TopSpec - 829157

WP7 - Signal detection and data processing

Deliverable: D7.2 **-** Top-down analysis software.
Top-down data analysis software distributed to participants.
Includes implementation of an approach to mass spectra deconvolution via transient decay rates.

Task: Develop a software suite for analysis of top-down mass spectrometry data focused on the monoclonal antibodies structural characterization.

Content:

Author: Yury Tsybin, Spectroswiss

Deadline: 11/30/2021

1. Executive summary.

- The selected developed data analysis tools and approaches have been integrated into a software suite with a graphic user interface, Peak-by-Peak BioPharma;
- The current functionality of the Peak-by-Peak BioPharma supports data processing and analysis required for structural characterization of monoclonal antibodies (mAb) of a direct interest to TopSpec, including: intact mass analysis of whole mAbs, middle-up analysis of mAb subunits, top-down analysis of whole mAbs, and middle-down analysis of mAb subunits;
- The implemented workflows support the two modern approaches for mAb analysis: with and without deconvolution;
- The modular architecture of Peak-by-Peak BioPharma allows embedding additional functionalities developed by TopSpec partners, specifically those implemented by KI, TNTU, IP, and FasmaTech;
- The innovative CHARDA (charge determination analysis) workflow is available as an option in Peak-by-Peak BioPharma. The workflow computes the signal decay rates in the time-domain transients to assign charge states to the individual peaks in mass spectra. The latter facilitates analysis of complex product ion patterns in the top-down analysis of mAbs;
- The compiled software suite, Peak-by-Peak BioPharma, has been distributed to the interested parties of the TopSpec consortium and is available for installation to all TopSpec partners.

2. Description of the action.

The reach the objectives of the WP7's deliverable D7.2, the leading partner, SPS, collaborated with the involved TopSpec partners, including KI, IP, NTNU, and Fasmatech. The main focus of D7.2 is on the structural analysis of monoclonal antibodies (mAbs) with top-down MS approaches.

The biotherapeutics-grade mAbs are generally from the immunoglobulin G (IgG) class. An individual IgG molecule is structurally a tetrameric glycoprotein complex composed of two ~50 kDa heavy chains and two ~25 kDa light chains. Four IgG isotypes exist and are defined by their heavy chain amino acid sequence: IgG1, IgG2, IgG3, and IgG4. The IgG3 mAbs are not used as therapeutics due to their significantly faster clearance than IgG1/IgG2 isotypes (7 vs. 21 days in some cases). The disulfide bridges (16 for IgG1 and IgG4; 18 for IgG2) and non-covalent interactions maintain the three-dimensional structures of mAbs. The light and heavy chains (Lc and Hc) are linked by a single disulfide bond and the heavy chains by two (for IgG1 and IgG4) or three (for IgG2) disulfide bonds located in a short hinge domain. The other 12 cysteine bridges are intramolecular and delimit six different globular domains, variable and constant. Antigen binding is mediated by the variable domains, mainly by the three loops connecting individual β-strands in each complementarity determining region (CDR). Like the natural IgGs, all recombinant antibodies contain an -Asn-X-Ser/Thr-Y- consensus sequence for N-glycosylation in their heavy chain constant domain (where X and Y are amino acids different from proline). As a result, a typical mAb-based biotherapeutic is represented by a defined set of proteoforms (glycoforms).

Following the structural details of mAbs and mAbs-derived biotherapeutics, the following information needs to be provided by mass spectrometry. First, a general elemental composition (amino acids and modifications) and the amino acid sequence of the mAbs are to be verified. This analysis shows the structural completeness of a mAb molecule in the development or production process. Second, a relative percentage of the main glycoforms (proteoforms) is to be deduced (glycosylation profiling). Third, expected and unexpected modifications (including deamidation, isomerization, oxidation, and glycation) are to be identified, quantified, and, preferably, localized. Forth, integrity and scrambling of the inter- and intra-molecular disulfide bonds are to be controlled. Further analysis may include charge variant analysis, *de novo* sequencing, light and heavy chain pairing, estimation of the drug-to-antibody ratio (DAR) values for antibody-drug conjugate (ADC) samples, mAb aggregation extent, *etc*.

The high structural complexity of the mAbs requires a combination of the MS approaches for their comprehensive structure description. In line with the TopSpec action vision and activities, the commonly employed bottom-up approaches are to be necessarily complemented with the "top-down view" on the mAbs: mass measurements at the intact and subunit (middle-up) levels, and product ion identification following gas-phase fragmentation of precursor ions at the intact (top-down) and subunit (middle-down) levels, **Figure 1**. The mass spectrometric analysis of the mAb subunits is supported by advances in the sample preparation area offering novel, highly-specific enzymes, particularly for mAbs processing, e.g., hinge-region specific enzymes IdeS and KGP.

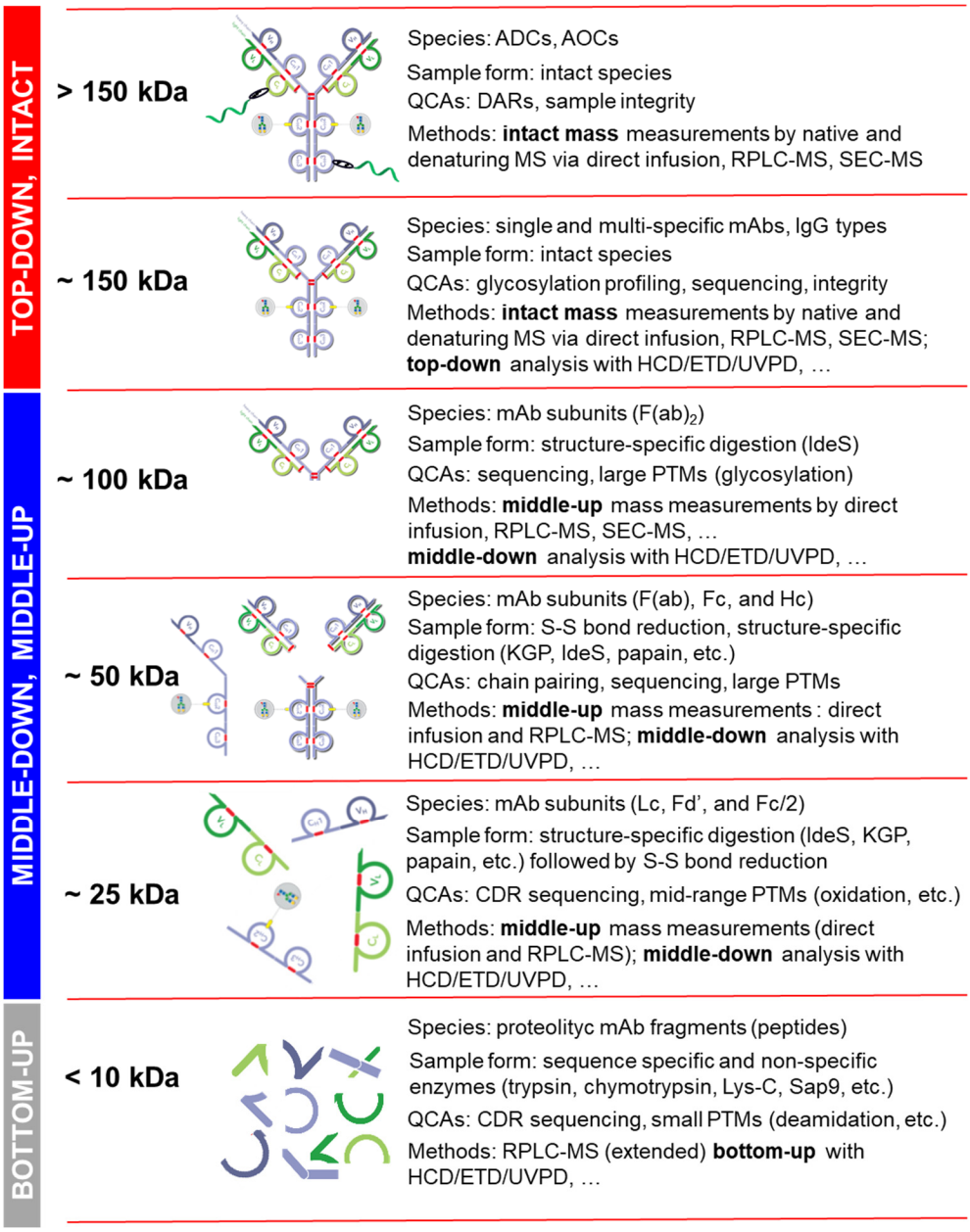

Figure 1. A molecular weight-organized overview of the MS-based approaches to the structural analysis of mAbs and mAb-derived species. The acronyms and abbreviations are the following. mAb: monoclonal antibody; IgG: immunoglobulin G; ADC: antibody-drug conjugate; AOC: antibody-oligonucleotide conjugate; QCA: critical quality attribute; DAR: drug-to-antibody ratio; RPLC: reversed-phase liquid chromatography; SEC: size exclusion chromatography; MS: mass spectrometry; HCD: higher-energy collision-induced dissociation; ETD: electron transfer dissociation; UVPD: ultraviolet photodissociation; CDR: complementarity determining region; PTM: post-translational modification; Lc: light chain; and Hc: heavy chain.

To support the mAbs' structural characterization needs and workflows overviewed in Figure 1, the following activities related to top-down data analysis have been realized within TopSpec project:

- Several TopSpec project participants (SPS, TNTU, IP, KI, and Fasmatech), have developed algorithms and software capabilities to process and analyze the envisioned and acquired experimental data on mAbs analysis following the originally proposed top-down strategy;

- Based on our benchmarking and evaluation of the software tools, and having reviewed various integration options, as well as following the recommendations from the external experts formulated following July 2021 project reporting, it was decided that Peak-by-Peak BioPharma (from SPS) will be the backbone of the TopSpec data analysis software, **Figure 2**, whereas other tools and functionalities developed by TopSpec partners will be added following the "nodes" principles of the modern software, e.g., Proteome Discoverer;

- The thus selected top-down data analysis software tool, Peak-by-Peak BioPharma, was developed using Python, C/C++, and Qt on top of the Peak-by-Peak framework for FTMS data processing and data analysis (Spectroswiss). The employed versions of the third-party components were contemporary with the Python release 3.8.2;

- SPS, KI, and NTNU co-developed algorithms for the time-domain transient decay analysis which constituted the core of the CHARDA workflow. After a series of original implementations by all partners, KI took the lead in programming the main CHARDA functionality. The final implementation was provided as an executable module to SPS for embedding into the framework – Peak-by-Peak BioPharma.

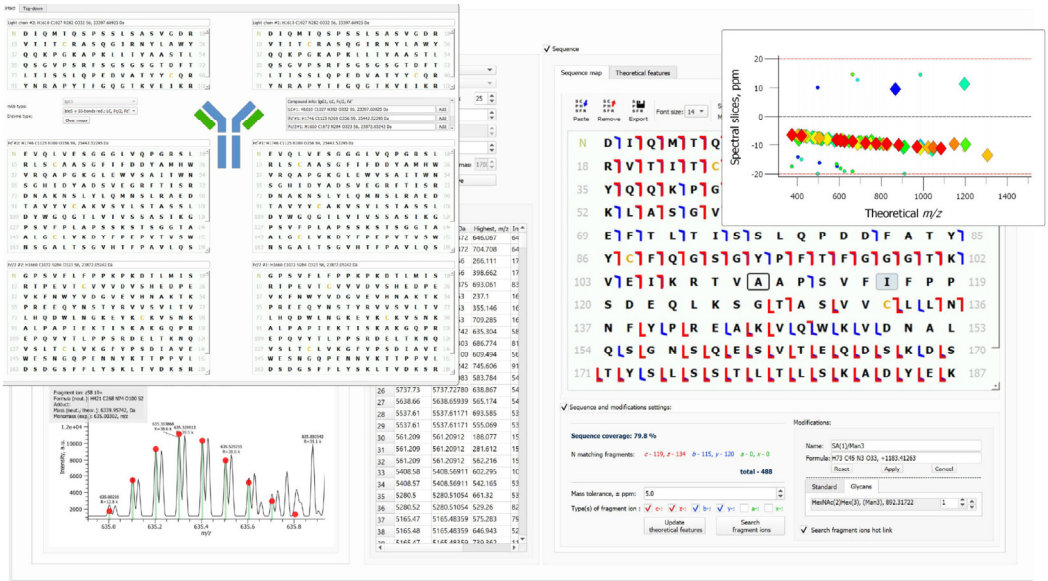

Figure 2. A representation of Peak-by-Peak BioPharma's graphic user interface and options.

3. Performance and functionality.

The key features of Peak-by-Peak BioPharma software suite are the following (the focus is made on Orbitrap data, as TopSpec platforms are built around Orbitraps):

- A wide and comprehensive support of input Orbitrap data formats: time-domain transients, reduced and full profile mass spectra in magnitude mode FT (mFT), absorption mode FT (aFT), and enhanced mode FT (eFT). The latter format corresponds to the .RAW file format which represents the commercial file format.

- A complete workflow to process and analyze the time-domain transients. That includes options to truncate the original time-domain transients to the use-defined period and obtain the corresponding (lower resolution) mass spectra. Importantly, the transient decay calculations, which enable CHARDA workflow, are supported in Peak-by-Peak BioPharma as a "node" based on the implementation by KI.

- Advanced mass spectrometry data processing and data analysis solutions for mAb analysis, including complex ADCs/AOCs. As demonstrated, for example, in the TopSpec paper "Drug-to-Antibody Ratio Estimation via Proteoform Peak Integration in the Analysis of Antibody–Oligonucleotide Conjugates with Orbitrap Fourier Transform Mass Spectrometry" by Nagornov *et al., Anal Chem* 2021.

- Superior quality data owing to proprietary methods and techniques for Orbitrap FTMS data. These include a capability to process and analyze the unreduced data.

- Integration of intact mass, middle-up, middle-down, top-down, and bottom-up mass spectrometry approaches for the structural analysis of mAbs

- Data averaging of LC-MS/MS and MS/MS data from multiple experiments. The advantages of these approaches have been outlined in the following papers:

- Recalibration of intact (MS data) and middle/top-down (MS/MS data) mass spectra. TopSpec project has demonstrated that mass accuracy for product ion identification is to be substantially improved compared to the current state-of-the-art (in the published literature, which is 10-20 ppm). TopSpec partners, specifically Fasmatech, have demonstrated that top-down mass spectra of mAbs may contain a very large part of the total signal, above 40%, distributed among the internal, and not terminal, product ions. Therefore, options and capabilities to re-calibrate the top-down mass spectra is seen as the absolute must of modern data analysis workflows.

- Proteoform-specific accurate simulation of Orbitrap mass spectra of mAbs. The structural complexity of mAbs creates an even more complex response function from Orbitrap mass

spectrometers. In fact, "assuming" or "estimating" the outcome of the experimental data may very easily be wrong. That is well exemplified by a number of statements made in peer-reviewed papers. To remove the "guess" from the design of Orbitrap experiments and enhance quality of data acquisition and analysis, we implemented a special workflow to rigorously simulate the mAbs mass spectral data for specific models of Orbitraps. The tool is now a part of FTMS Simulator and constitutes a part of Peak-by-Peak BioPharma.

- Embedded database of common biotherapeutic-grade mAbs with sequences and modifications. This database facilitates user operation with the input of mAb's data into the FTMS Simulator.

- Comparison of experimental and simulated isotopic envelopes (narrowband data) and mass spectra (broadband data) of mAbs – can be performed on the intact mass level or on the subunit level.

- Work with mAbs intact mass data: high-resolution and low-resolution deconvolution, charge state grouping, and quantitation. Interestingly, the main workflow required by BioPharma nowadays is still based on the intact mass measurements supported by the low-resolution deconvolution. We thus evaluated several low-resolution deconvolution workflows. The Peak-by-Peak BioPharma offers to perform low-resolution deconvolution using the UniDec workflow (licensed to Spectroswiss) as well as via proprietary workflows developed by NTNU (David Kilgour) as a part of TopSpec action. The NTNU workflows are supported as "nodes" in TopSpec tool.

- In-depth work with mAbs top-down data: Intact mass and product ion assignment to the deconvolved data, as well as an interactive graphical representation of sequence maps (add proteoforms, remove assignments, etc.).

- Advanced and envisioned functionality: several advanced algorithms for top-down data analysis have been implemented by TopSpec partners. These functionalities may be converted into the "nodes" of Peak-by-Peak BioPharma upon need and agreement at the later stage of project progression. The examples of the advanced nodes specifically developed for top-down analysis include: internal product ion management (Fasmatech and NTNU) and *de novo* mAb sequencing (Fasmatech, NTNU, and Spectroswiss).

- Ease of use and ergonomic aspects: Peak-by-Peak BioPharma is realized in the graphical user interface (GUI) form and supports high-speed computation and visualization, some of which are implemented from the first principles.

- The output of the Peak-by-Peak BioPharma is multiple, including a graphical output (images), mass spectral data and peak lists.

4. Installation and Operation Protocols.

The first version of a user guide dedicated to the Peak-by-Peak BioPharma has been prepared to guide the user through the installation and operation protocols. During the remaining phase of TopSpec project, the user guide will be further completed with more detailed description and address the needs of both the first time and the experienced user. The user guide will be a part of the deliverable D7.3 which will also describe the use of the external data acquisition systems FTMS Booster TD.