MABS
2020, VOL. 12, NO. 1, e1791399 (10 pages)
https://doi.org/10.1080/19420862.2020.1791399

Taylor & Francis
Taylor & Francis Group

REPORT

A general evidence-based sequence variant control limit for recombinant therapeutic protein development

Aming Zhang[a], Zhengwei Chen[a], Meinuo Li[a], Haibo Qiu[a], Shawn Lawrence[b], Hanne Bak[b], and Ning Li[a]

[a]Analytical Chemistry, Regeneron Pharmaceuticals, Inc., Tarrytown, New York, USA; [b]Preclinical Manufacturing and Process Development, Tarrytown, New York, USA

ABSTRACT

Sequence variants (SVs) resulting from unintended amino acid substitutions in recombinant therapeutic proteins have increasingly gained attention from both regulatory agencies and the biopharmaceutical industry given their potential impact on efficacy and safety. With well-optimized production systems, such sequence variants usually exist at very low levels in the final protein products due to the high fidelity of DNA replication and protein biosynthesis process in mammalian expression systems such as Chinese hamster ovary cell lines. However, their levels can be significantly elevated in cases where the selected production cell line has unexpected DNA mutations or the manufacturing process is not fully optimized, for example, if depletion of certain amino acids occurs in the cell culture media in bioreactors. Therefore, it is important to design and implement an effective monitoring and control strategy to prevent or minimize the possible risks of SVs during the early stage of product and process development. However, there is no well-established guidance from the regulatory agencies or consensus across the industry to assess and manage SV risks. A question frequently asked is: What levels of SVs can be considered acceptable during product and process development, but also have no negative effects on drug safety and efficacy in patients? To address this critical question, we have taken a holistic approach and conducted a comprehensive sequence variant analysis. To guide biologic development, a general SV control limit of 0.1% at individual amino acid sites was proposed and properly justified based on extensive literature review, SV benchmark survey of approved therapeutic proteins, and accumulated experience on SV control practice at Regeneron.

ARTICLE HISTORY
Received 18 March 2020
Revised 21 May 2020
Accepted 8 June 2020

KEYWORDS
Amino acid depletion; amino acid misincorporation; Chinese hamster ovary cell line; control limit; critical quality attributes; mammalian cell culture; mass spectrometry; monoclonal antibody; sequence variant; therapeutic protein

Introduction

Recombinant therapeutic proteins, particularly monoclonal antibodies (mAbs), have become a major class of biotherapeutics to treat various human diseases.[1,2] To date, more than 90 mAbs have been approved.[3] To ensure product quality, attributes that can potentially affect drug safety and efficacy, known as critical quality attributes, need to be well characterized and monitored during the drug development and manufacturing process. Among the various product quality attributes, sequence variants (SVs) resulting from unintended amino acid substitution are an increasing concern and have been discussed by both the biopharmaceutical industry and regulatory agencies.[4–6] Such SVs have been shown to exist in both natural and recombinant proteins,[7] and are believed to be caused by a number of mechanisms including DNA mutations during replication, and transcriptional and translational errors during the protein biosynthesis process.[5] Thanks to the high fidelity of biologic systems, which evolved to prevent the occurrence of such spontaneous errors, the SVs are usually present at a very low level (<0.1%) in natural biologic proteins.[5,8] However, during therapeutic protein drug development, companies continuously aim to increase the protein titer and process productivity to meet global demand and reduce the cost of goods for expanded patient access. This has led to the wide use of the so-called intensified bioreactor manufacturing systems, which are designed to maximize cell density and specific productivity for the target therapeutic proteins during the cell culture process. Such intensified production systems can impose higher than normal expression machinery stress to the production cell lines. If not fully optimized, elevated levels of SVs could be generated in the protein products.[9–11] In addition, to further increase the product titer, cell line development usually goes through multiple rounds of selection with increasing selective stresses to find the top-producing cell clone. This selection process could potentially introduce DNA mutations to the cell lines. If not properly screened, it could lead to unexpectedly high levels of SVs in the final drug products.[7,12,13]

Given these concerns regarding how elevated SVs might affect drug quality, both the industry and regulatory agencies have started to pay more attention to SVs. Over the past decade, substantial efforts and resources have been invested across the industry to better understand the causes of SVs and their control in biologic development. As a result of these collective efforts, several control strategies have been developed and proposed by different companies to best monitor and mitigate the SV issue during their product and process development.[6,8,13] As expected, these proposed strategies highlighted the importance of a multi-assay,

CONTACT Haibo Qiu haibo.qiu@regeneron.com Analytical Chemistry, Regeneron Pharmaceuticals, Inc., , Tarrytown, New York 10591-6707, USA; Aming Zhang aming_102@hotmail.com CMC Analytical, GlaxoSmithKline, Collegeville, Pennsylvania 19426, USA

This article has been republished with minor changes. These changes do not impact the academic content of the article.
Supplemental data for this article can be accessed here.

multi-tier SV screening approach to guide process development from early cell line selection to small-scale cell culture process development to scale-up confirmation. Together, these strategies have provided a valuable and industry-wide framework and high-level guidance toward the goal of establishing some common best practices in terms of SV control. However, to reach this goal, we also have to acknowledge that there is still a lack of clarity and consensus across the industry, and probably within regulatory agencies as well, on a variety of important aspects. These include: 1) the selection and combinatory use of multiple SV-relevant analytical technologies (e.g., next-generation sequencing-based DNA or RNA sequencing, liquid chromatography (LC)-mass spectrometry (MS)/MS, surrogate amino acid analysis); 2) the selection of stage-(s) and degree to implement SV monitoring and control during the product and process development considering both overall control strategy effectiveness and development timeline; 3) the appropriate assessment of SV risk on product safety and efficacy; 4) determination of a rational SV control limit or acceptable level in process development and in the final drug products; and 5) reporting of the SV data in regulatory filing. To fill some of these knowledge gaps, the results of a survey of industry practices on SV analysis and control in their biologic development were published recently by the International Consortium for Innovation & Quality in Pharmaceutical Development.[4] In the survey, one of the most critical questions asked is the level of SVs that individual companies set as an action limit (or control target) for their product and process development. This is also the central question that needs to be first answered before developing any SV control strategy for therapeutic protein development.

Here, we attempt to address this important question by proposing and justifying an appropriate SV control limit for recombinant therapeutic protein development from three different perspectives: 1) review of the typical ranges of SV observed from different expression systems based on major underlying causes; 2) survey of the SV benchmark for approved therapeutic protein drugs; and 3) evaluation of process capability in terms of the SV control. Specifically, we conducted an extensive literature review to summarize the typical ranges of SVs observed from three different scenarios: 1) a natural biologic system; 2) a recombinant protein production system with well-optimized process and conditions; and 3) a production system with identified process deficiencies leading to elevated SVs. We also performed an SV benchmark survey on 15 selected commercial therapeutic protein drugs (mainly mAbs) that could potentially represent most approved biologics developed in the past two decades. Finally, we also evaluated the manufacturing process capability in the control of SVs for multiple Regeneron mAbs when a designed SV control strategy was incorporated into the whole manufacturing process development. Through such a holistic approach, an evidence-based SV control limit was proposed and justified for the potential adoption of biologic development by the biopharmaceutical industry.

Results

Literature review of SV ranges under different scenarios

In the past decade, literature reports pertaining to SVs have been mounting, probably driven by the increased number of therapeutic proteins and associated developmental activities in the pipeline and the advances in analytical technologies, particularly in high-resolution MS instrumentation and data processing bioinformatics. A comprehensive review of the relevant literature is beneficial to summarize the necessary background information to establish a rational target for SV control. In general, SVs resulting from unintended amino acid substitution in both natural and recombinant proteins have been suggested to be attributed to three broad mechanisms: 1) DNA mutation in the coding sequence; 2) transcriptional error from DNA to mRNA; and 3) translational error from mRNA to protein sequence. As reviewed in literature, due to the finite fidelity in the DNA replication and protein biosynthesis process, unintended amino acid substitution occurs naturally in any natural biologic system in a spontaneous manner.[5] However, in normal biologic systems, the chance for such spontaneous errors to occur is expected to be extremely low, with the range of 10^{-11}-10^{-8} during DNA replication,[14] 10^{-6}-10^{-4} during mRNA transcription[15,16] and 10^{-5}-10^{-4} during protein translation,[17] as illustrated in Figure 1. In prokaryotic systems, like *E. coli*, the translational error could be higher, up to 10^{-3}, or 0.1% of SVs relative to its native form.[18] Due to the spontaneous nature of the events, these very low levels of SVs resulting from transcriptional or translational errors are usually inevitable, and thus can be considered as the biologic noise in protein expression.

For recombinant therapeutic proteins, this situation will be slightly different as the production systems are designed and developed to achieve maximum productivity. This goal is

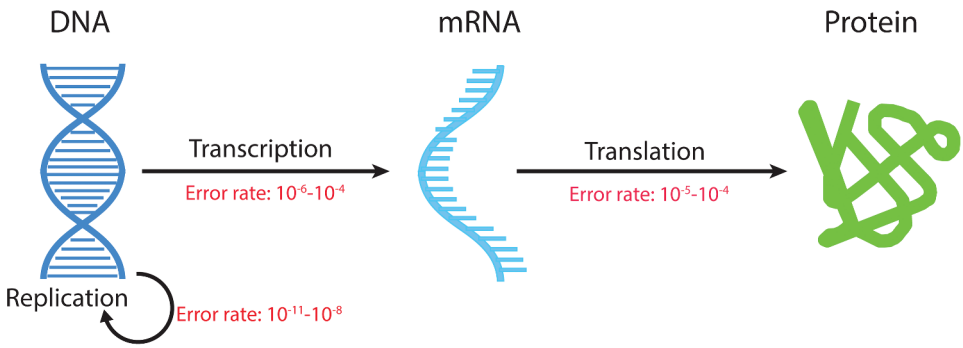

Figure 1. Elucidation of the central dogma and typical error rate in each step.

usually achieved by selecting high-producing cell lines, increasing both cell density and specific productivity for the target proteins, and extending cell culture time in bioreactors. Under these conditions, the expression cell lines are expected to experience higher stress than in normal biologic system in nature, which could lead to elevated levels of transcriptional and translational errors and the resulting SVs. Therefore, to establish an appropriate SV control limit for recombinant protein production, it is critical to understand the level of SV elevation relative to the biologic noise under the expected expression stress. To achieve that, we conducted an extensive literature review to categorize the levels of SVs observed under three general scenarios: 1) a biological system in nature; 2) a fully optimized recombinant protein production system; and 3) a less optimized production system with identified deficiencies causing elevated SVs.

The review of literature is summarized in Table 1. For the biological system in nature, Zhang et al.[7] selected the endogenous human serum albumin protein that was purified from three healthy human subjects and analyzed their SV profiles. A total of 66 SVs resulting from 15 different amino acid substitution types were identified. The levels of these SVs were measured in the range of 0.0013–0.0103%, consistent with the expected translational error of 10^{-5}-10^{-4} in the natural biologic system. The results suggested that these naturally occurring SVs were mainly caused by the spontaneous translational error. For the recombinant therapeutic protein production systems, the authors also surveyed the SV profiles of multiple therapeutic proteins selected from their development pipeline (presumably all with fully optimized production systems), including six small and Fc fusion proteins produced using an *E. coli* system, and several (exact number unspecified) mAbs from a Chinese hamster ovary (CHO) expression system.[7] The averaged SV levels from the same amino acid substitution type were observed to be in the range of 0.001–0.406% for the 6 proteins from the *E. coli* expression system, and 0.001–0.08% for the mAbs from the mammalian CHO expression system. The observed SV levels were higher in the proteins expressed by the *E. coli* system compared to those expressed by the CHO system, likely due to the higher translational error rate (10^{-5}-10^{-3}) expected in the prokaryotic expression system than in the mammalian system (10^{-5}-10^{-4}). In addition, for both expression systems, the upper range of the measured SVs (0.406% and 0.08% for the *E. coli* and CHO systems, respectively) is higher by approximately 4–8 fold compared to the expected upper level from spontaneous translational error rates (0.1% and 0.01% for the *E. coli* and CHO systems, respectively). The level of SV elevation observed in recombinant therapeutic proteins may reflect the degree of rate increase in translational error within the production cell lines when experiencing higher than normal expression machinery stress.

When we set a general SV control limit for recombinant protein manufacturing process development, it is important to take into account the fact that SVs in recombinant proteins will likely be elevated compared to naturally expressed proteins due to the high titer demand and resultant expression stress. The following question is how much margin should be set as an appropriate SV limit to control for therapeutic proteins. A rational margin should not only sufficiently consider the

Table 1. Summary of the reported SV levels in natural and recombinant proteins under three scenarios.

Occurrence Scenario	Expression System	Reported SV Level	Translational Error Rate	Comment	Company	Reference
Nature biological system	Human	**0.0013–0.0103%**	**10^{-5}-10^{-4}**	Level consistent with translational error rate	Amgen	Zhang Z. et al.[7]
Well optimized recombinant therapeutic protein production system	CHO	**0.001–0.080%**	**10^{-5}-10^{-4}**	Level higher than expected translational error rate	Amgen	Zhang Z. et al.[7]
	E. coli	**0.001–0.406%**	**10^{-5}-10^{-3}**		Amgen	Zhang Z. et al.[7]
Under-optimized recombinant protein production system with identified deficiency causing elevated SVs	CHO	**1–27%**	**10^{-5}-10^{-4}**	DNA mutation	Genentech	Harris RJ. et al.[12]
	CHO	**~2%**	**10^{-5}-10^{-4}**	DNA mutation	AbbVie	Zhang S. et al.[13]
	CHO	**7.8–9.9%**	**10^{-5}-10^{-4}**	DNA mutation	BMS	Fu J. et al.[19]
	CHO	**~5-50%**	**10^{-5}-10^{-4}**	DNA mutation	Pfizer	Degruttola H. et al.[20]
	E. coli	**0.6–1.9%**	**10^{-5}-10^{-3}**	Rare codon	BMS	Huang Y. et al.[21]
	E. coli	**0.1–1.0%**	**10^{-5}-10^{-3}**	Rare codon	Amgen	Hutterer KM[22]
	CHO	**1.4–6.7%**	**10^{-5}-10^{-4}**	AA depletion	Biogen	Wen DY. et al.[23]
	CHO	**0.2–1.5%**	**10^{-5}-10^{-4}**	AA depletion	Genentech	Feeney L. et al.[24]
	CHO	**0.7%**	**10^{-5}-10^{-4}**	AA depletion	Amgen	Raina M. et al.[25]
	E. coli	**0.8–49.2%**	**10^{-5}-10^{-3}**	AA depletion	Sanofi	Ni J. et al.[26]

normal level of SV elevation in recombinant therapeutic protein production with well-optimized systems, but also allow the detection of abnormal levels of SVs when a manufacturing process has certain deficiencies that cause higher levels of SVs. The SV survey done by Zhang et al.[7] suggested ~4–8 fold of SV elevation in therapeutic proteins using presumably well-optimized production systems (Table 1). To address the other part of the question, we also reviewed the reported SVs in therapeutic proteins produced from under-optimized production systems, for which a specific factor was identified as the cause of the elevated SVs. As summarized in Table 1, a variety of mechanisms, including DNA mutation, codon usage and nutrient depletion, could exist and lead to elevated SVs in therapeutic protein production systems. The level of these SVs spans a very wide range, from 0.1% to ~50%. Interestingly, the low end of the range (0.1%) is slightly higher than the upper level of SVs (0.08%) observed in the therapeutic proteins produced from CHO as surveyed by Zhang et al.[7] These observations indicate that 0.1% may be a good limit to separate the level of SVs occurring between well-optimized and under-optimized production systems.

SV benchmark analysis of approved therapeutic proteins

The review of SVs reported in literature as shown in Table 1 provides the general ranges of SV levels occurring under different expression systems and scenarios. However, it is important to note that most of the SVs reported in literature were probably measured in selected model or therapeutic proteins under development. For these proteins, we can anticipate that the processes used for manufacturing may not be fully optimized to meet all the product quality requirements, including the control of SVs. To understand the level of SVs present in well-established manufacturing processes, we performed an SV benchmark survey in 15 approved mAb and Fc fusion protein drugs (Table 2). These protein drugs were approved during 1997–2017. They were developed and manufactured by 11 pharmaceutical companies using three major mammalian expression systems, including CHO, NS0 and Sp2/0 mouse hybridoma cell lines. Therefore, a survey of the SVs in these 15 protein drugs could presumably provide a representative SV benchmark for the approved therapeutic protein drugs. In addition, the NIST mAb, which has been extensively characterized was also included for the LC-MS method verification and SV benchmark analysis.

Before the SVs were analyzed for the 15 approved therapeutic proteins, we first evaluated the performance of the LC-MS method developed internally for SV analysis. The NIST mAb was used for the method evaluation because it has also been well characterized, including the analysis of SV profile by two independent laboratories.[5] A summary of results comparing the SVs identified by our LC-MS method with the reported SVs by two separate laboratories is shown in Table 3. The results suggested that all three laboratories were able to detect and identify low-level SVs in the range of 0.01–0.1%. However, the sets of SVs identified by the three laboratories did not completely overlap with each other. As shown in Figure 2, all three testing laboratories identified a very similar number (21–23) of SVs in the NIST mAb. However, only 12 of them were commonly identified by all three testing laboratories. These results suggested that there is a large method-based variation in detecting these low-level SVs. This variation, however, is considered inevitable, given the extremely low abundance of the SVs, and the fact that any small differences in the chromatography separation and/or MS data acquisition settings across the testing laboratories can

Table 2. Approved therapeutic proteins selected for the SV benchmark survey

Drug Trade Name	International non-proprietary Name	IgG Subclass	First US Approval Year	Company	Host Cell Platform
N/A	NIST antibody	IgG1	NA	MedImmune	NS0
Rituxan	Rituximab	IgG1	1997	Biogen	CHO
Herceptin	Trastuzumab	IgG1	1998	Genentech	CHO
Remicade	Infliximab	IgG1	1998	J&J	Sp2/0
Humira	Adalimumab	IgG1	2002	AbbVie	CHO
Xolair	Omalizumab	IgG1	2003	Amgen	CHO
Orencia	Abatacept	CTLA4-IgG1	2005	BMS	CHO
Soliris	Eculizumab	IgG2/4k	2007	Alexion	NS0
Yervoy	Ipilimumab	IgG1	2011	BMS	CHO
Cyramza	Ramucirumab	IgG1	2014	Eli Lilly	NS0
Repatha	Evolocumab	IgG2	2015	Amgen	CHO
Keytruda	Pembrolizumab	IgG4	2015	Merck	CHO
Portrazza	Necitumumab	IgG1	2015	Eli Lilly	NS0
Cosentyx	Secukinumab	IgG1	2015	Novartis	CHO
Darzalex	Daratumumab	IgG1	2015	J&J	CHO
Imfinzi	Durvalumab	IgG1	2017	AstraZeneca	CHO

Table 3. Comparison of SV identification and quantification across three testing laboratories.

Amino Acid Substitution	Chain	Position	Relative Percentage (%) Regeneron Lab	External Lab 1	External Lab 2
A→T	HC	A51		0.01	
	HC	A132	0.01	0.01	0.009
	HC	A144		0.02	
	LC	A192	0.01		
	HC	A381	0.01	0.02	
G→D	HC	G125	0.01	0.01	0.01
	HC	G141	0.03	0.04	0.01
	LC	G156			0.02
	HC	G284	0.02	0.03	0.03
	LC	G199	0.07	0.07	0.02
H→D	LC	H197	0.01		
R→K	LC	R60	0.01	0.01	
	LC	R210			0.09
K→R	HC	K150	0.04	0.04	0.02
S→I/L	LC	S59	0.01		
S→N	HC	S30	0.03		0.02
	LC	S158			0.02
	LC	S170/173	0.07	0.09	0.05
	LC	S181	0.07	0.03	0.04
	LC	S207		0.10	
	HC	S270	0.01	0.02	0.10
	HC	S301	0.01		
	HC	S307	0.02	0.04	
	HC	S386			0.01
	HC	S411	0.08	0.06	0.02
	HC	S443	0.08		
	LC	S201	0.01		
V→I/L	LC	V57	0.01	0.02	0.03
	HC	V81	0.02		
	HC	V128		0.01	0.002
	LC	V190			0.01
	HC	V308/311	0.06	0.06	0.04
N→K	HC	N318		0.02	
S→R	HC	S443		0.07	
R→G	HC	R68			0.01
T→S	LC	T10			0.01
T→N	HC	T138			0.01

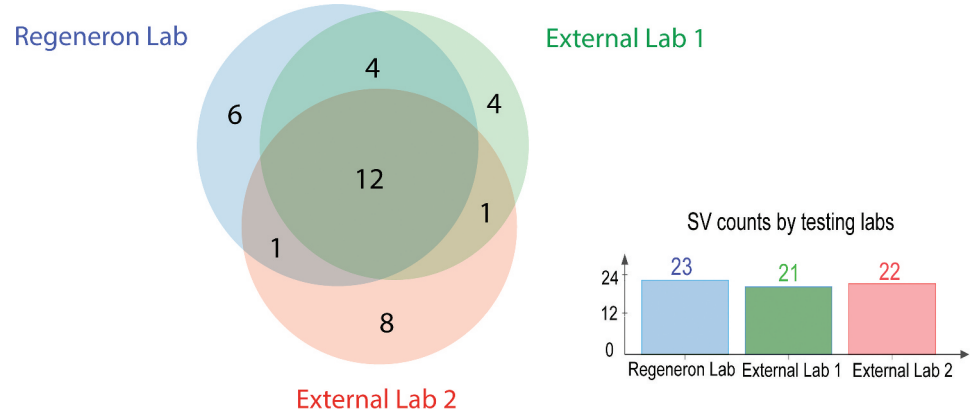

Figure 2. Comparison of the SV identification across three testing laboratories.

potentially affect the detection. The method-based variation is expected to decrease when analyzing elevated SVs, which is partially supported by the fact that none of the three laboratories reported SVs above 0.1% in the NIST mAb. In addition, based on our long-term experience, the run-to-run assay variation for our LC-MS method is relatively small in terms of both the SV identification and quantification (data not shown).

Following the method evaluation, the 15 approved therapeutic proteins listed in Table 2 were analyzed to determine their SV profiles. In the SV analysis, the identified sequence coverage is similar to that of a typical tryptic peptide mapping LC-MS/MS method used in product characterization and reaches greater than 97% across the 15 therapeutic proteins and NIST mAb (Supplemental Material). This high sequence coverage for individual protein ensures there is no large gap to preclude any SV identification. A total of 566 SVs were identified and manually verified across these 16 proteins, including the NIST mAb. SVs with the same type of amino acid substitution occurring to the adjacent or closely located positions were counted as two separate SVs if they could be well separated and unambiguously confirmed by MS2 fragmentation. Otherwise, they were counted as one SV. For SV benchmark analysis, the levels of the identified SVs were plotted against the amino acid substitution type for all 16 proteins, as shown in Figure 3a. In this plot, the 15 therapeutic proteins and NIST mAb listed in Table 1 are shuffled in order to mask the protein identities, and then designated as mAb-1 to mAb-16, although not all of these proteins are mAbs. The data in Figure 3a showed that the majority of SVs identified from the 16 proteins were present at very low levels, mostly less than 0.1%, consistent with the surveyed SV range by Zhang et al.[7] in Table 1. However, a small subset of SVs displayed elevated level above 0.1%, with the highest up to approximate 0.75% for some doublets (HC Y59/60F and HC Y94/95F in mAb-16, both in EU numbering). More interestingly, most of these elevated SVs resulted from the Tyr→Phe substitution. And furthermore, they are all from the same therapeutic protein, mAb-16. By contrast, two other identified Tyr→Phe SVs, HC Y295F in mAb-2 and HC Y206F in mAb-9 (both in EU numbering), were in the range below 0.1%.

Elevated Tyr→Phe SVs have been reported and extensively studied previously.[9,24] In these cases, depletion of Tyr in the cell culture media during bioreactor production was identified, which likely caused the subsequent mischarging of $tRNA^{Tyr}$ by Phe given their structural similarity upon Tyr depletion. The observation of elevated Tyr→Phe SVs at many different sites of the mAb-16 HC and LC sequences suggests that they may result from the same cause, i.e., amino acid depletion during the cell culture. For our SV benchmark analysis, we are mostly interested in the level of SVs that are present in recombinant therapeutic proteins produced from a fully optimized manufacturing process. Therefore, mAb-16 was excluded for the SV benchmark analysis below.

Figure 3b shows the levels of 522 SVs in the remaining 14 therapeutic proteins and NIST mAb included in the final benchmark analysis. They range from 0.001% to 0.154% (HC S440N in mAb-9, EU numbering). Of the 522 SVs, there is only a total of 9 SVs (~1.7% of total SVs) with their levels above 0.1% if we count the SV of LC L46/47V (0.122%) in mAb-2 as one. In addition, unlike the Tyr→Phe SVs observed in mAb-16, these nine SVs result from four different amino acid substitutions. Further, they are distributed across six different mAbs. These observations indicate that these nine SVs are unlikely to be caused by any systematic deficiency in the manufacturing process. Overall, the data in Figure 3b suggest that, in the approved therapeutic proteins with an optimized manufacturing process, the SVs are mostly well-controlled, with the majority of them below 0.1%, and with only a small percentage of outliers. This is further elucidated by Figure 4, where the level distribution of the identified SVs is plotted for each amino acid substitution type in a Boxplot. It is clear from Figure 4 that eight of the nine SVs (with the exception of HC S408N in mAb-9) are outliers to their corresponding distributions. In addition, across all the observed amino acid substitution types, the maximum level of their distributions is either very close to or fairly below the 0.1% limit. These results suggested that 0.1% may represent a reasonable upper limit of the SV benchmark for the approved therapeutic proteins.

SV control limit and process capability evaluation at Regeneron

Both the literature review and SV benchmark analysis of the approved therapeutic proteins suggested that 0.1% may present the upper limit of normal SVs to be expected in recombinant therapeutic proteins with a well-optimized manufacturing process. Before it was selected as a general SV control limit for new therapeutic protein development, we evaluated the process

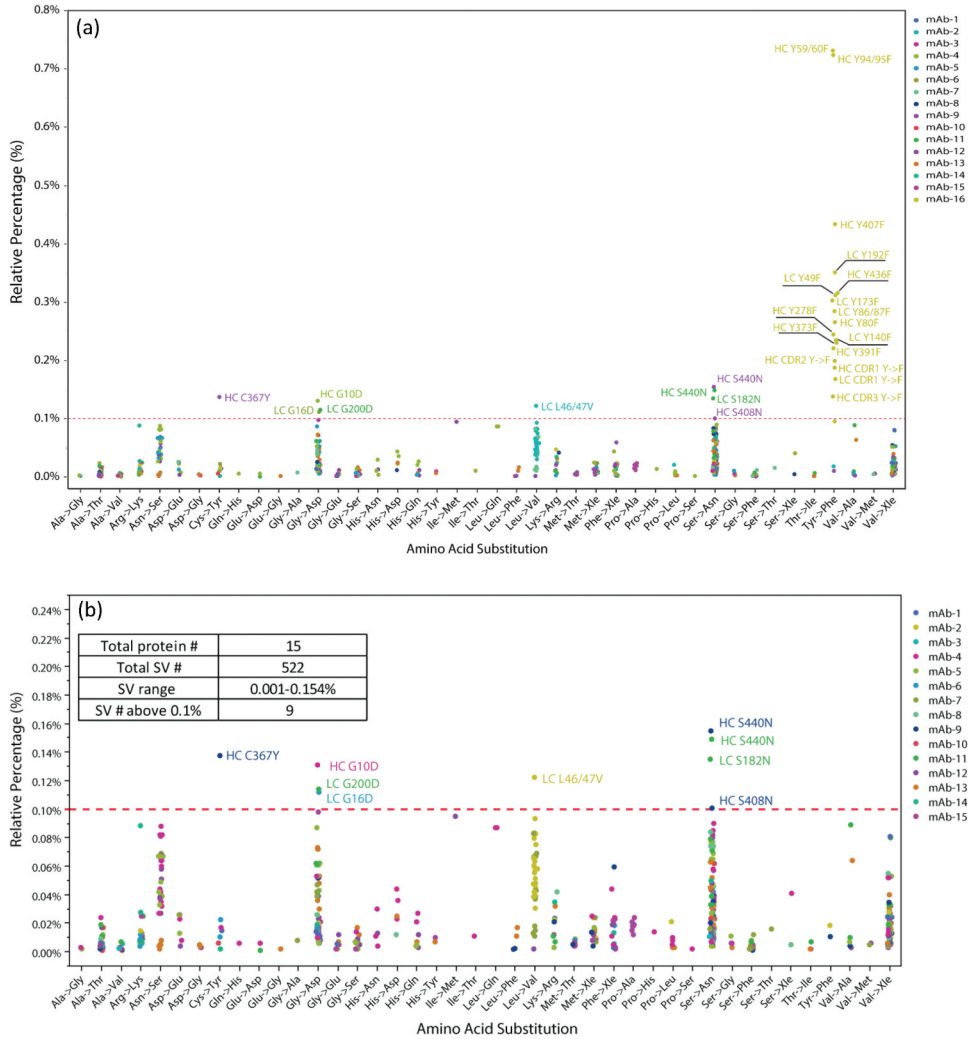

Figure 3. (a) The level of SVs and type of amino acid substitutions identified across 15 therapeutic proteins and NIST mAb. All the SVs above 0.1% are labeled with the substitution locations, all in EU numbering. For these occurring to the CDR regions, the SVs are labeled with the CDR number instead of the specific sequence location. (b) The level of SVs and type of amino acid substitutions identified across 14 therapeutic proteins and NIST mAb with the exclusion of mAb-16.

capability to determine how practical it is to meet such a stringent SV target in general. Five Regeneron mAbs (denoted as REGN-mAb1 to REGN-mAb5) were selected for this process capability evaluation. All five mAbs have reached the CMC development stage where the intended commercial manufacturing process was fully developed and characterized. Throughout their process development, from initial cell line selection, medium and feed strategy development, cell culture

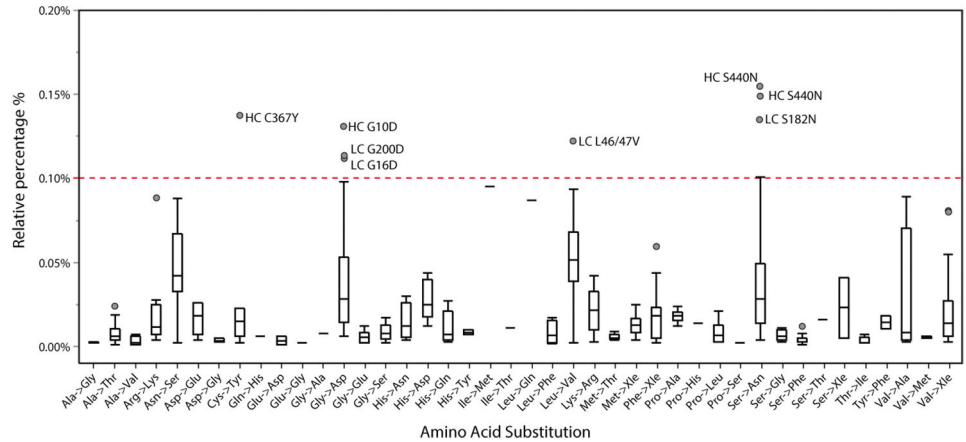

Figure 4. SV distribution boxplot across all the identified amino acid substitution type measured from 14 therapeutic proteins and NIST mAb. The SVs that are above the 0.1% limit can be considered as outliers of the distribution.

design-of-experiments optimization and scale-up confirmation at pilot scale, the SV profiles were closely monitored as one of the critical product quality attributes, along with important process performance parameters such as titer and viable cell density. If SVs above 0.1% were observed in any development stage, the underlying causes were investigated, and corresponding mitigation strategy was designed through process optimization. In addition, all the other critical quality attributes of the five mAbs were also met with respect to their predefined targets (data not shown).

The SV profiles of all five mAbs from the final confirmation pilot and GMP batches were measured (Figure 5). For each mAb, three lots of the confirmation batches were included in the analysis. In total, 132 distinct SVs were identified across the five mAbs. Consistent with the 14 surveyed therapeutic proteins in Figure 3b, the majority of the SVs observed in the five Regeneron mAbs are below the 0.1% limit. There are only four of the 132 SVs above the 0.1% limit (the two HC N389/390S SVs labeled in Figure 5 were counted once because they were the same SV in REGN-mAb1 measured in two confirmation batches). In addition, the highest SV level observed was 0.116%, only slightly above the 0.1% limit. These results aligned very well with the measured SV benchmark from the approved therapeutic proteins developed by the industry (Figure 3b). It is worth noting that elevated SVs significantly higher than 0.1% were observed for most of the five Regeneron mAbs during the early stages of process development. After proper process optimization, majority of them were below the 0.1% control limit in the final confirmation batches. These results, along with the approved therapeutic protein SV data, suggested that 0.1% is a practical SV control limit for biologic development. It can be achieved along with both other predefined critical quality attribute targets and the high titer goal when appropriate control strategy is implemented into the manufacturing process development.

Discussion

SVs have become a quality attribute that is increasingly discussed and studied during the development of biologics. Biopharmaceutical companies across the industry are developing their own control strategies to manage and mitigate the potential SV risk. Although no official guidance has been issued, regulatory agencies have started to show higher expectation with respect to the extent of data and evidence provided by sponsors to demonstrate sufficient control of SV level and associated risk in their manufacturing process development and final product characterization.[27] Currently, one major gap among the industry and regulatory agencies in this field is the lack of some consensus or general guidelines on how to appropriately assess the SV risks and subsequently set rational control limits during product and process development. This is largely due to the extreme heterogeneity of SV profiles that can potentially occur with all possible amino acid substitutions to the protein sequence, which makes it essentially impossible to fully assess the impact of different types of SVs on the safety and efficacy of the drug. As a result, compared to most other quality attributes like post-translational modifications, there has been limited prior knowledge accumulated so far to support SV risk assessment, particularly in terms of the clinical experience. With this gap, the first and probably most difficult challenge in the design and implementation of an effective SV control strategy is to set an appropriate SV control limit that is rational and can be widely adopted by the industry to guide the process and product development.

Here, we attempted to address this challenge by proposing a general SV control target for biologic development. The proposed SV numeric limit was justified by a holistic review of SVs from three different aspects: 1) typical levels of SVs observed under different scenarios of production systems; 2) SV benchmark analysis and clinical experience from approved therapeutic proteins; and 3) manufacturing process capability in terms of SV control. The literature review in Table 1 suggested that SVs can naturally occur in human proteins, which means they are not novel quality attribute associated with therapeutic proteins, thus alleviating the potential risk concern to some extent. However, it is important to note that the levels of these naturally occurring SVs are extremely low, and in the range consistent with expected spontaneous translational error rate of 10^{-5}-10^{-4} (equivalent to 0.01% SV or below).[7] By contrast, for recombinant therapeutic proteins, the SVs as surveyed by Zhang et al. showed an elevated level up to ~0.08% with the CHO expression system, representing an increase of ~10 fold from the biological noise level of SVs in natural proteins. This level of SV increase, which was also observed

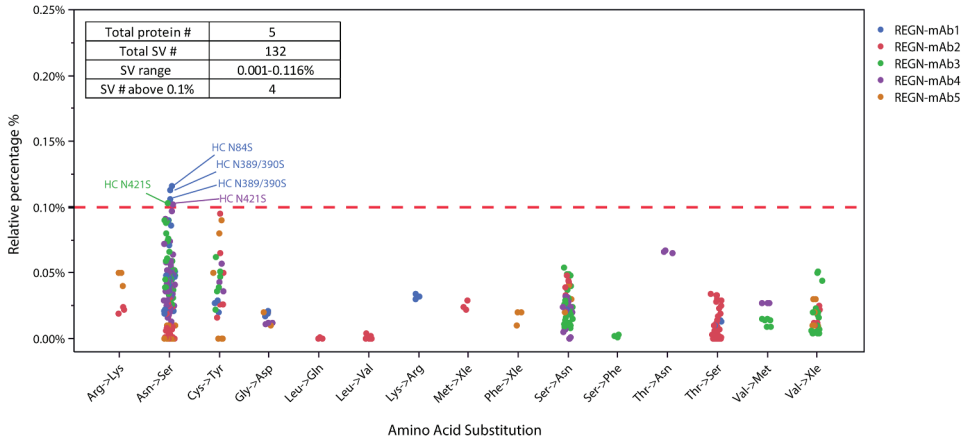

Figure 5. The level of SVs and type of amino acid substitutions identified across 5 REGN antibodies under developments. All the SVs above the 0.1% limit are labeled with the substitution locations, all in EU numbering.

in the benchmark analysis of the 14 approved therapeutic proteins and NIST mAb (Figure 3b), is likely a common phenomenon in recombinant therapeutic proteins. It may be related to the expression machinery stress that the cell lines experience in high-productive recombinant protein expression systems, thus leading to increased translational errors than in a normal biologic system. For cases where the production systems were not fully optimized in terms of SV control, such as the presence of certain DNA mutation in selected production cell line or depletion of critical nutrients during cell culture process, the levels of the resultant SVs could increase dramatically to span a wide range from 0.1% to ~50% as reported in literature (Table 1). This highlights the importance of implementing an appropriate strategy to monitor and control the SVs like all other common quality attributes throughout product and process development. More importantly, the observations of different SV range with different production systems and manufacturing processes suggest that it is critical to select an appropriate SV control limit in therapeutic protein development. Such an SV control limit should consider the normal level of SV increase observed in recombinant therapeutic proteins relative to the natural biologic proteins. It should also be able to distinguish the abnormal level of SVs resulting from non-fully optimized manufacturing process, such as those from DNA mutation or amino acid depletion.

The SV benchmark analysis of approved therapeutic proteins further supports this notion. As shown in Figure 3b, the majority of the SVs identified across the 14 approved therapeutic proteins and NIST mAb are in the range of 0.01–0.1%, which also represents an ~10-fold increase relative to the biological noise level of SVs in natural biologic proteins. Very few SVs (9 of 522 SVs, or 1.7%) showed levels above 0.1%. They could be considered as the outliers based on the distribution analysis in Figure 4. However, even for these SV outliers, the highest level observed was 0.154%, which was only slightly higher than the 0.1% limit. These data suggested that 0.1% may be a good representation of the SV upper limit (or industry standard) in recombinant therapeutic proteins when the employed manufacturing process is well optimized. Besides the evidence from the approved therapeutic proteins, the 0.1% SV control limit was also supported by the process capability evaluation of the Regeneron mAbs under development. As shown in Figure 5, when an appropriate control strategy was incorporated into the process development, the level of SVs could be successfully managed to meet the 0.1% control limit for all of the five mAb development programs. When the manufacturing process was not fully optimized, thereby leading to abnormal level of SVs in the final therapeutic protein products, the 0.1% control limit was also shown to be sensitive enough to distinguish them from the SV benchmark, as demonstrated by the case of mAb-16 in Figure 3a. This is particularly important because if the selected SV control limit is too high, the designed SV control strategy may not be able to detect abnormal SVs from the benchmark, and thus fail to direct the necessary manufacturing process optimization.

The SV survey of approved therapeutic proteins also provides valuable clinical experience to support SV risk assessment and control limit setup for therapeutic protein development. As suggested by the survey, SVs are usually present in very low levels (0.01–0.1%) in the final protein products. Therefore, their potential impact on the product's biologic activity, efficacy, and pharmacokinetics/pharmacodynamics is expected to be low, if any. The main risk concern with SVs is that immunogenicity may be potentially enhanced by generation of some neoepitopes in the therapeutic protein generated from the unintended amino acid substitutions. Evidence has been reported in literature that a single amino acid substitution in a protein sequence can lead to significantly enhanced immunogenicity in both a vaccine[28] and a therapeutic mAb.[29] However, the actual immunogenicity risk associated with SVs is unknown. Of the 14 approved therapeutic proteins (i.e., with mAb-16 and NIST mAb excluded) included in the SV survey discussed above, all demonstrated efficacy and safety, including acceptable immunogenicity, in both clinical trials and post-approval clinical use. Our SV survey provided accumulated clinical evidence in a variety of disease settings that the presence of SVs within a limit of 0.1% in final drug products can be generally considered safe with low immunogenicity risk.

In summary, the results from this study suggest that 0.1% at an individual amino acid site is an appropriate SV control limit for the biologic product and process development. It reflects an approximate 10-fold margin for the normal SV elevation in recombinant therapeutic proteins compared to the biologic noise in natural proteins. It is also well aligned with the SV benchmark surveyed from the 14 approved therapeutic proteins and NIST mAb (Figures 3b and 4). Therefore, the associated risk with SVs up to the 0.1% limit can be justified by the accumulated clinical experience from these therapeutic proteins. Furthermore, the 0.1% control limit was also shown to be sensitive enough to detect abnormal SV profiles when the manufacturing process was not well optimized, critical for the design of a control strategy to guide process development. Finally, the assessment of internal process capability further suggested that 0.1% is a practical SV control limit for biologics development when appropriate control strategy is implemented into process development. However, it is worth clarifying that the 0.1% control limit we are proposing here is a general development target (or action limit) during the manufacturing process development. It is not necessarily meant to be a stringent acceptance criterion for release or clinical use of therapeutic proteins under development. As shown in both approved therapeutic proteins and Regeneron mAbs, a small percentage of SV outliers above the 0.1% limit may occasionally exist. When such cases rise, extensive risk assessment will probably need to be conducted to consider the nature of the associated amino acid substitution, their locations, the dosage level, disease indication, the immune environment in targeted patients and other factors to determine the specific risk and acceptable criteria. Finally, it is also worth noting that the SV benchmark analysis and the proposal of a 0.1% action limit in this study are fully based on the amino acid substitution analysis at the protein level using the LC-MS/MS method. To implement a more comprehensive SV control strategy in biologic development, particularly during the cell line development, genetic mutations that can also lead to SV formation should also be considered and monitored at proper stages using

technologies such as next-generation sequencing in combination with protein level SV analysis using LC-MS/MS as described before.[13]

Materials and methods

Materials

The NIST reference mAb (Catalog #: RM8671) was purchased from Millipore-Sigma. 15 approved therapeutic proteins, including 14 mAbs and one Fc fusion protein are commercially available. The five Regeneron mAbs analyzed were manufactured and purified internally. Sequencing grade-modified trypsin (Catalog #: V5111) was purchased from Promega. PNGase F (Catalog #: P0704L) used for N-linked glycan removal was purchased from New England BioLabs. All other chemical reagents and lab supplies used in the protein digestion were from Sigma unless otherwise specified.

Sample preparation

A typical reduced and alkylated trypsin digestion procedure was used for the sample preparation in the peptide mapping-based SV analysis. Briefly, 1.0 mg of each protein sample (including both commercial therapeutic proteins and Regeneron mAbs) was aliquoted and buffer exchanged into 5 mM acetic acid solution using Nanosep® 10 K centrifugal filter (Pall Corporation). After buffer exchange, a 100 µg aliquot of each antibody sample was denatured and reduced in 5 mM TCEP-HCl (Catalog #: 20491, Thermo Scientific) by heating at 80°C for 10 minutes. After denaturation and reduction, the samples were diluted with 1 M UltraPure pH 7.5 Tris-HCl solution (Catalog #: 15567–027, Invitrogen) containing 8 M urea. The diluted samples were then alkylated and digested simultaneously with 2 mM iodoacetamide and trypsin at an enzyme-to-substrate ratio of 1:20 (w/w) at 37°C in the dark for 3 hours. Following trypsin digestion, each sample was further deglycosylated with 10 µL of 1 mU/µL PNGase F at 37°C for another 1 hour, then quenched with 5% trifluoroacetic acid solution.

LC-MS/MS method and data acquisition

The tryptic digests of each protein sample were analyzed using a Waters ACQUITY UPLC system coupled to a Thermo Scientific Q Exactive Plus mass spectrometer for the SV analysis. Waters CSH C18 column (1.7 µm particle size, 2.1 mm × 150 mm) was used for the peptide separation. Two mobile phases: 0.1% formic acid in water (mobile phase A) and 0.1% formic acid in acetonitrile (mobile phase B), were used with an acetonitrile gradient of 0.1–40% over 85 minutes to separate and eluate the tryptic peptides at the flow rate of 0.25 mL/min. The column was then washed with 90% acetonitrile for 10 min and re-equilibrated in mobile phase A. Between two sample injections, one water blank run was performed to eliminate potential carryovers from previous sample, which could potentially introduce false positive in SV identification. The MS data were acquired on Thermo Scientific Q Exactive Plus mass spectrometer with one full scan at resolution of 70 k (at m/z 400), followed by 5 data-dependent MS/MS scans at a resolution of 17.5 k. The automatic gain control (AGC) for the MS/MS scan was set to $1e^5$ with maximum IT of 300 ms to improve fragmentation spectra for low abundance SVs.

SV data analysis

The acquired LC-MS/MS data were first analyzed using Byonic™ software (Protein Metrics) to identify all potential SVs. Common post-translational modifications along with all the possible amino acid substitutions resulting from single nucleotide change were included in the search. The identified candidates of SVs by Byonic™ software 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 native form using Byologic™ (Protein Metrics). After removal of all the possible false positive results, the relative abundance of each verified SV was determined using Skyline software with selection of all available charge states and three isotopic peaks for both of the SV-containing peptide and its native peptide.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

ORCID

Haibo Qiu http://orcid.org/0000-0002-8449-4233

References

1. Ecker DM, Jones SD, Levine HL. The therapeutic monoclonal antibody market. MAbs. 2015;7:9–14. doi:10.4161/19420862.2015.989042.
2. Weiner GJ. Building better monoclonal antibody-based therapeutics. Nat Rev Cancer. 2015;15:361–70. doi:10.1038/nrc3930.
3. The Antibody Society. Antibody therapeutics approved or in regulatory review in the EU or US. [accessed 2020 June 7]. https://www.antibodysociety.org/resources/approved-antibodies/
4. Valliere-Douglass J, Marzilli L, Deora A, Du Z, He L, Kumar S, Liu YH, Martin-Mueller H, Nwosu C, Stults J, et al. Biopharmaceutical industry practices for sequence variant analyses of recombinant protein therapeutics. PDA J Pharm Sci Technol. 2019. doi:10.5731/pdajpst.2019.010009.
5. Borisov OV, Alvarez M, Carroll JA, Brown PW. Sequence variants and sequence variant analysis in biotherapeutic proteins. State-of-the-art and emerging technologies for therapeutic monoclonal antibody characterization Volume 2. Biopharmaceutical characterization: the NISTmAb case study. ACS Publications; 2015. p. 63–117.
6. Lin TJ, Beal KM, Brown PW, DeGruttola HS, Ly M, Wang W, Chu CH, Dufield RL, Casperson GF, Carroll JA, et al. Evolution of a comprehensive, orthogonal approach to sequence variant analysis for biotherapeutics. MAbs. 2019;11:1–12. doi:10.1080/19420862.2018.1531965.
7. 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. doi:10.1021/bi401002c.
8. Wong HE, Huang CJ, Zhang Z. Amino acid misincorporation in recombinant proteins. Biotechnol Adv. 2018;36:168–81. doi:10.1016/j.biotechadv.2017.10.006.

9. Zeck A, Regula JT, Larraillet V, Mautz B, Popp O, Gopfert U, Wiegeshoff F, Vollertsen UEE, Gorr IH, Koll H, et al. Low level sequence variant analysis of recombinant proteins: an optimized approach. PLoS One. 2012;7:e40328. doi:10.1371/journal.pone.0040328.
10. Khetan A, Huang YM, Dolnikova J, Pederson NE, Wen D, Yusuf-Makagiansar H, Chen P, Ryll T. Control of misincorporation of serine for asparagine during antibody production using CHO cells. Biotechnol Bioeng. 2010;107:116–23. doi:10.1002/bit.22771.
11. Harris RP, Mattocks J, Green PS, Moffatt F, Kilby PM. Determination and control of low-level amino acid misincorporation in human thioredoxin protein produced in a recombinant Escherichia coli production system. Biotechnol Bioeng. 2012;109:1987–95. doi:10.1002/bit.24462.
12. Harris RJ, Murnane AA, Utter SL, Wagner KL, Cox ET, Polastri GD, 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. Biotechnology (N Y). 1993;11:1293–97. doi:10.1038/nbt1193-1293.
13. Zhang S, Bartkowiak L, Nabiswa B, Mishra P, Fann J, Ouellette D, Correia I, Regier D, Liu J. Identifying low-level sequence variants via next generation sequencing to aid stable CHO cell line screening. Biotechnol Prog. 2015;31:1077–85. doi:10.1002/btpr.2119.
14. Kunkel TA. DNA replication fidelity. J Biol Chem. 2004;279:16895–98. doi:10.1074/jbc.R400006200.
15. Imashimizu M, Oshima T, Lubkowska L, Kashlev M. Direct assessment of transcription fidelity by high-resolution RNA sequencing. Nucleic Acids Res. 2013;41:9090–104. doi:10.1093/nar/gkt698.
16. Gout J-F, Thomas WK, Smith Z, Okamoto K, Lynch M. Large-scale detection of in vivo transcription errors. Proc Natl Acad Sci US A. 2013;110:18584–89. doi:10.1073/pnas.1309843110.
17. Ellis N, Gallant J. An estimate of the global error frequency in translation. Mol Gen Genet. 1982;188:169–72. doi:10.1007/BF00332670.
18. Edelmann P, Gallant J. Mistranslation in E. coli. Cell. 1977;10:131–37. doi:10.1016/0092-8674(77)90147-7.
19. Fu J, Bongers J, Tao L, Huang D, Ludwig R, Huang Y, Qian Y, Basch J, Goldstein J, Krishnan R, et al. Characterization and identification of alanine to serine sequence variants in an IgG4 monoclonal antibody produced in mammalian cell lines. J Chromatogr B Analyt Technol Biomed Life Sci. 2012;908:1–8. doi:10.1016/j.jchromb.2012.09.023.
20. DeGruttola HS, Marzilli L, Beal KM, Anderson K, Rouse JC. Sequence variant analysis of antibody biotherapeutics with ultrahigh-resolution mass spectrometry. Biotherapeutics analytical summit. Baltimore (MD); 2015.
21. Huang Y, O'Mara B, Conover M, Ludwig R, Fu J, Tao L, Li ZJ, Rieble 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. doi:10.1002/pro.2046.
22. Hutterer KM, Zhang Z, Michaels ML, Belouski E, Hong RW, Shah B, Berge M, Barkhordarian H, Le E, Smith S, et al. Targeted codon optimization improves translational fidelity for an Fc fusion protein. Biotechnol Bioeng. 2012;109:2770–77. doi:10.1002/bit.24555.
23. Wen D, Vecchi MM, Gu S, Su L, Dolnikova J, Huang Y-M, Foley SF, Garber E, Pederson N, Meier W, et al. Discovery and investigation of misincorporation of serine at asparagine positions in recombinant proteins expressed in Chinese hamster ovary cells. J Biol Chem. 2009;284(47):32686–94. doi:10.1074/jbc.M109.059360.
24. Feeney L, Carvalhal V, Yu XC, Chan B, Michels DA, Wang YJ, Shen A, Ressl J, Dusel B, Laird MW, et al. Eliminating tyrosine sequence variants in CHO cell lines producing recombinant monoclonal antibodies. Biotechnol Bioeng. 2013;110(4):1087–97. doi:10.1002/bit.24759.
25. Raina M, Moghal A, Kano A, Jerums M, Schnier PD, Luo S, Deshpande R, Bondarenko PV, Lin H, Ibba M. Reduced amino acid specificity of mammalian tyrosyl-tRNA synthetase is associated with elevated mistranslation of Tyr codons. J Biol Chem. 2014;289:17780–90. doi:10.1074/jbc.M114.564609.
26. Ni J, Gao M, James A, Yao J, Yuan T, Carpick B, D'Amore T, Farrell P. Investigation into the misincorporation of norleucine into a recombinant protein vaccine candidate. J Ind Microbiol Biotechnol. 2015;42:971–75. doi:10.1007/s10295-015-1613-x.
27. Rogstad S, Faustino A, Ruth A, Keire D, Boyne M, Park J. A retrospective evaluation of the use of mass spectrometry in FDA biologics license applications. J Am Soc Mass Spectrom. 2017;28:786–94. doi:10.1007/s13361-016-1531-9.
28. Kumar SR, Prabakaran M, Ashok Raj KV, He F, Kwang J. Amino acid substitutions improve the immunogenicity of H7N7HA protein and protect mice against lethal H7N7 viral challenge. PLoS One. 2015;10:e0128940. doi:10.1371/journal.pone.0128940.
29. Joubert MK, Deshpande M, Yang J, Reynolds H, Bryson C, Fogg M, Baker MP, Herskovitz J, Goletz TJ, Zhou L,et al. Use of in vitro assays to assess immunogenicity risk of antibody-based biotherapeutics. PLoS One. 2016;11:e0159328. doi:10.1371/journal.pone.0159328.