**Introduction**

While many mass-spectrometry (MS) software tools are freely available, the large majority of them are designed to solve small, well-defined problems and require complex manual steps to translate data into meaningful biological results. Users have to learn how to install and use each tool and need to be aware of all the possible pitfalls of connecting disparate algorithms that were not designed to work together, often resulting in exceedingly high False Discovery Rates or missing many positive identifications. ProteoSAFe is a software platform designed to bridge this gap by allowing bioinformaticians to integrate their novel algorithms into workflows (Flexibility), which are then seamlessly executable on either desktops or compute clusters (Scalability), and easily accessed via a user-friendly interface (Accessibility).

**Methods**

ProteoSAFe addresses three key sources of volatility: Flexibility in integration of tools, Scalability in leveraging distributed computing, and an Accessible user interface that minimizes tool learning curves. Flexibility is achieved mainly through the concept of workflows, which are defined with a series of simple XML specification files and are decoupled from the software itself to accommodate easy addition and maintenance without the need to modify any code. The increasing demands of resource-intensive tools are addressed with a scalable infrastructure that can delegate analysis to distributed computing resources, such as compute clusters and clouds. Finally, the system is made accessible by a universal user interface for all workflows, which simplifies learning by hiding the details of each tool’s native interface.

**Results**

We demonstrate ProteoSAFe’s integration and search capabilities using multiple identification/quantification tools, and evaluate its scalability using a publicly available HEK293 human embryonic cell line dataset. ProteoSAFe currently integrates MS database search (InsPecT, MS-GFDB), PTM analysis (MS-Align, SpectralNetworks, MODa), genome annotations (Proteogenomics), spectral clustering (MS-Cluster, SpectralArchives), De novo sequencing (PepNovo, Shotgun Protein Sequencing), spectral library searches (M-SPLIT), quantification (OpenMS) and Top-Down analysis (MS-Align*-).

ProteoSAFe’s scalable infrastructure is capable of delegating analysis to compute clusters and thus speeds up time-consuming tasks such as proteogenomics or blind searches. Its simple, user-friendly interface has made the tools accessible to over 1,000 researchers at various institutions, who have now used ProteoSAFe to altogether search over 300 million spectra in more than 10,000 separate search jobs.

To enable a fair comparison of the available search tools, HEK293 search results were compared on a search space supported by all tools and after imposing 1% False Discovery Rates at the level of Peptide Spectrum Matches (PSMs). The diversity of ProteoSAFe search tools makes it necessary to use such a strategy since it is otherwise unclear how to compare single-peptide unmatched identifications with M-SPLIT’s mixture-spectra (2 peptides per spectrum) and MS-Alignment’s blind searches (where unclear PTM site assignments complicate assessment of unique peptides).

By enabling the redistribution and reproducible execution of search workflows, ProteoSAFe allows users to obtain the same search results as reported in published manuscripts and greatly simplifies the application of the exact same search procedures to new data. The absence of such features has traditionally complicated the transfer of MS software from developers to third-party users, and is a major reason why labs without dedicated bioinformatics staff prefer to wait before adopting new MS tools.

**Legend**

A) XML Workflow Specifications

B) Architecture

C) User Interface

This work was supported by the National Institutes of Health grant 1-P41-RR024851 from the National Center for Research Resources