

## **Table 5: Best Practices for Sequence Variant Analysis**

**Facilitator:** Yuko Ogata, *Seagen Inc.*

**Scribe:** Elsa Gorre, *Janssen Research & Development, LLC*

**Key Words:** Sequence variants, amino acid misincorporations, mutation

### **Table Scope:**

Sequence variants, or unintended amino acid substitutions in protein therapeutics, can arise from 2 main sources: 1) Amino acid misincorporation due to culture conditions, and 2) mutation, mistranslation, or truncation at DNA/RNA. Identification and quantitation of these variants are a critical part of clone selection, process optimization and product release as they can affect target binding, stability, antibody clearance, and immunogenicity at high levels. In addition, recognition of the source of variants (process vs. DNA level) is critical in identifying the best strategy for control. A careful evaluation of DNA mutation during clone selection is particularly important because such mutations are not generally mitigatable. Next-Generation Sequencing provides orthogonal and complimentary detection of DNA mutations, but it is not practical to use this technique for all top clones. Multi-Attribute Method (MAM) or reduced peptide mapping is a commonly used technique for sequence variant analysis (SVA) in biopharma industry, but identification and localization can be challenging. This roundtable discussion will focus on the opportunities and best practices for phase appropriate SVA and reporting.

### **Discussion Questions:**

- What are the most effective ways to minimize false positives - multi-enzyme digestion, computational approaches, or other innovative strategies?
- What should be the minimal levels of sequence variants/misincorporations reported to the process team and regulatory agencies?
- What are the limitations in the current software programs for SVA? How can they be improved?
- What can be done to reduce the time required for SV identification, monitoring, and validation?
- What is the best strategy for differentiating misincorporation vs. mutation at the DNA level? Should we take different approaches for identification and monitoring?

### **Discussion Notes:**

- Sequence variance analysis is implemented differently across multiple biopharmaceuticals. In some pharmaceutical companies, SVA is a new process that they are implementing primarily in the early stages, while most companies perform SVA to guide their Cell line development group in their clone selection process.
  - o This assay is done to either top 24,8 or 6 clones.
  - o From initial 48 clones, other product quality attributes are used to cut down the number of clones.
- There was discussion regarding what level of SV is reported back to the team. Different groups would report anything greater than or equal to 0.1%, while some were reporting 0.5% or greater.

- In addition to SVA, complementary assays such as amino acid analysis (AAA) or spent media analysis are performed to verify amino acid depletion levels that may lead to misincorporation.
- Molecules with well-known misincorporations are used as controls (positive controls) in some cases.
  - o Used to set specification in limit of detection.

#### Multi-enzymatic approach for SVA

- Usually, more than 1 enzyme is used in sequence variance analysis.
- The most preferred enzyme is trypsin, and a secondary enzyme can be chymotrypsin or GluC.
- Secondary enzymes are used to verify true positive hits detected from trypsin.
- Some CLD groups use Next Generation Sequencing (NGS) to verify the variant detected by mass spec.
- If a sequence variant is found across multiple sites, that SV is thought to be a misincorporation due to an amino acid depletion.
- There are cases where SVA can be more challenging due to protein heterogeneity, such as in fusion proteins with multiple glycosylation sites.
  - o In those cases, the recommendation is to deglycosylate the protein before digestion.

#### Software used to Analyze Sequence Variance

- There are many different software programs available to use for sequence variant analysis.
  - o Some groups have built in-house programs to help them with either visualizing sequence variants or to quickly identify false positives.
- There is a need for software to help with identifying false positives as that is unanimously the most time-consuming step in this analysis.
- Some software programs have an auto-flagging script to highlight potential false positives; to get the script, people may need to reach out to a technical support.
  - o Visualization of RT shifts from WT peptides along with predicted RT shifts for the modifications can be helpful.
  - o The stacked plots of Wt and Var is useful during validation.
- A participant from a software company mentioned efforts being made for further improvement.
  - o They are building flags/ logics steps that can be employed to minimize the false positives observed.
  - o During their pre-processing, an automatic background subtraction is performed.
- New peak detection feature can be helpful in identifying variants in some cases.