Tools for the analysis and characterization of therapeutic protein species

Marceline Manka Fuh
Pascal Steffen
Hartmut Schlüter

Mass Spectrometric Proteomics, Institute for Clinical Chemistry and Laboratory Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Abstract: A continuously increasing number of therapeutic proteins are being released into the market, including biosimilars. In contrast to small organic drugs, therapeutic proteins require an extensive analysis of their exact chemical composition because of their complexity and proof of the absence of contaminants, such as host cell proteins and nucleic acids. Especially challenging is the detection of low abundant species of therapeutic proteins because these species are usually very similar to the target therapeutic protein. However, the detection of these species is very important for the safety of patients because a very small change of the exact chemical composition may cause serious side effects. In this review, we give a brief overview about the most important analytical approaches for characterizing therapeutic protein species and their contaminants and focus on the progress in this field during the past 3 years. Top-down mass spectrometry of intact therapeutic proteins in the future may solve many of the current problems in their analysis.

Keywords: therapeutic protein species, biosimilars, liquid chromatography, mass spectrometry, capillary electrophoresis

Introduction
Therapeutic proteins have become the leading drugs in the biologic class among vaccines and diagnostics in the drug market. They gained ground in the 1980s and have rapidly grown since then with over 200 marketed products greatly benefiting the quality of life of millions of patients worldwide. The biotherapeutic market consists of peptides, small therapeutic proteins, and nonimmune proteins with the fastest growing product class being therapeutic antibodies and their related products which include antibody–drug conjugates (ADC), Fc–fusion proteins, and antibody fragments. Production and formulation of these therapeutic proteins exhibit very critical and technical challenges from those set by the traditional small molecule drugs. These complex molecules are produced from a very broad platform of expression systems such as bacteria like Escherichia coli, yeast, mammalian, insect, and plants. These generate a variety of process-related impurities, such as contamination with host cell proteins (HCPs), as well as diverse species of the therapeutic target protein, which can copurify with the therapeutic target protein thereby requiring very robust, sensitive, and selective identification as well as effective and economic purification techniques. Depending on the therapeutic proteins created and the host cell system, the response to HCPs could potentially vary from negligible to quite severe, including anaphylactic shock or cytokine release syndrome.
The structure of therapeutic proteins range from relatively unordered to highly ordered, multimeric states. Their activity and side effects are highly affected by their exact chemical composition. Product-related species, such as aggregates, unwanted glycosylated species of monoclonal antibodies, or charge variants, are of great concern. Such changes in the exact chemical composition could not only affect the final product but also the efficiency of the purification steps involved if not detected. The exact chemical composition of the final product is therefore of critical concern as very minor changes in the composition can lead to the appearance of unwanted protein species, which could lead to dangerous outcomes, such as decrease or total loss of activity of the active pharmaceutical ingredient resulting in the change of the intended efficacy or potential toxicity to patients. These variant protein species are mostly present in lower concentrations in relation to the main target product species thereby making identification and separation even harder. In order to ensure product consistency, quality, and purity, the manufacturing process has to be tightly monitored since it is quite challenging to fully characterize the finished product in the laboratory.

It was estimated that 32 biologics, with a combined US $51 billion of sales in 2009, will lose patent protection by 2015; this gives room for biosimilars to emerge. Unlike generic small molecules, there are a lot of difficulties associated with the production of biosimilars, which is heightened by their high degree of complexity. Since biosimilars are similar but not identical to the reference product, these slight differences have to be extensively characterized to ensure that these deviations from the original validated target species have no effect on the dosage and patient safety. The approval process for biosimilars relies primarily on comprehensive analytical data at quality, preclinical, and clinical levels to establish comparability and high similarity with the originator biologic. Basically, for these drug classes, “the product is the process.” Excellent reviews giving comprehensive overviews about the analysis of therapeutic proteins have been published recently. In Table 1, an overview of the methods applied for the analysis of therapeutic proteins, including further important reviews and some comments about benefits and pitfalls, is given.

Thus, in this review, we focus on research papers published mainly in the last 3 years with a focus on the analysis of therapeutic proteins. In addition, we give the reader a general brief overview of the current analytic tools available for their characterization and discuss the most recent developments in the field.

**Process-related impurities**

Upstream process conditions are responsible for a lot of impurities accompanying the protein of interest, such as host cell derived molecules, cell culture media components, solvents, and buffers used for chromatographic purifications. These can nonetheless be reduced by accurate and well-monitored manufacturing procedures.

A petrifying drawback in therapeutic protein production for clinical use is HCP and DNA contamination and their detection, identification, and elimination in the final drug product to acceptable levels. There also exists the possibility of HCP-associated product damage in cases where they possess proteolytic activities. As the identification rate of HCP and sensitivity of the methods employed still remain ambiguous, the most common high-throughput methods used are the immunoassays, such as enzyme-linked immunosorbent assay (ELISA) and Western blot. However, a major limitation to the HCP ELISA methods is the inability to detect non- or low immunoreactive HCP and its dependence on antibody– antigen binding conditions leading to the under-quantification of particular HCPs which can be potentially deadly. The rise of other orthogonal methods, such as liquid chromatography (LC) and mass spectrometry (MS), in the pharmaceutical industry has alleviated these shortcomings and greatly complemented HCP ELISA. MS-based quantitative proteomics incorporates methods, such as two-dimensional difference gel electrophoresis and two-dimensional high pressure LC, and HCP enrichment, combined with tandem mass spectrometry (MS/MS), for identifying HCP. Zhu-Shimoni et al reported a scenario whereby low Chinese hamster ovary protein ratios were reported for the final pools of a therapeutic monoclonal antibody drug “mAb1” after Chinese hamster ovary protein ELISA. But, an additional peak was detected upon analysis with capillary electrophoresis using sodium dodecyl sulfate, which was later identified with sodium dodecyl sulfate polyacryl amide gel electrophoresis and matrix assisted laser desorption ionization-peptide mass fingerprint. Capillary electrophoresis typically uses background electrolytes, such as epsilon-aminocaproic acid (EACA), to enhance electrophoretic mobility, which makes online MS impossible. However, Redman et al developed an integrated microfluidic capillary electrophoresis-electrospray-ionization (ESI) approach with online MS analysis for intact antibodies using top-down MS. For validation, they used a commercially available antibody infliximab and IgG1 and IgG2. Using both MS and ELISA is a better strategy that has been used to study the interactions of HCP with monoclonal antibod-
<table>
<thead>
<tr>
<th>Sample</th>
<th>Technique</th>
<th>Information/application</th>
<th>Quality</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixture of intact species of the therapeutic protein and contaminants, such as HCPs</td>
<td>Size exclusion chromatography; gel filtration</td>
<td>Purity, presence of Therapeutic proteins (TP) species of different sizes, Aggregates, Other proteins (contaminants)</td>
<td>+ Fast, + Detection of aggregates, − Not suited for quantification, − Poor resolution</td>
<td>55</td>
</tr>
<tr>
<td>Affinity chromatography with biospecific affinity ligands (protein A, protein G, protein L)</td>
<td>Fast enrichment of antibodies</td>
<td>+ High specificity, + High degree of purity, − Coelution of some HCPs also possible</td>
<td>7,56</td>
<td></td>
</tr>
<tr>
<td>Affinity chromatography with PTM specific binders</td>
<td>Sample preparation step for investigation of PTMs</td>
<td>+ High specificity for defined PTMs</td>
<td>57,58</td>
<td></td>
</tr>
<tr>
<td>Sodium dodecyl sulfate poly-acryl amide gel electrophoresis</td>
<td>Purity, resolves species and contaminants with different size</td>
<td>+ Detection of abundant impurities which differ in size, + Fast, + Sensitive, + Specific, − Only known proteins can be detected, − Development of time consuming and expensive assays, − Inability to detect low/nonimmunogenic proteins, − Quantification difficult</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Enzyme-linked immunosorbent assay; Western blots</td>
<td>Detection of defined proteins</td>
<td>+ Fast, + Sensitive, + Specific</td>
<td>24,25</td>
<td></td>
</tr>
<tr>
<td>Two-dimensional gel electrophoresis</td>
<td>Purity, presence of species of different size and charge</td>
<td>+ High resolution, − Laborious, time consuming, − Requires skilled staff, − Hydrophobic proteins often are not displayed</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Sodium dodecyl sulfate capillary gel electrophoresis/MS (mass spectrometry)</td>
<td>Purity, presence of TP species with different sizes, Other proteins (contaminants)</td>
<td>+ Detection of abundant and less abundant impurities, + High resolution, − Not easy to operate</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>IEF (isoelectric focusing gel electrophoresis)</td>
<td>Purity, resolves species and contaminants with different charge</td>
<td>+ Fast, + Low resolution, − Not suited for quantification</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Purified product</td>
<td>Molecular weight of intact species, partial sequences, protein complexes/assemblies, PTM, disulfide bonds</td>
<td>+ Identification of different protein isomers (different PTM), + Fast, + Complexes and protein assemblies can be detected, + Near native-state analysis, + Disulfide bond identification, − High resolution needed (Fourier-transform-ion-cyclotron [FTICR], orbitraps), − Small changes in mass (i.e., deamidation) nearly unresolvable (for high mass proteins), − Because of complex spectra, high purity is required, − High amount of sample</td>
<td>62,63</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Sample</th>
<th>Technique</th>
<th>Information/application</th>
<th>Quality</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC-MS/MS by data independent acquisition, for example, sequential window acquisition of all theoretical mass spectra (SWATH)</td>
<td>Presence and relative quantities of proteins (gene products)</td>
<td>+ Multiplex quantification + Unbiased MS/MS acquisition + No prior assay development − Need for spectral libraries − Sample preparation takes time (trypsin digestion) − Only peptides present in protein libraries will be recognized − Modified peptides might be missed − Insufficient information about quantities of protein species</td>
<td>64–66</td>
<td></td>
</tr>
<tr>
<td>Bottom-up analysis of proteins by LC-MS/MS; before and after removal of glycans</td>
<td>Presence of proteins (but not protein species), common PTMs (ie, deamidation, oxidation)</td>
<td>+ Fast (if only one protease will be used and no enrichment is performed) + High sequence coverage will be achieved, several different proteases will be used + Well-established workflow + Site location of the glycans (in case of glycopeptides) − Coeluting glycopeptides may be missed − Insufficient information about quantities of protein species</td>
<td>67,68</td>
<td></td>
</tr>
<tr>
<td>Top-down analysis of intact proteins by LC-MS/MS</td>
<td>Relative quantities of species, composition of species</td>
<td>+ Allows relative quantification − Challenging − Still difficult to establish</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>Native MS</td>
<td>Conformation, presence of adducts and complexes</td>
<td>+ Important information about conformation and complexes − Laborious − Requires special mass spectrometers tuned for native MS in case of large protein complexes</td>
<td>17,70</td>
<td></td>
</tr>
<tr>
<td>Ion mobility MS</td>
<td>Secondary and tertiary structure</td>
<td>+ High speed (millisecond timescale) + Sensitive + Lower limit of detection</td>
<td>41,42,71</td>
<td></td>
</tr>
<tr>
<td>Capillary IEF</td>
<td>Purity, species with different charge variants</td>
<td>+ High resolution + Minimal development time + Reduced sample volume + Fast run times</td>
<td>51–53,72</td>
<td></td>
</tr>
<tr>
<td>ERLIC with tandem IMAC/TiO₂ enrichment and LC-MS/MS</td>
<td>Identification of phosphorylated sites</td>
<td>+ Allows relative quantification − More time required for enrichment</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Diverse sample preparation steps including, for example, release of glycans by PNGase F, LC-MS/MS</td>
<td>Glycoprotein composition (N and O glycosylation)</td>
<td>+ Highly quantitative + Isomer-sensitive + Site-specific glycoprotein analysis</td>
<td>37,38,60</td>
<td></td>
</tr>
<tr>
<td>Bright-field differential dynamic microscopy</td>
<td>Detection of submicron particles in protein-rich solutions</td>
<td>+ Low sample volume + Lower detection limits</td>
<td>49</td>
<td></td>
</tr>
</tbody>
</table>
ies, which cause the problem of copurification of HCP on Protein A affinity chromatography.\textsuperscript{18} However, Levy et al\textsuperscript{18} reported to have identified HCP–monoclonal antibody impurities by cross-interaction chromatography followed by two-dimensional gel electrophoresis and MS (matrix assisted laser desorption ionization tandem time-of-flight mass analyzer [TOF/TOF]).

Multiple protein analytes in the same sample can be rapidly analyzed and identified in a high-throughput fashion with MS, also giving light on what is present or not present in the samples, although absolute quantification is still a challenge. Good knowledge of the therapeutic protein expression system, upstream conditions implemented, and target protein itself gives a better insight on which HCPs are to be expected and therefore widens the ideas on which detection methods are preferable and at which point to employ these methods in the protein production process.\textsuperscript{33}

**Product-related impurities**

Therapeutic proteins designed for clinical use are characterized thoroughly to be able to detect molecular variants which could either be of genetic origin or emerge at the protein level\textsuperscript{34} possibly during the manufacturing process and/or storage of the drug. From post-translational modifications (PTMs) to truncated forms to aggregates, which occur during the formulation process or even during administration, protein therapeutics pose problems which need to be carefully monitored.

PTMs, such as phosphorylation, can be detected via electrostatic repulsion hydrophilic interaction chromatography with tandem immobilized metal affinity chromatography/titanium dioxide (IMAC/TiO\textsubscript{2}) enrichment and identification of the subsequent phosphopeptides by LC-MS/MS.\textsuperscript{35} Fc glycosylation of Fc fusion proteins can be analyzed by cleaving with IdeS protease and the resulting fragments analyzed by LC-MS.\textsuperscript{36} Protein glycans can generally be identified by chromatographic staining methods (eg, sodium dodecyl sulfate poly-acryl amide gel electrophoresis stained with fluorescent stains) or affinity-based methods (eg, lectin blot).\textsuperscript{37} MS-based approaches, such as nano-LC-MS/MS, have been used to analyze glycoproteins with respect to site-specificity.\textsuperscript{38} Most PTMs (acetylation, methylation, and ubiquitination) can be detected by immunoaffinity techniques using motif antibodies to enrich for the specific PTM followed by LC-MS/MS analysis.\textsuperscript{39} With high-resolution mass spectrometers becoming more affordable (such as orbitraps), a more comprehensive analysis of the microheterogeneity of glycosylation and PTMs of intact protein analysis using top-
down or middle-down MS became possible. One limitation of top-down analysis by MS is the need for high purity of the analyte because otherwise the highly complex spectra cannot be interpreted properly.\textsuperscript{16}

The biophysical characterization of drugs, such as ADCs (a potent cytotoxic agent which is covalently linked via a linker to an antibody),\textsuperscript{40} which are either cysteine or lysine conjugates is still a challenge. Accurate information on the conjugation profiles and the drug-loading distribution of these highly complex drugs is important for ADC engineering as these affect the pharmacokinetics, toxicity, and clearance of the drug. Native MS is a powerful approach for studying the lysine conjugates.\textsuperscript{41} Native MS in combination with ion mobility-MS was used to directly determine the drug to antibody ratio and drug loading profiles of trastuzumab–emtansine.\textsuperscript{42} Debaene et al\textsuperscript{43} also illustrated the use of native MS and ion mobility-MS to rapidly assess ADC structural heterogeneity and how they can be implemented into MS workflows for in-depth ADC analytical characterization.

Another point of interest is the introduction of other protein species due to aggregation, which can occur during all stages of the lifetime of a therapeutic protein (expression, refolding, purification, sterilization, shipping, storage, and delivery processes). Changes in therapeutic protein formulations, such as temperature, pH, and salt content, can cause aggregation or precipitation and thus detection of aggregates in protein drug products, especially on the subvisible size range, is important. It is greatly faced during formulation as concentration is increased to decrease administration volume as these aggregates can highly jeopardize patient safety. Many biophysical techniques are available for doing so, but each method has a series of shortcomings leading to inconsistency of results across platforms. Methods include the traditional size exclusion chromatography dynamic light scattering, differential scanning calorimetry, field-flow fractionation, atomic force microscopy, resonant mass measurement, sedimentation velocity analytical ultracentrifugation, Coulter counting, microflow imaging, and nanoparticle tracking analysis.\textsuperscript{43–48} Size exclusion chromatography is most often the method of choice as it is relatively fast and cheap. Recently, methods such as bright-field differential dynamic microscopy have also been developed and used to quantify the dynamics of submicron particles in protein-rich liquid clusters.\textsuperscript{49}

**Biosimilars and follow-on biologics**

The meaning of a biosimilar varies by jurisdiction but often refers to a biologic product that is comparable (European Union) or highly similar (the USA) to a previously approved biologic.\textsuperscript{50} Due to the fact that the manufacturing process for biologics always changes, the concept of biosimilarity needs to be demonstrated by extensive analytical methods before preclinical and clinical data are used.\textsuperscript{50} Verifying charge variants by determining the pI value is a great way for product identification, stability monitoring, and as a purity assay for quality control release. The biopharmaceutical industry generally relies on methods such as ion-exchange chromatography, isoelectric-focusing gel electrophoresis, and capillary equivalents such as capillary isoelectric focusing and imaged capillary isoelectric focusing to characterize charge variants. Imaged capillary isoelectric focusing is more sensitive and reliable as it takes into account not only the surface-exposed but also the intrinsic net charges.\textsuperscript{51–53} Recently, Stoll et al\textsuperscript{54} were able to characterize isoforms and variants of rituximab using selective comprehensive two-dimensional separation by liquid chromatography (2D-LC) combined with online MS analysis for the intact monoclonal antibodies and its partially digested and reduced forms using a middle-up approach. Here they used ion-exchange chromatography (in this case a cation-exchange chromatography [CEX] column) in the first dimension, which is the gold standard for separation of charged forms but has a low resolution compared to other separation techniques, and coupled it in the second dimension to a reversed-phase chromatography (RPC). They identified three major species and 19 minor species of the intact rituximab. For the partially digested forms, they were able to identify six major and 14 minor species and for the partially digested and reduced forms five major and 16 minor species. Some of these differences include different glycoforms as well as C-terminal lysine.

**Outlook**

Meanwhile, many tools have been established for analyzing the exact chemical composition in depth. However, until now, it is still very challenging to differentiate variant species, which differ only in one or a few moieties, from the therapeutic target protein, because the differences in the chemical properties are very small. Thus, the separation of these species from the therapeutic target protein is very difficult. With modern high-resolution mass spectrometers, these species can often be detected, even if their separation is not possible. Nevertheless, this analytical approach requires top-down mass spectrometric methods by which the species are infused in an intact form into the mass spectrometer. This requirement still can be troublesome depending on the nature of the therapeutic protein. Especially, large proteins, such as therapeutic antibodies, are troublesome regarding.
the top-down mass spectrometric analysis. Thus, we need new methods for making top-down MS more easy, reliable, faster, and automatizable.

Disclosure
The authors report no conflicts of interest in this work.

References


