb HER2) was spiked to another unrelated recombinant humanized mAb [17]. In the other study, two mAbs of the same putative amino acid sequence, but manufactured differently (*e.g.* different cell lines, production scale) were mixed in a series of samples and analyzed [2]. However, the mixture of these in-house manufactured mAbs might not be the idealist spike-control samples as their inherent micro-heterogeneity, and process related impurities were not fully characterized and subjected to change when manufactured in different scales [2, 17]. Use of a reference standard such as NISTmAb, as discussed above, might be an alternative approach as it is currently used by many laboratories to benchmark instrument and analytical method performance. But inconsistent results of SV identification and quantitation by multiple laboratories have indicated it might be a better system suitability control for other analytical methods than SVs analysis [1, 19]. To this end, we spiked low levels of one fully characterized mAb to another fully characterized highly homologous mAb to investigate whether the analytical procedure is suitable for the identification of “unknown” SVs and the quantitation of the identified SVs.

At the identification stage of the current study, we spiked 0.025%, 0.1% and 0.5% Trastuzumab to Pertuzumab. The spike-in levels were designed and based on a recently published paper that described an evidence-based sequence variant control limit for recombinant therapeutic protein development [19], which proposed a general SV control limit of 0.1% at individual amino acid sites. Currently, there is no consensus among industry and regulators on scientifically rational control limits of SVs during product and process development of recombinant protein drugs. This lack of consensus is largely due to the extreme heterogeneity of SVs, non-harmonized SV analysis practice among industry members, and limited prior knowledge on the functional significance (immunogenicity, safety and efficacy) of SVs. The 0.1% control limit was proposed because it represents a ~tenfold increase relative to the biological noise level of SVs in natural biologic proteins [18] and a majority of the SVs identified across the 14 approved therapeutic proteins and NISTmAb are in the range of 0.01–0.1% [1, 19]. During identification, 0.1% was used as a starting point, however a 0.025% spike-control sample was also included in the identification stage of the study to assess a more stringent standard.

The in-house data presented here, along with information on analytical methods for SVs, biological mechanism, and functional significance, as documented in literature and regulatory submissions, have led to several topics of interest for further investigation and discussion. 1) Selection of spike-control proteins. Spike control proteins should be highly homologous to the recombinant protein drug to be assessed, and the pair of spike control proteins (in this case Trastuzumab and Pertuzumab) should be also highly homologous. 2) Determination of critical steps in SV analysis, including sample preparation, digestion, LC separation, MS/MS and MS data acquisition and bioinformatic analysis that need to be optimized to achieve desirable sensitivity yet minimize false positives. 3) Selection of SV quantitation approach. The level of SVs is typically reported as relative abundance but in at least one case absolute quantitation was performed using a stable isotope labeled internal standard in a regulatory

submission (proprietary information). Whether relative or absolute quantitation, the selection of calibrators and internal standards (if used) should be justified based on development stage. 4) Determination of the impact of SVs on product safety, efficacy, and quality. There might be “hot spots” for SVs in certain recombinant protein drugs and some SVs might have functional significance (*e.g.*, decreased stability), which have been reported in literature [16, 20] and documented in regulatory submissions (proprietary information).

We acknowledge the limitation of our study, that this spike-control approach using two highly homologous mAbs, does not necessarily establish suitability of the system or the procedure for identifying and quantitating SVs in all recombinant protein drug products. However, the steps taken provide an approach for developing a procedural platform that sponsors and regulators can use to guide future characterization approaches. Overall, the current study highlights that analytical method(s) for SVs should incorporate control sample(s) which demonstrate that sample preparation, LC separation, mass spectrometry and bioinformatic software are purposeful as intended and suitable for the identification and quantitation of SVs at a pre-determined threshold.

Acknowledgements This work was funded by the U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Office of Pharmaceutical Quality. The authors thank Dr. Sarah Rogstad, Senior Scientific Advisor, Office of Testing and Research, Office of Pharmaceutical Quality, FDA/CDER, for critical comments.

Declarations

Conflicts of Interest None of the authors has any personal or financial conflict of interest or has entered into any agreement that could potentially result in a conflict of interest due to their access to the data presented here, or on their ability to analyze the data independently, or to prepare and publish future manuscripts, based wholly or partially on these data.

Disclaimer This publication reflects the views of the authors and should not be construed to represent FDA’s views or policies.

References

- Borisov OV AM, Carroll JA, Brown PW. (2015) Sequence variants and sequence variant analysis in biotherapeutic proteins. , in *State-of-the-art and emerging technologies for therapeutic monoclonal antibody characterization Volume 2 Biopharmaceutical characterization: the NISTmAb case study* pp 63–117, ACS Publications.
- Brady LJ, Scott RA, Balland A. An optimized approach to the rapid assessment and detection of sequence variants in recombinant protein products. *Anal Bioanal Chem.* 2015;407:3851–60.
- Easton RL. Structural characterization methods for biosimilars: fit-for-purpose, qualified or validated. *Generics and Biosimilars Initiative Journal (GaBI Journal).* 2022;11:41–4.
- FDA (2019) Development of Therapeutic Protein Biosimilars: Comparative Analytical Assessment and Other Quality-Related Considerations Guidance for Industry.
- Frye C, Deshpande R, Estes S, Francissen K, Joly J, Lubiniecki A, Munro T, Russell R, Wang T, Anderson K. Industry view on the relative importance of “clonality” of biopharmaceutical-producing cell lines. *Biologicals.* 2016;44:117–22.
- Guo D, Gao A, Michels DA, Feeney L, Eng M, Chan B, Laird MW, Zhang B, Yu XC, Joly J, Snedecor B, Shen A. Mechanisms of unintended amino acid sequence changes in recombinant monoclonal antibodies expressed in Chinese Hamster Ovary (CHO) cells. *Biotechnol Bioeng.* 2010;107:163–71.
- ICH (1999) ICH GUIDELINE Q6B: SPECIFICATIONS: Test procedures and acceptance criteria for biotechnological/Biological products.
- ICH (2022) International Council for harmonisation of technical requirements for pharmaceuticals for human use: Validation of Analytical Procedures Q2(R2).
- Liigand P, Kaupmees K, Kruve A. Influence of the amino acid composition on the ionization efficiencies of small peptides. *J Mass Spectrom.* 2019;54:481–7.
- Lin TJ, Beal KM, Brown PW, DeGruttola HS, Ly M, Wang W, Chu CH, Dufield RL, Casperson GF, Carroll JA, Friese OV, Figueroa B Jr, Marzilli LA, Anderson K, Rouse JC. Evolution of a comprehensive, orthogonal approach to sequence variant analysis for biotherapeutics. *MAbs.* 2019;11:1–12.
- Lund A, Ren D, Rogers RS, Rouse JC, Yu XC, Valliere-Douglass JF. Scientific Best Practices for Primary Sequence Confirmation and Sequence Variant Analysis in the Development of Therapeutic Proteins. *J Pharm Sci.* 2021;110:619–26.
- Rathore AS, Zydney AL, Anupa A, Nikita S, Gangwar N. Enablers of continuous processing of biotherapeutic products. *Trends Biotechnol.* 2022;40(7):804–15.
- Teams PMRDaCS (2021) Improving data review efficiency using automated labeling with byos validators: Sequence variant and post-translational modification analysis.
- Thakur A, Nagpal R, Ghosh AK, Gadamshtetty D, Nagapattinam S, Subbarao M, Rakshit S, Padiyar S, Sreenivas S, Govindappa N, Pai HV, MelarkodeSubbaraman R. Identification, characterization and control of a sequence variant in monoclonal antibody drug product: a case study. *Sci Rep.* 2021;11:13233.
- Valliere-Douglass J, Marzilli L, Deora A, Du Z, He L, Kumar SR, Liu YH, Mueller HM, Nwosu C, Stults J, Wang Y, Yaghmour S, Zhou Y. Biopharmaceutical Industry Practices for Sequence Variant Analyses of Recombinant Protein Therapeutics. *PDA J Pharm Sci Technol.* 2019;73:622–34.
- Wong HE, Huang CJ, Zhang Z. Amino Acid Misincorporation Propensities Revealed through Systematic Amino Acid Starvation. *Biochemistry.* 2018;57:6767–79.
- Yang Y, Strahan A, Li C, Shen A, Liu H, Ouyang J, Katta V, Francissen K, Zhang B. Detecting low level sequence variants in recombinant monoclonal antibodies. *MAbs.* 2010;2:285–98.
- Zhang Z, Shah B, Bondarenko PV. G/U and certain wobble position mismatches as possible main causes of amino acid misincorporations. *Biochemistry.* 2013;52:8165–76.
- Zhang A, Chen Z, Li M, Qiu H, Lawrence S, Bak H, Li N. A general evidence-based sequence variant control limit for recombinant therapeutic protein development. *MAbs.* 2020;12:1791399.
- Zhang Z, Shah B. Limited Proteolysis Coupled with Mass Spectrometry for Simultaneous Evaluation of a Large Number of Protein Variants for Their Impact on Conformational Stability. *Anal Chem.* 2021;93:14263–71.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.