

Bio Tool Kit - a complete set of tools for biomolecule characterization

In PeakView® Software or Explorer in SCIEX OS Software

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.

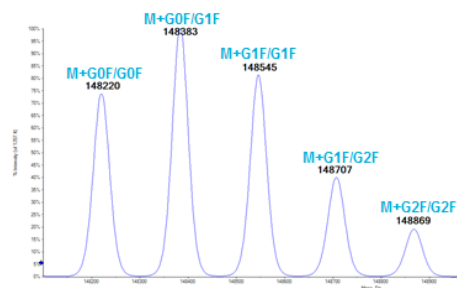
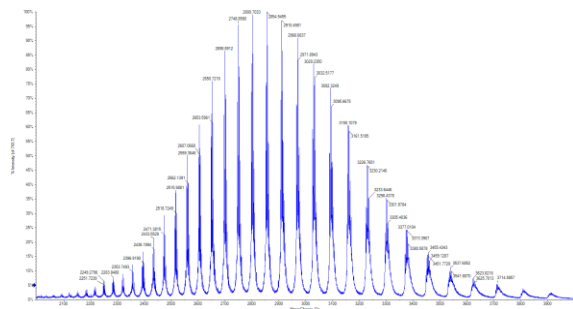
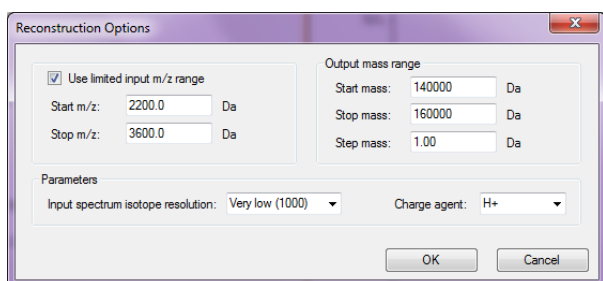



Figure 1: Protein reconstruction of intact monoclonal antibody. (Top) Raw spectrum. (Middle) Reconstructed spectrum showing different glycoforms of the mAb automatically detected in the raw data. (Bottom) Example of processing settings used in reconstruction, user enters the m/z range and spectral resolution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

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

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).

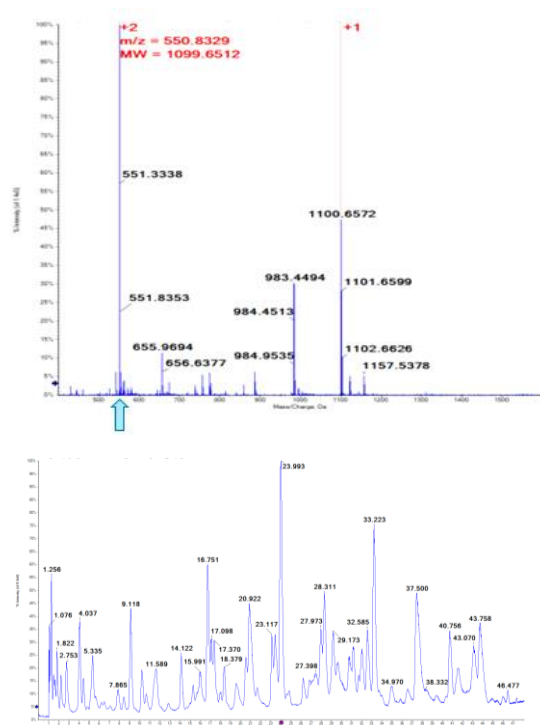


Figure 2: Manual reconstruction using isotope series. (Top) The 550 m/z peak was selected for reconstruction from a spectrum containing a mixture of tryptic peptides. The reconstruction tool determines the ion charge state and computes the molecular weight. Spectrum is labeled with the mass information 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).

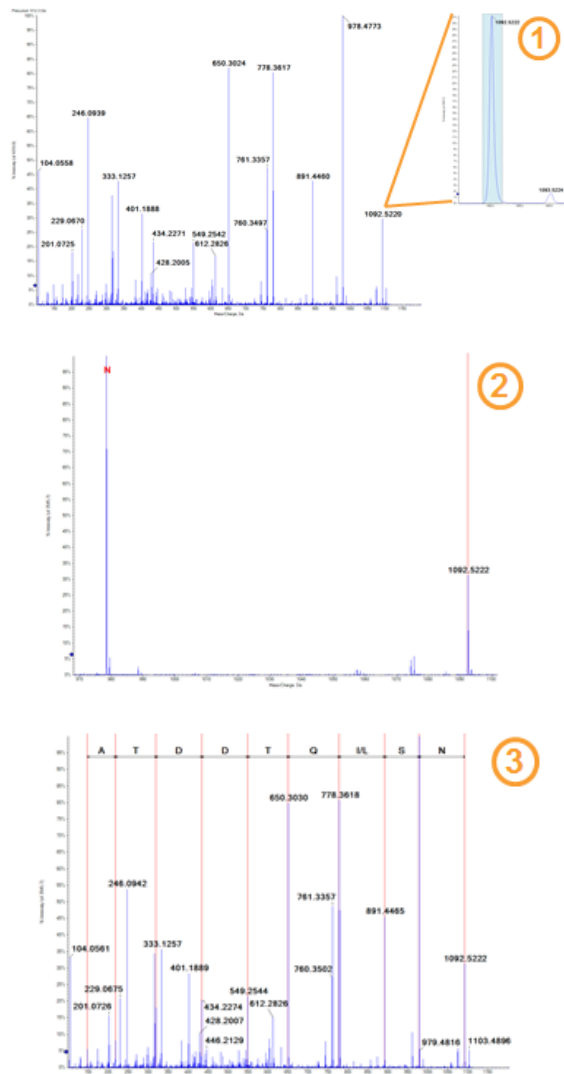


Figure 3. Manual peptide sequencing. The peak with nominal mass 1092 was highlighted as the starting point for sequencing the peptide from the MS/MS data. The next peaks were then selected for consideration, one after another, with the final sequence shown above the spectrum.

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.

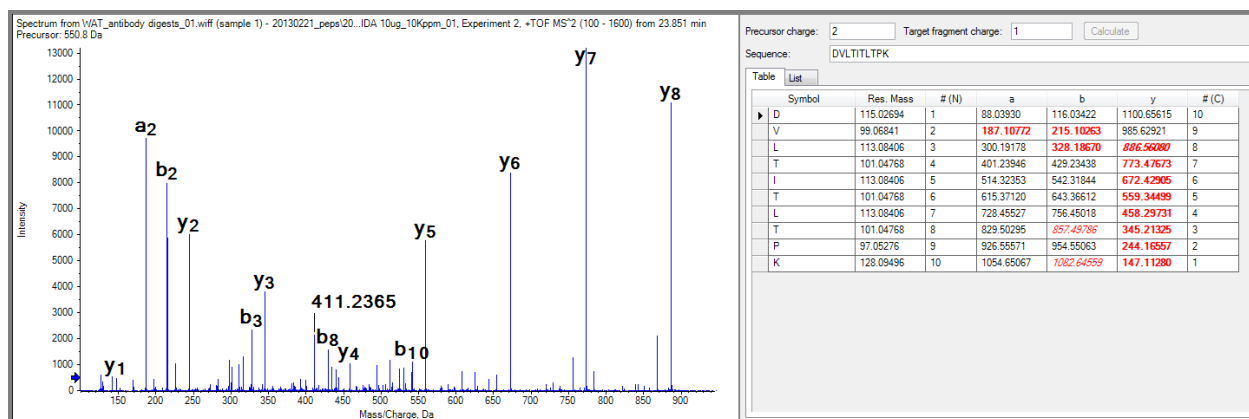


Figure 4: Peptide fragment pane. The theoretical fragment ions from the proposed sequence are matched to the fragment ions found in the MS/MS data. All matches are highlighted in the table and annotated in the spectrum.

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).

Reconstruction Results

Index	Ret. Time	Mass	Mass / Charge	Int. Sum	Num. Peaks	Match
12	31.23	2603.2872	2.89e3	4	T1. DVLMEDVIGTGPLSPVSGGASISQCAMR	
36	33.23	2603.2863	3.21e3	16	T1. DVLMEDVIGTGPLSPVSGGASISQCAMR	
47	26.21	3023.4797	5.99e4	12	T2. SSGYHNSAGNTLEWYLGKPGGSPK	
26	14.12	648.4203	5.52e4	4	T3. LLYIK	
42	23.42	2065.9848	9.79e4	7	T9. TS FSGVPRFSGGSGTDTLTK	
27	8.42	776.3305	8.89e3	4	T5. FSGVPRF	
34	17.12	1302.8087	1.73e3	8	T6. FSGGSGTDTLTK	
102	30.86	3200.9580	7.83e4	19	T7. TS VSRVNEKGVYVCAMFGGSHVPLTRAGTK	
85	32.96	3237.6416	1.40e4	5	T8. VSERDGVYVCAMFGGSHVPLTRAGTKLEK	
51	31.20	2844.3370	7.73e4	18	T8. VSERDGVYVCAMFGGSHVPLTRAGTK	
14	3.77	657.4186	1.26e3	2	T9. TIL LKLR	
18	5.17	501.3149	4.96e3	3	T9. LKLR	
82	41.53	3726.8204	6.11e4	16	T10. T11. RADAPVTSVFFPSSEGLTSGGASVVCAMFLNNYRK	
64	43.13	3570.7287	6.99e4	14	T11. ADAPVTSVFFPSSEGLTSGGASVVCAMFLNNYRK	
12	2.73	587.3263	7.49e3	3	T12. QIVAK	
23	10.66	389.4821	2.39e4	7	T13 - T14. RWDGSSR	
24	10.89	676.3171	1.89e3	2	T14. IDGSSR	
35	17.37	1590.7241	1.39e5	11	T15. GAGVNSWTDGSK	
39	23.09	2225.9990	9.53e4	8	T16 - T17. DSTYSRSLTLTKDEYER	
36	20.82	1533.7233	3.16e4	7	T16. DSTYSRSLTLTK	

Digest Protein Pane

Protein	Sequence	Peptide	Mass	Charge	Modifications
1	DAVLMEDVIGTGPLSPVSGGASISQCAMR	DAVLMEDVIGTGPLSPVSGGASISQCAMR	2603.2872	2	
2	SSGYHNSAGNTLEWYLGKPGGSPK	SSGYHNSAGNTLEWYLGKPGGSPK	3023.4797	3	
3	LLYIK	LLYIK	648.4203	1	
4	TS FSGVPRFSGGSGTDTLTK	TS FSGVPRFSGGSGTDTLTK	2065.9848	2	
5	FSGVPRF	FSGVPRF	776.3305	1	
6	FSGGSGTDTLTK	FSGGSGTDTLTK	1302.8087	1	
7	TS VSRVNEKGVYVCAMFGGSHVPLTRAGTK	TS VSRVNEKGVYVCAMFGGSHVPLTRAGTK	3200.9580	2	
8	VSERDGVYVCAMFGGSHVPLTRAGTKLEK	VSERDGVYVCAMFGGSHVPLTRAGTKLEK	3237.6416	1	
9	VSERDGVYVCAMFGGSHVPLTRAGTK	VSERDGVYVCAMFGGSHVPLTRAGTK	2844.3370	2	
10	TIL LKLR	TIL LKLR	657.4186	1	
11	LKLR	LKLR	501.3149	1	
12	T10. T11. RADAPVTSVFFPSSEGLTSGGASVVCAMFLNNYRK	T10. T11. RADAPVTSVFFPSSEGLTSGGASVVCAMFLNNYRK	3726.8204	2	
13	ADAPVTSVFFPSSEGLTSGGASVVCAMFLNNYRK	ADAPVTSVFFPSSEGLTSGGASVVCAMFLNNYRK	3570.7287	2	
14	QIVAK	QIVAK	587.3263	1	
15	T13 - T14. RWDGSSR	T13 - T14. RWDGSSR	389.4821	1	
16	T14. IDGSSR	IDGSSR	676.3171	1	
17	T15. GAGVNSWTDGSK	GAGVNSWTDGSK	1590.7241	1	
18	T16 - T17. DSTYSRSLTLTKDEYER	T16 - T17. DSTYSRSLTLTKDEYER	2225.9990	2	
19	T16. DSTYSRSLTLTK	DSTYSRSLTLTK	1533.7233	1	

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

Modifications in UniMod	1657
Modifications in SCIEX catalog	1378
Common Modifications	1126
Union of Modifications	1909

ProteinPilot Software-Related Statistics	
Modifications Used in Any Set (includes substitution)	1052
Workup Modification Set	63
Biological Modification Set	337
Quantitation-Related Modifications	228
Substitutions	460



Figure 6: Overview of the modifications in data dictionary.
Extensive list of biological, sample preparation, and substitution modifications (>1300 mods).

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 800 sample preparation and biological modifications
 - Add custom modifications

The SCIEX clinical diagnostic portfolio is For In Vitro Diagnostic Use. Rx Only. Product(s) not available in all countries. For information on availability, please contact your local sales representative or refer to <https://sciex.com/diagnostics>. All other products are For Research Use Only. Not for use in Diagnostic Procedures.

Trademarks and/or registered trademarks mentioned herein, including associated logos, are the property of AB Sciex Pte. Ltd. or their respective owners in the United States and/or certain other countries.

© 2017 DH Tech. Dev. Pte. Ltd. RUO-MKT-02-5248-A. AB SCIEX™ is being used under license.