ASMS 2026 Mass Spectral Libraries Interest Group Workshop Report

June 2, 2026 | San Diego Convention Center

Interest Group Chair: Tytus D. Mak (NIST)

Attendance: ~80 people

The Past, Present, and Future of Mass Spectral Libraries

The 2026 ASMS Mass Spectral Libraries Interest Group workshop brought together eight prominent researchers and developers representing major mass spectral library resources. The workshop explored the historical development, present capabilities, and future directions of mass spectral libraries, emphasizing their critical role in compound and peptide identification. Each panelist provided a brief presentation introducing their respective library or computational resource, followed by an open discussion addressing current challenges and opportunities facing the mass spectrometry community.

The invited speakers and their respective resources were:

- **Lewis Geer (NIST):** NIST peptide mass spectral libraries and associated informatics resources.
- **Stephen E. Stein (NIST):** NIST/EPA/NIH Mass Spectral Library and search technologies.
- **Henry Lam (HKUST):** SpectraST peptide spectral library searching.
- **Graeme Whitley (Wiley):** Wiley Registry of Mass Spectral Data.
- **Mingxun Wang (GNPS):** Global Natural Products Social Molecular Networking (GNPS).
- **Oliver Fiehn (MoNA):** MassBank of North America (MoNA).
- **Gary Siuzdak (METLIN):** METLIN metabolite spectral database.
- **Mathias Wilhelm (Prosit):** Prosit deep-learning-based peptide spectrum prediction.

Together, the presentations illustrated the evolution of mass spectral libraries from carefully curated experimental reference collections to large-scale, community-driven repositories and computationally generated spectral resources. The subsequent panel discussion explored the broader implications of these developments, particularly the need to balance expanding spectral coverage with identification reliability.

An early discussion centered on the fundamental limitations of compound identification. Panelists emphasized that reference spectra represent only the starting point for identification and that a high spectral similarity score does not necessarily establish the identity of an unknown compound. The persistent difficulty of identifying metabolites was illustrated by an example from a dietary biomarker study in which only one of 123 statistically significant and validated compounds could be identified. This underscored the considerable gap between detecting analytically meaningful features and assigning confident chemical identities.

A particularly substantive audience question concerned whether increasing spectral library size could inadvertently increase false-positive identifications as increasingly rare compounds are included. The

discussion highlighted differences between proteomics and small-molecule analysis, particularly the applicability of target-decoy approaches and false discovery rate (FDR) estimation. One panelist challenged the assumption that FDR is adequately controlled in proteomics, noting that target-decoy methods primarily address random incorrect matches rather than systematic misidentifications. Other panelists questioned whether conventional FDR concepts can be meaningfully applied across the structural diversity of small molecules. The exchange emphasized the importance of considering sample context, compound prevalence, and additional analytical evidence rather than relying exclusively on spectral matching scores.

Another discussion focused on the rapidly expanding role of artificial intelligence in spectral prediction. An audience member questioned whether models trained on existing experimental libraries could accurately predict fragmentation behavior across unfamiliar chemical space, particularly given the complexity of EI fragmentation and secondary ion reactions. Panelists expressed differing perspectives on whether AI models fundamentally interpolate between existing experimental observations or possess the potential to extrapolate beyond their training data. Nevertheless, the discussion emphasized the importance of understanding the chemical coverage of training datasets, characterizing prediction uncertainty, and maintaining experimentally measured reference spectra as essential ground truth for model development and validation.

The role of traditional mass spectrometry education was also examined. An audience member asked whether teaching students to manually interpret mass spectra and understand library search scores would remain relevant as AI-based identification tools become increasingly capable. Panelists strongly supported continued instruction in fragmentation chemistry, analytical fundamentals, and critical interpretation. They emphasized that the ability to independently evaluate an identification, recognize implausible results, and incorporate prior knowledge of a sample remains essential, regardless of advances in automated analysis.

Data interoperability and standardization emerged as additional practical concerns. An audience member described the substantial effort required to convert data between incompatible formats, prompting discussion of efforts to establish common spectral data models. Panelists noted that reaching consensus on metadata standards has been challenging and that widespread adoption remains a greater obstacle than defining the formats themselves. The use of controlled vocabularies was identified as an important step toward improving metadata consistency, although differences in the completeness and availability of metadata across libraries remain unresolved. AI-assisted software development was also suggested as a potentially useful means of facilitating data conversion and implementing common standards.

Finally, the discussion addressed the integration of orthogonal analytical measurements, including chromatographic retention behavior, ion mobility, collision cross sections (CCS), and multistage fragmentation. Panelists emphasized the value of acquiring additional structural evidence while acknowledging practical limitations. In response to an audience question about building in-house libraries, stepped collision energies were recommended as a useful strategy for improving fragmentation coverage. Although CCS and ion mobility measurements offer additional identification capabilities, the ability of current technologies to distinguish structurally similar isomers was cautioned against being overstated.

Overall, the workshop highlighted the continued importance of high-quality experimental reference data alongside the growing capabilities of community-driven libraries and computational prediction. The

discussion reinforced that advances in spectral coverage and artificial intelligence must be accompanied by improvements in identification confidence, data standardization, and analytical interpretation. Continued collaboration among library developers and users will be essential for ensuring that next-generation spectral libraries remain reliable, interoperable, and scientifically meaningful.