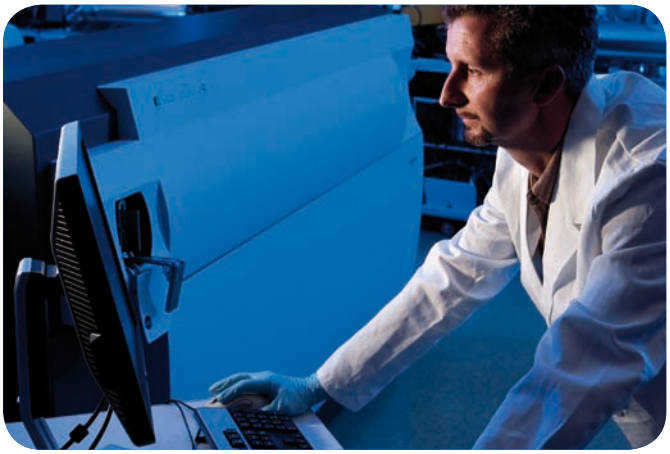
Bruker Daltonics



Technical Note # TN-37

Towards 100% sequence coverage in protein QC: In-depth characterization of monoclonal antibodies using the ProteinScape database software

QC related analysis of biotherapeutic proteins requires experimental verification of the complete amino acid sequence, including any modifications that are present. This is typically not achieved in a single LC-MS/MS experiment and multiple analyses are therefore necessitated. However, this significantly complicates the downstream data analysis due to distribution of relevant information in multiple datasets. A new bioinformatics platform, ProteinScape™, is applied here to organize and analyze substantial amounts of LC-MS/MS data obtained from the sequence analysis of monoclonal antibodies (mAb) at an unparalleled level of efficiency. ProteinExtractor™, a unique algorithm implemented within ProteinScape, merges data obtained from various complementary analytical techniques into a single analysis report. For the recombinant antibody sample analyzed here, 100 % protein sequence coverage was achieved for the light chain and 99 % for the heavy chain together with the rapid, unambiguous assignment of the N-terminal HC pyroglutamylation.

Introduction

Monoclonal antibodies represent one of the most important classes of biotherapeutics. According to common requirements in quality control (QC), recombinant protein products need to be characterized in detail. This includes complete (or as complete as possible) confirmation of the expressed protein sequence and all modifications using experimental data. To meet these requirements, various analytical methods often need to be applied simultaneously. Such combined approaches may involve the use of multiple protein and peptide separation techniques (e.g. SDS-PAGE, HPLC), various enzymes and complementary LC-MS/MS platforms. For such in-depth protein sequencing projects, seamless integration, organization and efficient analysis of the resulting data represents a major challenge. The availability of suitable bioinformatics tools is therefore of crucial importance. Bruker´s ProteinScape database software provides a new level of efficiency and convenience for the organization and analysis of large scale MS and MS/MS data acquired throughout extensive protein QC experiments. The software allows the generation of a flexible project hierarchy that reflects in detail all the analytical workflows applied in the lab. Furthermore, the unique ProteinExtractor algorithm implemented in ProteinScape compiles multiple experiments (e.g.. various enzymes, separation techniques, MS/MS platforms) into a single, non-redundant overall result by merging all information generated from the individual workflows utilized. To illustrate the capabilities of ProteinScape in typical protein characterization applications, a monoclonal

antibody (mAb) sample was used here for in-depth sequence confirmation by LC-MS/MS following multiple enzymatic digests.

Experimental

Sample preparation: A human IgG1 expressed in Chinese hamster ovary cells was analyzed.
The mAb was reduced and alkylated using dithiothreitol and iodoacetamide as reagents. An aliquot of the reduced and alkylated sample was separated by SDS-PAGE. The mAb heavy (HC) and light (LC) chains were subjected to in-gel digestion using trypsin as an enzyme. Further sample aliquots were digested in solution using various enzymes (trypsin, chymotrypsin, GluC, LysC).

LC-MS/MS: Automated LC-MALDI-TOF/TOF experiments were performed by coupling an EASY-nanoLC™ to a PROTEINEER fc II™ fraction collector (both Bruker Daltonics) and analyzing the spotted fractions with the Bruker ultrafleXtreme™ MALDI-TOF/TOF. Electrospray LC-MS/MS measurements were carried out using the Ultimate3000™(Dionex) nanoLC coupled to either a maXis™ UHR-TOF or a 12 Tesla solariX™ FTMS (both Bruker Daltonics).

Data analysis: All LC-MS/MS data was organized and analyzed in Bruker´s ProteinScape database software. Final compilation of results generated across the various analytical workflows and instrument platforms was performed using the ProteinExtractor algorithm.

Results

Fig. 1 shows the graphical user interface of ProteinScape. During development, particular attention was paid to ensure that the software can be navigated in a highly intuitive manner for optimal data organization and analysis. ProteinScape follows the concept of mirroring in-silico both, the project structure and the workflows typically applied in a protein analysis lab. To achieve this, the project navigator (left upper part of Fig. 1) allows the user to create a customized project tree containing multiple sublevels. For the analysis of the monoclonal antibody presented here, the generated project hierarchy reflects all the steps applied throughout sample preparation (SDS-PAGE; enzymatic digestion performed in-gel and in-solution, respectively), LC-MS/MS measurement and final data analysis. For each of these steps, all relevant method parameters can be stored in the ProteinScape database. One of the great benefits of ProteinScape´s data warehousing concept is that there is no need to load individual datasets before working on them. All data is permanently accessible by simple navigation through the project treeview interface. Furthermore, all information details included in a particular dataset (protein level, peptide level, chromatographic perspective, single MS and MS/MS spectra) are accessible by a single mouseclick. To ensure high processing speed even for large datasets, ProteinScape handles peaklists instead of raw data. However, raw data is available on demand at any time along the data analysis workflow.

ProteinExtractor algorithm: Generating integrated results from multidimensional experiments

For maximum sequence coverage multiple enzymes of complementary cleavage characteristics (trypsin, chymotrypsin, GluC, LysC) were used to digest the mAb. Subsequent LC-MALDI-TOF/TOF analyses yielded individual sequence coverages between 64 and 88 % (Fig. 2).

Intuitive navigation through the project

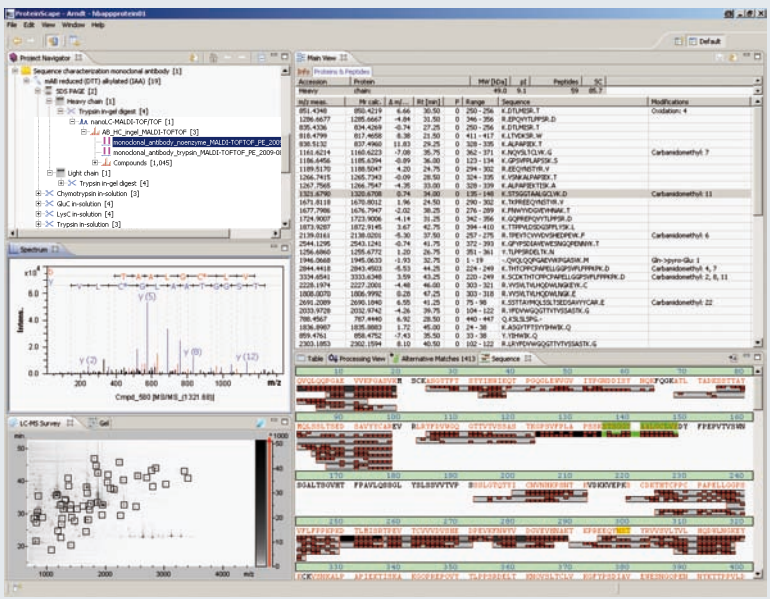


Fig. 1: Navigating the virtual lab: ProteinScape allows organizing data in an appropriate structure representing the analytical workflows applied in a particular project. All information featured by a dataset (protein level, peptide level, chromatographic perspective, single MS and MS/MS spectra) is accessible by a single mouse click.

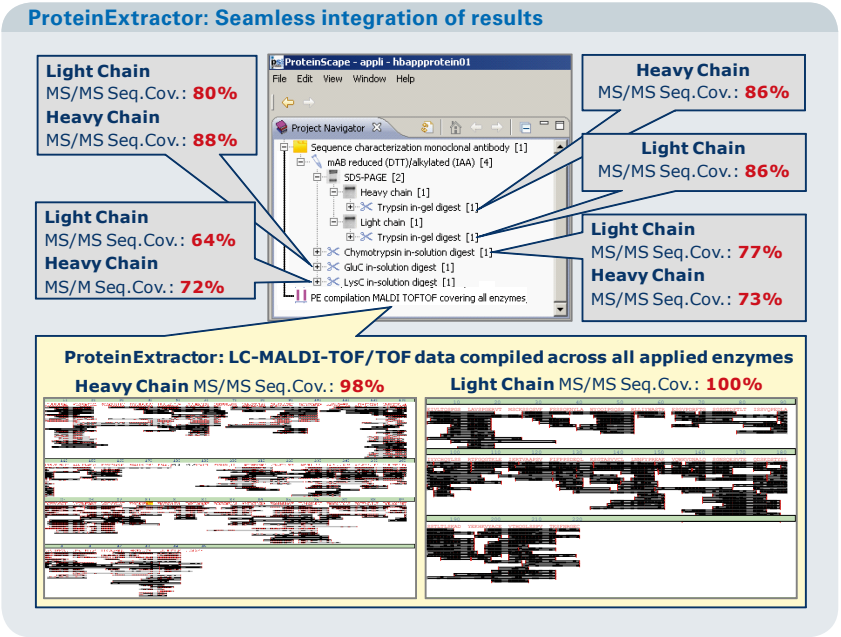


Fig. 2: The unique ProteinExtractor algorithm integrates results obtained from multiple experiments into a single sequence coverage report: Compilation of individual LC-MALDI results obtained from four enzymatic digests results in 100 % sequence coverage for the LC and 98 % for the HC. The sequence coverage by peptides (grey bars) and by MS/MS fragment ions (red bricks) is immediately visible in the sequence plot representing the compiled dataset.

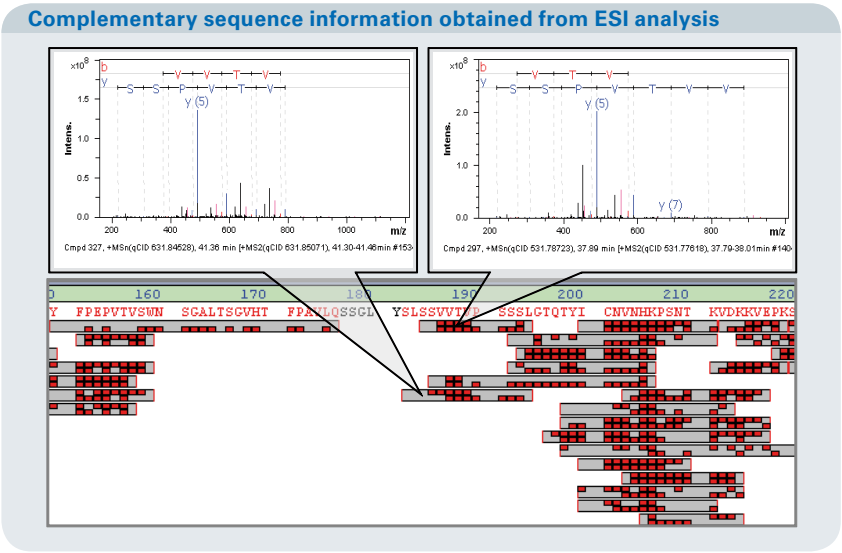


Fig. 3: Partial sequence view displaying the MS/MS sequence coverage achieved for the HC based on ESI and MALDI results compiled by ProteinExtractor: ESI-FTMS measurements added 2 unique peptide MS/MS spectra covering amino acid residues HC-[182-184] that were not accessible by MALDI.

The sequence coverage increased significantly when these individual digest analyses were merged into a single sequence confirmation report. This otherwise demanding compilation step is easily performed by the ProteinExtractor algorithm with a few mouse clicks, and thereby the MALDI data yielded 100 % sequence coverage for the light chain and 98% for the heavy chain (see Fig. 2).
Further sequence information for the HC range 177-184 (which went undetected with MALDI) was obtained by additional electrospray measurements. Again, ProteinExtractor combined ESI and MALDI data into an overall sequencing result, which further increased the sequence coverage for the HC to 99%. As shown in Fig. 3, ESI-MS/MS measurements contributed MS/MS spectra for 2 additional peptides, providing coverage for 3 out of the 8 amino acid residues (HC 182-184) that had not been previously detected by LC-MALDI.

Direct access to specific information

The ProteinScape database software provides rapid access to specific analytical and biological information through querying, sorting, filtering and comparing data by protein and peptide features (e.g. modifications, sequences, analysis parameters, and others). For example, a search for peptide sequences assigned to the HC N-terminus yielded a list of peptides that contain the expected N-terminal pyroglutamylation [Gln-1 to pyroGlu-1] (see Fig. 4). Evaluation of the respective MS/MS spectra, which are directly accessible in ProteinScape, confirmed the presence of this N-terminal modification based on consistent ESI and MALDI data.

Direct access to specific information stored in the database

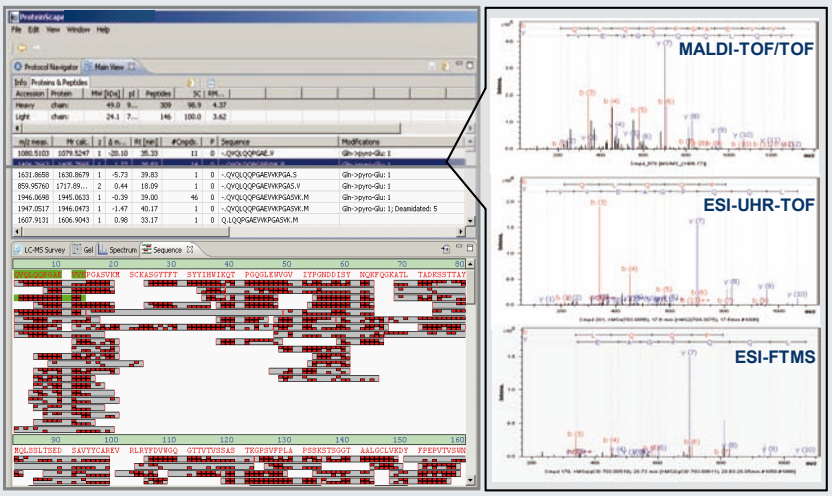


Fig. 4: ProteinScape provides direct access to specific analytical and biological information: The figure shows compiled results from 3 different LC-MS/MS platforms queried for peptides representing the N-terminus of the antibody´s heavy chain, confirming N-terminal pyroGlu modification. A direct link to the corresponding MS/MS spectra reveals consistency of the MS/MS data obtained from multiple MS platforms. As an example, MS/MS spectra of the N-terminal peptide pyroQVQLQQPGAEVVK are displayed here.

Summary

ProteinScape was applied to the sequence analysis of a monoclonal antibody and served as a powerful tool for the streamlined analysis of data obtained from multiple LC-MS/MS analyses of various enzymatic digests performed on both MALDI and ESI mass spectrometers. The unique ProteinExtractor algorithm implemented within ProteinScape merges data obtained from various complementary analytical techniques (different separation methods, various MS/MS instruments etc.) into a single analysis report to fully exploit the benefits offered by multidimensional analytical strategies. Extensive query and sorting functionalities implemented in ProteinScape ensure instant access to specific analytical and biological information featured by a particular dataset.
For the monoclonal antibody sample analyzed here, the overall combined results provided 100% protein sequence coverage for the light chain and 99% for the heavy chain. In addition, ProteinScape facilitated direct screening of the data obtained from multiple MS/MS platforms to confirm Gln to pyroGlu conversion of the heavy chain´s N-terminus.

References

Bruker Daltonics Application Note #ET-17/#MT-99: Characterization of the N-glycosylation Pattern of Antibodies by ESI - and MALDI mass spectrometry

Authors

Arndt Asperger, Ulrike Schweiger-Hufnagel, Peter Hufnagel, Romano Hebeler, Matthias Witt, Pierre-Olivier Schmit, Marcus Macht

Keywords
monoclonal antibody
protein characterization
sequence coverage
LC-MS/MS
ESI
MALDI
bioinformatics

Instrumentation & Software
ultrafleXtreme
maXis
solariX
EASY nanoLC
PROTEINEER fc II
ProteinScape
microTOF-Q II
amaZon

For research use only. Not for use in diagnostic procedures.

www.bdal.com

Bruker Daltonik GmbH

Bremen · Germany
Phone +49 (0)421-2205-0
Fax +49 (0)421-2205-103
sales@bdal.de

Bruker Daltonics Inc.

Billerica, MA · USA
Phone +1 (978) 663-3660
Fax +1 (978) 667-5993
ms-sales@bdal.com