Top-down characterization of biopharmaceuticals

Marcus Macht

Applied mass spectrometric techniques can fundamentally be divided into those starting from intact proteins (top down) and those starting from peptides derived by chemical or, more commonly, enzymatic digestion (bottom up). This article primarily covers top-down analysis and the information that it can obtain. It therefore covers electrospray ionization and matrix-assisted laser desorption/ionization techniques and top-down fragmentation techniques.

Keywords: Biopharmaceutical; Biosimilar; Deamidation; Fragmentation; Glycoprotein; Glycosylation; Ionization; Mass spectrometry (MS); Protein characterization; Top-down analysis

Abbreviations: CID, Collision-induced dissociation; ESI, Electrospray ionization; FT-ICR, Fourier transform ion cyclotron resonance; ICH, International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals in Human Use; LC, Liquid chromatography; MAb, Monoclonal antibody; MALDI, Matrix-assisted laser desorption/ionization; MS, Mass spectrometry; PTM, Post-translational modification; TOF, Time-of-flight

Marcus Macht*
Bruker Daltonik GmbH,
Fahrenheitstrasse 4, 28359
Bremen, Germany

*Tel.: +49 421 2205416;
Fax: +49 421 2205104.;
E-mail: mma@bdal.de,

1. Introduction

The pharmaceutical industry currently faces a trend towards the use of proteins as active pharmaceutical ingredients. Proteins typically have molecular weight significantly higher than typical active pharmaceutical ingredients and their molecular structure is much more flexible and variable than the usually relatively rigid structure of small molecules. As slight changes in the structure of these molecules can induces significant changes in their physical as well as pharmaceutical properties, the structure and any modifications occurring have to be determined thoroughly before their use in humans. Mass spectrometry (MS) has over the years emerged as one of the primary techniques for studying the molecular weight, the homogeneity or the heterogeneity, to verify amino-acid changes due to mutations or proteolytic processing during production, and to monitor any natural or artificial post-translational modifications (PTMs) [e.g., glycations, N- or O-glycosylation, lipidation or artificial loading with active groups (e.g., toxins or chelating reagents)].

When proteins are used for diagnostic or therapeutic purposes, their structure has to be thoroughly characterized. Apart from the amino-acid sequence, all modifications, sequence alterations or changes taking place upon production, changes in the production process or upon storage have to be characterized, as described in industry guidelines ICH Q5E and Q6B [1]. One of the most widely-used techniques for this characterization is MS, due to its sensitivity and flexibility. Nowadays, the most widespread technology for the analysis is electrospray ionization MS (ESI-MS) [2], but, recently, matrix assisted laser desorption/ionization MS (MALDI-MS) [3] was also applied for specific types of analysis. The dominant type of mass analyzers is time-of-flight (TOF) because of sensitivity, resolution and mass accuracy, particularly due to its fundamentally unlimited mass range, which makes it especially suited to the analysis of protein complexes or aggregates or the analysis of proteins under native conditions [4]. The second type of mass analyzer, often used in pharmaceutical protein characterization, is FT-based, either FT-ICR or Orbitrap.

Currently, there are two main trends in MS protein characterization: bottom-up and top-down analysis. While the first starts typically with an enzymatic, or alternatively chemical, digestion of the

0165-9936/$ - see front matter © 2013 Elsevier Ltd. All rights reserved. doi:http://dx.doi.org/10.1016/j.trac.2013.01.016

intact protein into smaller peptides followed by separation and analysis of these peptides, the latter performs MS analysis directly on the level of the intact protein [5–7]. Bottom-up analysis currently is the technology of choice for rapid protein identification and quantification of large numbers of proteins because of the relatively easy handling of peptides. Peptide separation can easily be achieved on a chromatographic level, providing access to huge numbers of proteins from any given organism.

However, there are intrinsic limitations in this technique which arise from the so-called ''protein inference problem'' [8,9]. The MS analysis identifies the peptides, not the proteins. The proteins from which these peptides originate are assigned solely based on statistics. This approach also means that mutations in or modifications on the protein sequence can be assigned only if they are explicitly observed. No statement is possible on non-observed deviations from the expected sequence. In addition, due to the upfront digestion, the context between the occurrence of a modification or a sequence mutation and a specific protein isoform is lost (e.g., if a 50% phosphorylation on a given sequence position is observed, no statement can be made whether this position is always occupied by 50% or whether there are two equally abundant isoforms, one unphosphorylated, one completely phosphorylated).

In contrast to bottom-up analysis, the top-down approach starts from the intact protein sequence, typically in the first step by determination of the intact molecular weight. This commonly reveals some initial information {e.g., protein homogeneity [are different isoforms/modifications present or not], the comparison between the expected theoretical protein mass [which is, in the case of recombinant biopharmaceuticals, usually known] and the observed mass [which allows to conclude about truncations, missing removal of signal sequences or purification tags or the distribution of heterogeneous modifications (e.g., glycosylations)]}. However, the limitations of top-down analysis are that, currently, only a few MS techniques are available to start the fragmentation in a comprehensive way on the intact protein sequence [10–12] and the separation of protein mixtures on the protein level is much more difficult and shows lower resolution than on the peptide level. In addition, bottom-up analysis is typically more sensitive than top-down analysis, which still requires a couple of pmol for a comprehensive analysis. However, particularly for the analysis of biopharmaceutical compounds, this technology has its advantages. The complexity of the samples is usually low so that the protein separation is not a major issue. Usually, there is no sensitivity problem (at least, not for the major compounds). In addition, the analysis of biopharmaceuticals requires a comprehensive analysis of the protein sequence, which is an intrinsic feature of the top-down analysis, whereas, in bottom-up analysis, this usually requires a combination of several complementary digests to cover the full sequence, and, even then, it might still be possible for no complete sequence coverage to be achieved.

In this review, we discuss the current state of top-down analysis for the analysis of biopharmaceuticals. The typical applications are intact mass confirmation, sequence confirmation, modification-heterogeneity analysis and identification of unknown or unexpected modifications or mutations. Some of these changes (e.g., the formation of isoaspartic acid after asparagine deamidation) can only be revealed by the techniques typically applied for top-down analysis.

2. Intact protein-based analysis

2.1. Intact molecular-weight determination

Typically, the first step in any type of top-down analysis is the determination of the intact molecular weight of the analyte of interest. This might be a protein, a glycoprotein, a protein complex or modified proteins in the case of protein conjugates or degradation analysis [13,14]. As the amino-acid sequence of biotherapeutics or biosimilars is well known, comparison of theoretical and experimental masses provides a direct indication for the homogeneity and the molecular integrity of the analyte. To do so, intact protein-mass determination by MALDI-TOF, ESI-TOF or ESI in conjunction with FT-based detection (either ICR or Orbitrap) techniques can be used [13–19]. For simple cases, MALDI-TOF is often the method of choice, which even allows us to measure very large compounds (e.g., IgMs) [20], but, when higher mass accuracy is required, ESI is usually used.

Currently, LC-ESI-TOF is by far the most commonly applied technology for the determination of the intact average mass of the proteins. The reason why what is typically regarded as less accurate average mass is used is that current instrumentation is resolution limited so that isotopically-resolved data cannot be obtained. Due to mass shifts, the average mass is also a sensible indicator (e.g., for deamidations causing a nominal 1 Da mass shift) whereas the same mass shift is usually hardly observable in a mixture of proteins with wider isotopic patterns. Also, the determination of the monoisotopic mass of a protein becomes tricky when the mass exceeds a certain limit, as the monoisotopic peak disappears in the spectral noise and the position has to be inferred from the overall isotopic pattern – a task that becomes increasingly difficult with increasing protein mass (and consequently wider isotopic distribution and resulting poorer ion statistics for each individual isotopomer) and complexity induced by deamidation or adduct formation. By using average masses, mass accuracy down to a few ppm can be achieved, corresponding to absolute errors of 0–2 Da for intact proteins, as shown in Fig. 1A for an