

### Table 13: Software Analysis of MS Data - Emerging Approaches and Best Practices

**Facilitator:** Simon Letarte, *Merck & Co., Inc.*

**Scribe:** Jason Gilmore, *Seattle Genetics, Inc.*

#### Scope:

Software is a critical component of protein characterization by mass spectrometry. The large volume of data generated by modern mass spectrometers gets translated into product knowledge and ultimately distilled to a few numbers and figures in a regulatory submission. Available software exists on a continuum, from the basic vendor-supplied ones for viewing spectra and integrating peaks, to proteomics software and to sophisticated, database-backed solutions designed for biopharmaceutical applications. Most laboratories, depending on the diversity of their hardware, will end up using a variety of MS software for their protein characterization needs. This represents challenges in terms of training, access, maintenance, compliance, acquisition cost and method documentation. At the same time, with the recent advances in AI and big data, exciting opportunities exist to make our workflows smarter and more efficient.

#### Topics for Discussion:

1. How many different MS software are you using in your laboratory?
2. Do you see your organization increasing or decreasing the number of different MS software next year?
  1. year?
  2. Are you standardized on a vendor platform or are you using multiple ones?
  3. Do you use mostly vendor-specific or vendor-agnostic software?
  4. Is pricing and subscription-based software an issue within your organization?
  5. Is finding CFR 11 part 21 compliant software a challenge?
  6. What compliant software do you use today?
  7. Is advanced software reducing approval time or enabling more in-depth characterization?
  8. Do you export the data into another software for statistics & figure creation? At what point?
  9. Do you see a role for AI (artificial intelligence, machine learning) and big data approaches in protein characterization?
  10. protein characterization?
  11. What would you like to see that does not exist yet?

#### Discussion Notes:

##### How many different MS software are you using in your laboratory?

Use of multiple tools is the norm:

- Xcalibur
- BioPharma Finder
- Proteome Discoverer
- ProMass
- MassLynx
- ProteinMetrics
- Genedata
- In-house solutions
- Python, Origin, R, JMP, etc.

Even when there is a primary vendor/tool, other software may be used for specialized tasks.

Do you see your organization increasing or decreasing the number of different MS software next year?

Some institutions have a preferred software suite and stay within it across groups. Others allow individual users to determine the tools they will use. There was no consensus on a desire to increase or decrease the number of software tools so long as assay-specific needs were met.

Do you use mostly vendor-specific or vendor-agnostic software?

No clear preference. Breadth of software available at each company depends on budget and willingness of each group to spend time evaluating new tools as they become available.

Is pricing and subscription-based software an issue within your organization?

Barriers to adopting new tools included cost of currently deployed tool suites, loss of time learning new systems, and the time it takes to establish confidence in new tools.

Several participants cited license structures/deployment as a pain-point of working with independent software suites. Some participants said that avoiding these license hassles was more important than having the “very best” software. Vendors say that constant support/updates requires a subscription service.

Is finding CFR 11 part 21 compliant software a challenge?

There is an abundance of well vetted software meeting these standards. The larger issue was establishing confidence that a software tool could be deployed and used to efficiently analyze data. Users desire more “trust” to be built with software vendors to avoid “black box” algorithms and software.

Is advanced software reducing approval time or enabling more in-depth characterization?

Emphatically both. Participants believe that more can be done in less time with existing tools.

Do you export the data into another software for statistics & figure creation? At what point?

Several participants export “final” data from existing tools and perform statistics and visualization using in-house solutions or other software. These users cite insufficient flexibility of built-in visualizations.

There is a tension between flexibility and ease of use for software tools. If the software is intuitive and simple to use, advanced users find it insufficiently flexible for complex analyses. If the software is highly flexible, there are more opportunities for errors and misinterpretation or users may have difficulty using the software. Vendors would prefer to identify “power users” and establish them as on-site contacts so that development can be focused.

Do you see a role for AI (artificial intelligence and machine learning) and big data approaches in protein characterization?

Consensus was that these approaches have potential but that more time is needed to see them develop and be validated. It is not clear how these would be implemented/deployed. As with existing tools, users require as much transparency as possible to understand how the tools are working.

What would you like to see that does not exist?

Unmet needs:

- Better retention time prediction
- Tools for Data-independent acquisition (DIA)