Table 10: Okay, Now That I Have All This HPLC/MS/MS Data, What Do I Do with It?

Facilitator: Xiaojuan Li, *Merck & Co., Inc.*

Scribe: Claire Bramwell, *Protein Metrics, Inc.*

Scope:
The high mass accuracy and high resolution makes tandem liquid chromatography-mass spectrometry (LC-MS/MS) an essential tool for identification, structure elucidation, relative and absolute quantitation analysis. It has wide applications in many fields for variety of analytes: small molecule, large therapeutics, proteomics, metabolomics, and mixture analysis, etc. The LC-MS/MS data normally contains many MS1 full sans and corresponding MS/MS scans, which makes the data interpretation very complex. This roundtable discussion with focus on understanding LC-MS/MS data, data quality evaluation, data analysis software, manual data analysis, and application of LC-MS/MS. From this discussion, better understanding of LC-MS/MS data will be achieved; data analysis software and manual analysis experience will be exchanged; and the application of LC-MS/MS technology will be discussed.

Questions for Discussion:

1. How to know if the experiment went well or not? How to evaluate LC-MS/MS data quality for different types of analysis?
    - internal standard;
    - system suitability;
    - check TIC, BPC;
    - collision energy, fragmentation efficiency, and MS^n;
    - different analyte: small molecule, large therapeutic, proteomics, metabolism, or mixture, makes difference);
    - types of analysis: bottom up, middle down, top down analysis.

2. How to analyze the LC-MS/MS data: software or manual interpretation?
    - raw data pre-process: noise removal, peak detection, etc.
    - list of software
    - manual analysis tools

3. What kinds of information can be achieved from LC-MS/MS analysis and how to achieve?
    - sequence coverage
    - identification
    - structure elucidation
    - quantitation

Discussion Notes:

- Applications used: All participants worked in characterization (biotherapeutics, gene therapy and capsid analysis). The following types of experiments are used:
    - Peptide Maps
    - PTM's
    - HCP's
    - Disulfide bond analysis

- Intact

- System Suitability-
  - Different methods of determining system suitability included:
  - Running peptide standards – different types of standards are used for different workflows.
  - Checking MS2 efficiency.
  - Frequency of standards for system suitability ranged from once a week, a few times a week, to running bracketing standards for each run.
  - Some participants also used system suitability software "QC" checks, such as Skyline.
  - Some also didn't do any formal system suitability experiments but inspected the data manually to determine if there is an issue with the instrument. It was felt that running system suitability adds unnecessary processing and run time.

- Data Processing- most used software (both mass spec vendor and third-party software). It was suggested by a few that there is no substitute for manually processing data.

- Software used by members at the round table:
  - BioPharma Finder (Qualitative)
  - Chromelean (Quantitative)
  - Excalibur (Qualitative)
  - Gene Data (Quantitative and Qualitative)
  - Pin Point (Quantitative)
  - Protein Metrics (Quantitative and Qualitative)
  - Skyline (Quantitative)

  It was felt by members of the round table that it is difficult to find software for effective processing of disulfide bond data. It was found that most of the software they had evaluated for disulfide bond analysis returned a number of false positives, and the analyst would have to integrate manually. Also looking for scrambled disulfides gave a lot of false positives. A lot of this work is processed manually using reduced and non-reduced peptide maps.

  The location of data used for processing varies. Some are able to process data from the server where it is stored, and others need to process data on a local PC.

- Data Integrity

  Methods of storing and archiving data is as follows:
  - Enotebook
  - Any data used for filing goes into SAP
  - External vendour cloud
  - Server, either backed up manually or automatically by IT

- Vendour enterprise system

It was debated how long data should be archived for. There were suggestions from a few years to indefinitely.

- Assessment of Method:

  How to determine efficient digestion was discussed. This is typically determined by measuring missed cleavages. This was done both manually and using software. Trypsin is the most popular enzyme used, followed by Lys-C.

  It was also debated if CE and fragmentation efficiency needs to be optimized. For this assessment users integrate the data manually, especially to inspect DDA data.

- Data Review:

  For the raw sample files most had reviewers randomly choose a few data files to inspect. For PTM calculations in excel, just the % is reviewed rather than raw data.
  For quantitation experiments there is an example where multiple analysts are used.
  Some had no data review.

- Time spent processing data:

  For the software users this appeared to be approximately 20% of a FTE's time. If a project was new then this may be as much as 80%, although this decreases to 20% as the project progresses. For manual interpretation, data processing time is approximately 50% FTE time.