Bio Tool Kit - A Complete Set of Tools for Biomolecule Characterization

*In PeakView® Software or SCIEX OS Explorer*

In order to more fully comprehend the function and behavior of biomolecules it is essential to better understand the details of their molecular structure and sequence. The Bio Tool Kit provides a suite of tools for complete characterization of biomolecules such as proteins, peptides, and oligonucleotides using data generated from SCIEX mass spectrometers. The tools within the Bio Tool Kit enable intact molecular weight analysis, sequence analysis, as well as the detection, mapping, and quantitation of any post-translational modifications, mutations, or truncations. Together, these tools can provide a much more comprehensive view of the biomolecule under study.

### Key Benefits of the Bio Tool Kit

- **Reconstruction Workflows**
  - Determine the intact molecular weights of proteins, peptides, oligonucleotides, mAbs, and other biomolecules using the mass reconstruction tools and either a charge series or isotope series in positive ion or negative ion mode.

- **Manual Sequence Workflows**
  - Find the sequence of unknown peptides using the *de novo* sequencing tools and then match the proposed sequence to fragment ions within the MS/MS data.

- **Peptide Mapping Workflows**
  - Perform an *in silico* digest of a protein sequence and then map that sequence to LC-MS data. Confirm peptide matches using MS/MS data.

- **Extensive Data Dictionary**
  - All of the features within the Bio Tool Kit are backed by a database containing over 1300 modifications to enable a comprehensive analysis of any biological or chemical processing that may have occurred to the biomolecule.

---

Figure 1: Protein Reconstruction of Intact Monoclonal Antibody.
(Top) Raw spectrum. (Middle) Reconstructed spectrum showing different glycoforms of the mAb automatically detected in the raw data. (Bottom) Example of processing settings used in reconstruction, user enters the m/z range and spectral resolution of the raw data, then enters the mass range, step mass, and charge agent are entered for the reconstructed spectrum.
Reconstruction Workflows

Determining the intact mass of a biomolecule is often an important first step in biomolecule characterization. Users can choose to reconstruct raw data automatically or in a more manual fashion. With the protein reconstruction tool, the maximum entropy algorithm automatically detects all charge series envelopes in the raw data and produces a reconstructed spectrum annotated with the intact molecular weights of all biomolecules detected in the raw data (Figure 1).

With the manual reconstruction workflow, users have the choice to perform a deconvolution based upon a charge series or isotope series. For a manual reconstruction based upon an isotope series, the user simply selects multiple isotopes of the ion of interest and the tool will calculate the intact molecular weight and annotate the spectrum (Figure 2, top). Peptide masses can also be automatically determined from LC-MS (Figure 2 bottom) by using the LC-MS Reconstruct algorithm. Reconstruction is performed across the chromatographic run, peptide masses are computed and a table of masses is created for use in downstream workflows such as peptide mapping (Figure 5).

![Figure 1: Protein Reconstruction Workflow](image1.png)

**Figure 1.** Protein Reconstruction Workflow. The intact mass of a protein is determined using the maximum entropy algorithm to detect all charge series envelopes in the raw data and produce a reconstructed spectrum annotated with the intact molecular weights of all biomolecules detected in the raw data.

![Figure 2: Manual Reconstruction Using Isotope Series](image2.png)

**Figure 2: Manual Reconstruction Using Isotope Series.** (Top) The 550 m/z peak was selected for reconstruction from a spectrum containing a mixture of tryptic peptides. The reconstruction tool determines the ion charge state and computes the molecular weight. Spectrum is labeled with the mass information and other detected charge states are highlighted. (Bottom) For LC-MS data, the peak detection algorithm will find spectral peaks across the time dimension, determine the molecular weights from the isotope and charge state information, and output a table of masses (Figure 5).

**Figure 3. Manual Peptide Sequencing.** The peak with nominal mass 1092 was highlighted as the starting point for sequencing the peptide from the MS/MS data. The next peaks were then selected for consideration, one after another, with the final sequence shown above the spectrum.

**Manual Sequence Workflows**

Using the tools within the Bio Tool Kit, users can “walk” up or down an MS/MS spectrum to determine the sequence of a peptide (Figure 3). This process is straightforward and is accomplished by following these steps:

1. Highlight a fragment ion as the starting point for manual sequencing.
2. Double-click on the caption for the next peak to consider in the sequence.
3. Continue selecting peaks until finished. The final result shows the peptide sequence ladder.
Peptide sequences can also be matched to MS/MS data with the peptide fragment pane. Here, a table of theoretical fragment ions from the proposed sequence is matched to ions found in the data. All matches are highlighted within the table and labeled in the MS/MS spectrum. The tool considers multiple different fragment ion types and will also find and highlight modifications (Figure 4).

**Peptide Mapping Workflow**

Mapping a protein sequence to mass spectrometry data of a digested protein is one of the key ways to gain detailed insight into the structure and modifications on a protein. With the Bio Tool Kit, the first step to peptide mapping is to identify spectral peaks in the LC-MS data and then reconstruct them to neutral masses from the m/z data (Figure 2). The LC-MS Peptide Reconstruct feature uses a peak finding algorithm to identify groups of peaks that form isotope series and charge series.

Once the peaks have been found, it's time to map the data. Here a protein sequence is digested *in silico* to create a list of peptides. These peptides are then matched to the peaks found in the LC-MS data. Any sample preparation artifacts, missed cleavages, and biological modifications can also be selected and considered for the peptide matching. Matched peptides are highlighted in the reconstructed peak list and the protein sequence is highlighted with all found peptides (Figure 5).

If an LC-MS/MS workflow was used, the MS/MS data for any matched peptides can be compared with their sequence using the Peptide Fragment pane. As with the Sequencing workflow, all fragment ion matches are highlighted in the table and annotated in the spectrum (Figure 4).

---

**Figure 4:** Peptide Fragment Pane. The theoretical fragment ions from the proposed sequence are matched to the fragment ions found in the MS/MS data. All matches are highlighted in the table and annotated in the spectrum.

---

**Figure 5:** Peptide Mapping Workflow. The LC/MS data is automatically processed to calculate the masses of all peaks found in the data. Then a protein sequence is digested in-silico and the resulting peptides matched to the peaks found in the data. All peaks that are matched are highlighted in the reconstructed peak list table (bottom right). Additionally, the protein sequence is highlighted with all found peptides (bottom left). Mapping can be accomplished using a broad array of sample preparation and biological modifications.
Modification Catalog

Behind all of the features within the Bio Tool Kit lies the Data Dictionary. This dictionary contains detailed information about the building blocks for biomolecules such as amino acids and nucleotides. Additionally, the dictionary contains over 1300 biological, sample preparation, and substitution modifications including modifications from the UniMod database (Figure 6). Users may even edit the dictionary to add custom modifications if desired. This modification catalog is used in all relevant SCIEX software packages to ensure consistency across experiments and data processing:

- Bio Tool Kit (in either PeakView® Software or SCIEX OS Explorer) for peptide mapping, de novo sequencing, and reconstruction workflows
- ProteinPilot™ Software for discovery proteomics
- BioPharmaView™ Software for biologics characterization

Conclusions

The Bio Tool Kit is a micro application for use within PeakView Software or SCIEX OS Explorer and provides a suite of tools for biomolecule analysis. The kit provides the most commonly-needed functions for the characterization of biomolecules by mass spectrometry, including:

- Manual Sequence Workflow
- Peptide De novo sequencing
- Peptide MS/MS fragment matching
- Peptide Mapping Workflow
- Protein sequence in silico digest
- MS mapping with MS/MS confirmation
- Reconstruction Workflows
- Intact protein
- Peptide mixtures
- Manual determination
- Oligonucleotide in negative mode
- Extensive Data Dictionary
- Over 500 sample preparation and biological modifications
- Add custom modifications

Extensive list of biological, sample preparation, and substitution modifications (>1300 mods).

Figure 6: Overview of the Modifications in Data Dictionary.