DE GRUYTER

Igor A. Kaltashov, Shunhai Wang, Guanbo Wang

MASS SPECTROMETRY IN BIOPHARMACEUTICAL ANALYSIS

Igor A. Kaltashov, Shunhai Wang, Guanbo Wang
Mass Spectrometry in Biopharmaceutical Analysis

Igor A. Kaltashov, Shunhai Wang,
Guanbo Wang

Mass Spectrometry in Biopharmaceutical Analysis

DE GRUYTER

Authors
Prof. Dr. Igor A. Kaltashov
Department of Chemistry
University of Massachusetts-Amherst
240 Thatcher Way
Amherst, MA 01003
USA
kaltashov@chem.umass.edu

Dr. Shunhai Wang
Analytical Chemistry
Regeneron Pharmaceuticals, Inc.
777 Old Saw Mill River Rd
Tarrytown, NY 10591
USA
shunhai.wang@regeneron.com

Dr. Guanbo Wang
Biomedical Pioneering Innovation Center
Peking University
5 Yiheyuan Rd., Haidian District
Beijing 100871
China
guanbo.wang@pku.edu.cn

ISBN 978-3-11-054496-1
e-ISBN (PDF) 978-3-11-054618-7
e-ISBN (EPUB) 978-3-11-054506-7

Library of Congress Control Number: 2021943128

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Cover image: A three-dimensional structure of a monoclonal antibody showing post-translational modifications that are routinely analyzed using mass spectrometry.
Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Preface

The advent of modern biotechnology nearly 50 years ago has led to transformation of many aspects of human life, but it is the healthcare sector that continues to reap seemingly endless benefits offered by the ability to manipulate the genetic information within living organisms, turning them into factories producing complex yet well-defined macromolecules with highly specific biological properties and functions. The first recombinant protein drug – insulin – was approved in 1982, merely 10 years after the production of the first recombinant DNA molecule. The total number of protein therapeutics introduced in the following years into clinical practice in the USA alone exceeds 150, with over a thousand currently at various stages of development and testing. Moving beyond the confines of single-molecule medicines, biotechnology now also offers opportunities to treat previously intractable pathologies using objects as complex as intact viral particles and indeed re-engineered cells. In the past several years, we have witnessed a triumphant entry of several gene therapies into the ranks of approved treatments for a range of devastating genetic disorders, and extensive efforts are already underway to develop the next generation of virus-based medicines, such as oncolytic viruses. Even more complex objects generated with the use of biotechnological tools are now actively pursued, such as genetically reprogrammed bacteria capable of blocking the immune checkpoints of drug-resistant tumors and initiating durable anti-tumor immunity.

The list of the highly effective medicines designed and produced with the use of biotechnological tools continues to expand at an ever-accelerating pace, and so are the expectations/requirements for the analytical characterization of the biopharmaceutical products. Even the simplest entities within this class of medicines are highly complex compared to their small-molecule counterparts (that are usually generated using the organic synthesis tools) and require sophisticated and highly specialized analytical tools for their characterization. The potential of mass spectrometry as a means of addressing these tasks has become apparent as early as 1980s, and within a decade this technique has become a default tool in the analytical armamentarium of the biopharmaceutical industry. While its utility in the analysis of biopharmaceutical products was initially limited to various aspects of covalent structure characterization, it is now commonly used for a range of other highly demanding tasks, including the higher order structure analysis, quantitation of biopharmaceutical products and detection/identification of impurities, just to name a few. Many of such applications have become routine, but the relentless pursuit of ever-more challenging targets in the biopharmaceutical field continues to raise the bar for the analytical work, constantly stimulating refinement of the existing techniques and development of the novel ones.

This book aims at presenting both the commonly used approaches to characterization/analysis of biopharmaceutical products that are based on mass spectrometry, and the new technologies that are developed to address the emerging challenges in this field. The material is presented in a systematic way, guiding the reader from the

https://doi.org/10.1515/9783110546187-202

common tasks and applications to the more specialized ones. Chapter 1 serves as a brief introduction by presenting the field of biopharmaceuticals and the role of the analytical work at various stages of their discovery, design, development, and testing. Chapter 2 provides background on modern biological mass spectrometry, focusing on those concepts and techniques that are referred to in the subsequent chapters of the book. Chapters 3 and 4 discuss various approaches to characterization of covalent and higher order structure of therapeutic proteins, respectively. Chapter 5 delves into the issues of biosimilarity and comparability, and how they can be addressed using mass spectrometry. Chapter 6 explores in great detail characterization of impurities in biopharmaceutical products, including both product- and process-related ones. Quantitation of protein therapeutics is the focus of Chapter 7. Chapter 8 provides an overview of mass spectrometry-based methods for characterization of biopharmaceutical products that do not fall into the category of protein therapeutics and their derivatives (with the emphasis placed on the oligonucleotide-based medicines and the gene therapy products). Lastly, the concluding Chapter 9 briefly discusses two emerging areas of applications of mass spectrometry in biopharma: the process analytical technology and the personalized medicine. While our intent was to provide comprehensive coverage of the entire field, it is physically impossible to consider every approach or application that had been reported in the literature. Furthermore, the preparation of this book lasted over 3 years, and the field of biopharmaceutical analysis has experienced truly dramatic changes over this time period, mirroring the transformative changes of the entire biopharma sector (including the first approvals of cell and gene therapies in the USA). We made every effort to capture these changes and the new trends, but some omissions are obviously inevitable. Lastly, a consistent effort has been made throughout the entire book to avoid endorsing specific vendors and instrumentation as much as possible, focusing instead on general principles of the presented techniques and highlighting specific technical requirements when needed.

This book was inspired to a very significant degree by multiple collaborations and discussions with colleagues in the biopharmaceutical sector. Our special thanks go to Pavel Bondarenko (Amgen), Damian Houde (Relay Therapeutics), Steven Berkowitz (Biogen), Phil Savickas (Merk Pharmaceuticals), Paul Salinas (Takeda), John Thomas (Alexion), Olga Frieze and Justin Perry (Pfizer). The completion of this book would not be possible without the help of the past and present colleagues and collaborators at the University of Massachusetts-Amherst. We are particularly indebted to Yang Yang and Cedric E. Bobst, Ian Carrick (currently at Purdue University), Rinat R. Abzalimov (CUNY Advanced Science Research Center), Jake W. Pawlowski (Amgen), Chendi Niu (AstraZeneca), and Khaja Muneeruddin (Broad Institute), who provided unpublished material to illustrate various concepts and applications presented in this book. Finally, we really want to have a feedback from you – our readers. Please contact any of the authors if you have questions or comments on any concepts presented or ideas/opinions expressed in this book. We would love to hear from you and to learn from your own experience – and for that we want to thank you in advance.

Contents

Chapter 1
Analytical chemistry in biopharma

This introductory chapter discusses the unique aspects of analytical characterization of biopharmaceutical products, and highlights the role of analytics in various aspects of their discovery, development, and testing. In addition, we point out the unique and increasingly important role played by mass spectrometry (MS) in this endeavor, which is the focal point of this book.

The terms "biologics" and "biopharmaceuticals" are frequently used interchangeably to designate a growing segment of medicines that are distinct from both natural products and synthetically produced therapeutics in that they are manufactured using the tools of biotechnology [1]. The Food and Drug Administration (FDA) broadly defines biological products as a wide range of substances such as vaccines, blood and blood components, allergenics, somatic cells, gene therapy, tissues, and recombinant therapeutic proteins. They can be composed of proteins, carbohydrates, or nucleic acids, as well as combinations of these substances, or indeed may be living entities such as cells. The term "biopharmaceuticals" (or "biopharmaceutics") is typically understood in a narrower sense, i.e., as products generated by living organisms whose genetic machinery has been altered not only to maximize the yield of the desired product, but indeed to make the organism produce foreign entities for which the wild type species are not programmed. This term came into use with the advent of modern biotechnology in 1980s, although "biopharmaceutics" was commonly used in the literature as early as 1960s, when it was defined as "the study of the relationship between . . . physical and chemical properties of the [small-molecule] drug and its dosage forms and the biological effects observed following [its] administration" [2], a field which is now known as the pharmacodynamics (PD) studies. The biotechnological route of production makes biopharmaceuticals distinct from the natural products (obtained from living entities that had not been genetically modified). This definition should be further clarified by specifying the relatively large molecular weight of biologics (upward of several kilodaltons, Figure 1.1), which typically places these entities outside of the reach of organic synthesis. Therefore, small-molecule medicines produced using the biotechnological means are not considered as biopharmaceuticals. However, it is important to note that a number of biopharmaceuticals that are currently on the market were initially produced as natural products, and a smaller subset can be also manufactured using the tools of organic synthesis [3]. Furthermore, the vast armamentarium of organic chemistry is frequently employed to either enhance the pharmacokinetic properties of biopharmaceuticals (e.g., via PEGylation) or indeed endow them with therapeutic potency (e.g., antibody–drug conjugation or haptenation of recombinant carrier proteins in vaccines), making the biotechnology/organic synthesis dichotomy somewhat blurry. To further complicate the matter,

https://doi.org/10.1515/9783110546187-001

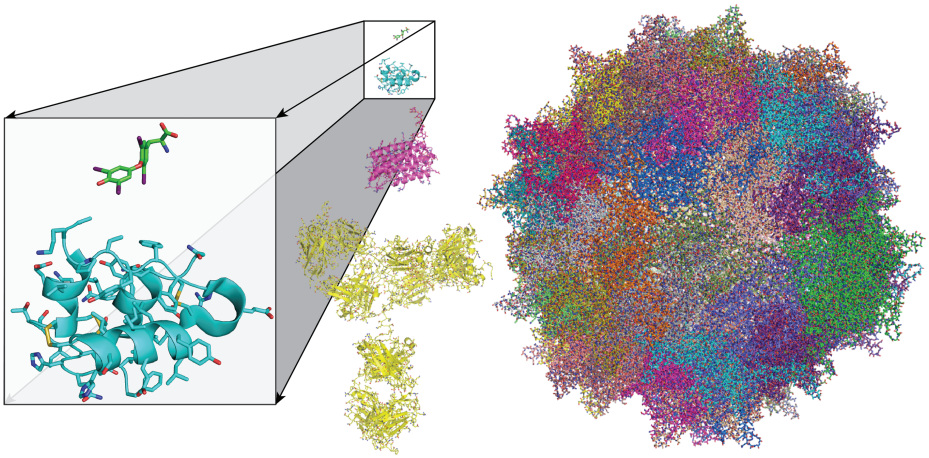


Figure 1.1: Size scale of biopharmaceuticals (insulin, cyan; interferon β1a, magenta; immunoglobulin γ, yellow; and adeno-associated virus, multi-color) in comparison with a small-molecule (a single amino acid) medicine thyroxine (green).

small nucleic acid–based medicines are usually considered biopharmaceuticals, despite being produced using the means of organic synthesis.

Proteins comprise the vast majority of biopharmaceuticals (hence, the frequent use of the terms "protein therapeutics" or "protein drugs," which until recently were considered to be synonymous with the "biopharmaceuticals"), although the share of oligonucleotide-based medicines continues to grow [4]. Furthermore, biotechnological routes of production are now actively evaluated for a range of carbohydrate-based medicines (e.g., heparin and heparin-derived products) that were traditionally manufactured as natural products. Modern biotechnology allows the biosynthetic machinery to be manipulated to enhance their therapeutic characteristics [5], making these polysaccharides chemically and biologically distinct from their cousins obtained from the animal sources; therefore, such entities should also be viewed as biopharmaceuticals. Lastly, the newest generation of biotechnology products used in medicine comprise complex and highly organized entities that incorporate multiple biopolymer units (e.g., a large number of proteins and nucleic acids making up viral particles in gene therapy applications [6]).

Currently, biopharmaceuticals constitute the fastest growing segment within the pharmaceutical market (Figure 1.2), with the number of new medicines in this class introduced each year since 2017 being consistently in the double digits. Since the FDA statistics reflected in Figure 1.2 pertains only to new medicines approved by the Center for Drug Evaluation and Research (CDER), it actually underestimates the total number of biologics license applications (BLAs) approvals, since a number of BLAs are approved by the Center for Biologics Evaluation and Research (CBER). Among the eight BLAs approved by CBER in 2020, one is a recombinant therapeutic

protein (Sevenfact™, a recombinant form of factor VIIa [7]), and one other is a cell therapy (Tecartus™, a CD19-directed genetically modified autologous T-cell immunotherapy [8]). Furthermore, one of the two vaccines approved by CBER in 2020 – MenQuadfi™ – is a so-called conjugate vaccine (a product of conjugation of polysaccharide antigens to a recombinant form of tetanus toxoid) [9], which should also be considered a biopharmaceutical product (Figure 1.3). Likewise, the 2019 approval numbers shown in Figure 1.2 do not include three CBER-approved BLAs for products designed and manufactured using the recombinant DNA technology (two vaccines – Ervebo [10] and Dengvaxia [11] – and a gene therapy product, Zolgensma [12, 13]).

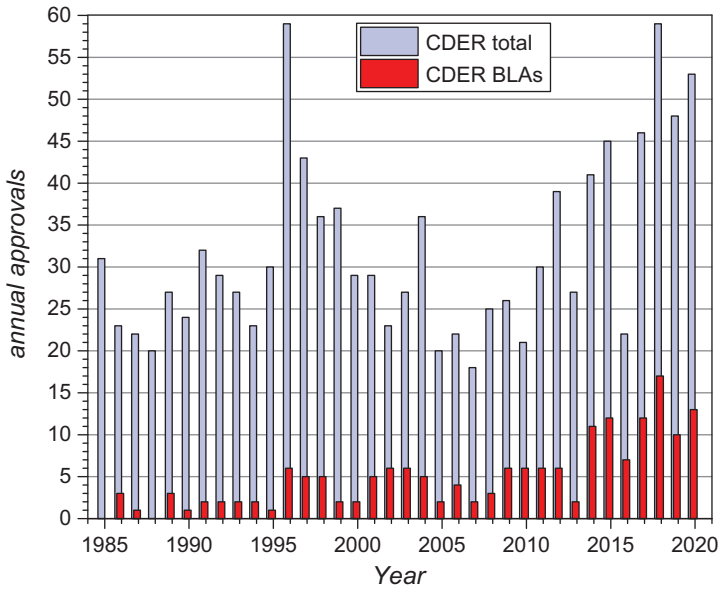


Figure 1.2: Total annual approvals issued by the FDA Center for Drug Evaluation and Research in the past 35 years and the fraction of biologics license applications (BLAs) among these approvals.

The clinical success rates of biologics are typically higher compared to the small-molecule drugs [14], a fact that is usually attributed to higher specificity of their interactions with therapeutic targets leading to higher efficacy and fewer side effects. Perhaps the best illustration for the superior effectiveness of the biopharmaceutical products and their ability to meet not only the existing but also emerging challenges in medicine is the fact that as of this writing all novel products that have received emergency approvals for the prevention and treatment of COVID-19 are biopharmaceuticals. While a few synthetically produced small-molecule medicines, such as ritonavir, and at least one natural product (heparin) have also been approved

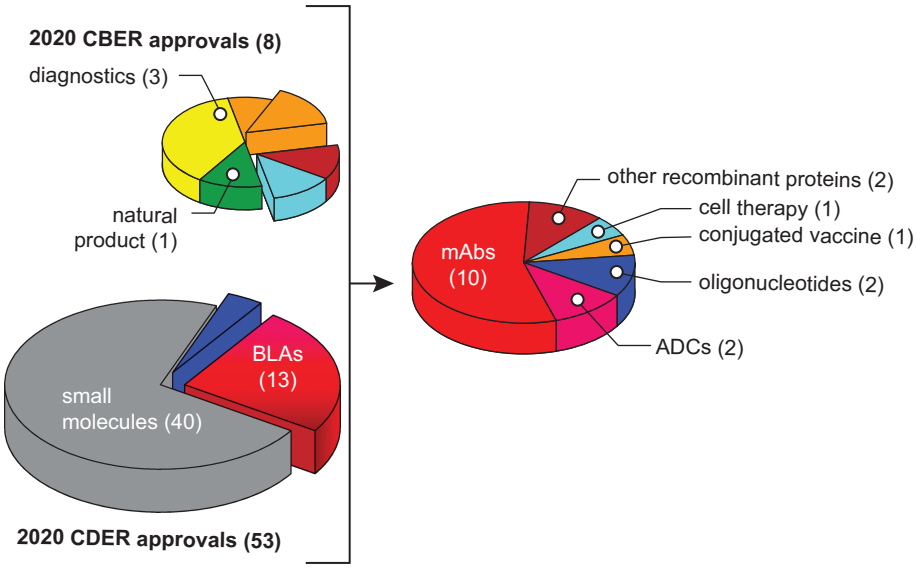


Figure 1.3: Left: the 2020 CDER and CBER approvals by modality; the sliced-out sectors represent medicinal products that are considered biopharmaceuticals (i.e., designed and manufactured using recombinant DNA technology, as well as synthetic oligonucleotides). Right: all biopharmaceutical products approved by the FDA in 2020 sorted by modality.

and are included in several treatment regiments of the novel coronavirus infection, these are repurposed medicines. In contrast, all SARS-CoV-2 targeting medicines that have truly become game-changers in combating the COVID-19 pandemic are novel products manufactured with the use of biotechnological tools, such as the Gamaleya Institute's Sputnik V [15] and the Pfizer/BioNtech BNT162b2 [16] vaccines, as well as a range of monoclonal antibodies targeting the novel coronavirus and effectively reducing the viral load [17].

While the diverse examples of biopharmaceutical products listed in the preceding paragraphs exhibit a wide range of chemical/biological properties, an important common feature of these entities is that all of them are biopolymers or combinations of biopolymers. An important consequence of this is the presence of multiple noncovalent intra- and inter-molecular bonds, which give rise to a higher order structure, or conformation. The latter is critical not only for the function (and, therefore, therapeutic activity) of the macromolecular medicines, but also for a variety of other interactions they may encounter post-administration, most notably with the host's immune system. While molecular conformation is an important characteristic of numerous small-molecule medicines as well, it only pertains to the formulation and solid dosage (where different polymorphs frequently display significant variation in their physicochemical properties, which affect the release/absorption rates [18]); once the drug is dissolved and absorbed, all memory of the solid-state conformation(s) is

lost. In contrast, the intricate three-dimensional architecture of a macromolecular drug must be maintained up until the point when its target is reached. This is needed not only to ensure the therapeutic effect, but also to eliminate the possibility of adverse reactions (such as the immune response). Therefore, the centrality of non-covalent intramolecular interactions vis-à-vis maintaining the proper conformation may be used as one of the defining criteria for biopharmaceutical products.

Another important distinction between the small-molecule medicines produced synthetically and the protein therapeutics manufactured using biotechnological tools is the significant structural heterogeneity frequently displayed by the latter. For example, the majority of therapeutic proteins are glycoproteins (i.e., they contain at least one glycan moiety that is enzymatically attached to the polypeptide chain after its translation has been completed); other post-translational modifications (PTMs) are also encountered frequently within a wide range of protein drugs. Although these enzymatic processes are tightly controlled in the course of the protein drug manufacturing, they nonetheless do not have the same fidelity as the translation off the genetic template. Therefore, instead of having a uniquely defined covalent structure, a protein drug is usually present as an ensemble of proteoforms, whose distribution in the final product must be controlled.

From the analytical point of view, the large size,the inherent (but nonetheless controlled) structural heterogeneity and the presence of the higher order structure (conformation) defined by the non-covalent interactions are the three main factors that set biopharmaceutical products apart from the classical small-molecule medicines and natural products. It is these three factors that create the framework for the analytical characterization of protein therapeutics and other biopharmaceutical products. The analyses are largely focused on the so-called critical quality attributes (CQAs), which are physical, chemical, biological, or microbiological properties or characteristics that must be within a certain limit, range, or distribution (Figure 1.4) in order for the specific product quality requirements to be met [19, 20]. MS is unique in that it can address a wide range of CQAs related not only to the biopharmaceutical product‘s potency and effectiveness (e.g., the amino acid sequence of a protein therapeutic and the integrity of its higher order structure) but also to its pharmacokinetic profile (e.g., oxidation of monoclonal antibodies affecting their affinity for FcRn receptors and, ultimately, their lifetime in circulation), safety (e.g., the presence of immunogenic PTMs and aggregation propensity) and purity (e.g., the presence of the host cell proteins in the final product).

A detailed discussion of CQAs and a related concept of the quality target product profile (QTPP) within the broader framework of the quality by design (QbD) approach to the development of biopharmaceutical products [19] is beyond the scope of this book. However, this section would be incomplete without mentioning the relevant analytical concepts, such as the analytical target profile (ATP) and analytical quality by design (AQbD), which were introduced by analogy to the QTPP and QbD concepts, but relate specifically to the analytical work. While

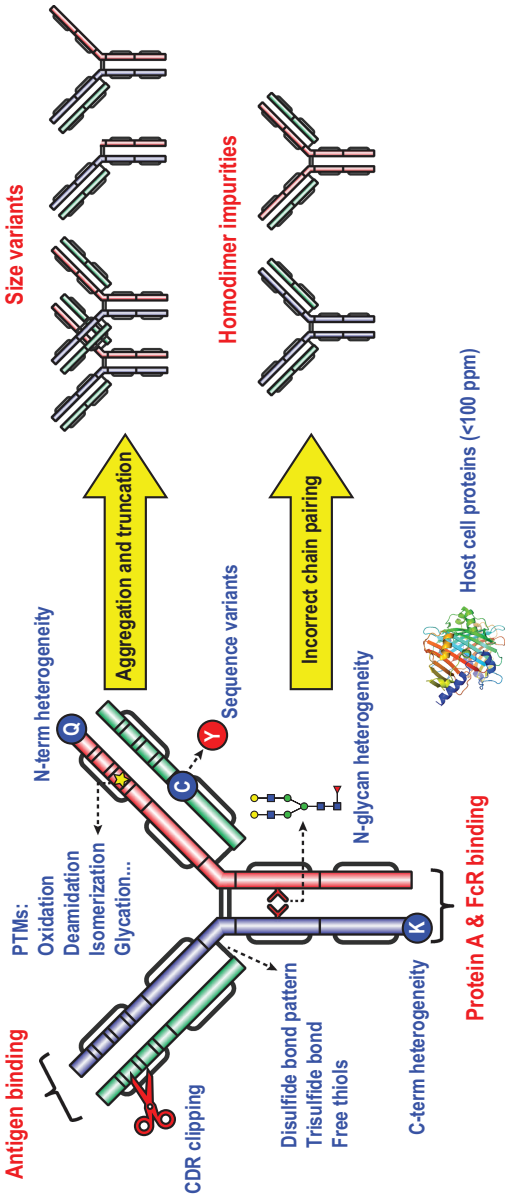


Figure 1.4: Quality attributes of biologics (using a bispecific monoclonal antibody as an example) that are typically monitored by mass spectrometry.

considered to be a subset of QbD, AQbD provides a mechanism to ensure that the analytical procedures are well understood, fit for purpose, robust, and consistently deliver the intended performance throughout their life cycle, and their application minimizes possibility of the analytical method failure [21]. In general, selection of a specific method of analysis (not necessarily limited to MS-based analyses) is made once CQAs are identified and ATP is specified. The choice of a specific method should be scientifically sound (e.g., take into account the required sensitivity and selectivity), and the method should be robust. Optimization of the method is frequently facilitated by evaluating the operating parameters within the multifactorial design space.The specific requirements that must be met by analytical methods evolve throughout the biopharmaceutical product's life cycle. For example, methods employed during Phase I studies are mainly used to characterize the product, support batch release, and facilitate stability testing; the emphasis is placed on identity, potency, and purity. As the product transitions to Phase II, the analytical procedure should be described, and the validation data should be available upon request. It is at this stage that the method qualification is expected (to confirm that it is fit for the purpose and the measurement results are reliable and reproducible). Full method validation is highly advisable (but not formally required) upon the product's transition to Phase III. According to the ICH-Q2 guidance, the validation characteristics include accuracy, reproducibility, specificity, limits of detections, and limits of quantitation. The BLA method submission should include not only the detailed description of the procedure (with sufficient detail to enable its reproduction), but also information on standards and controls, system suitability (to ensure that the entire system – including both the equipment and the analytical operations – will function correctly) and calculations (including both representative calculations and justification for correlation factors). The method is subject to re-validation at any time it is transferred (e.g., from the R&D department to QC, or to a different corporate entity). Although validation is not required for the so-called compendial methods (i.e., those listed in the Pharmacopeia), the methods are considered compendial only when they are applied to compendial products. An interested reader is referred to ICH Q2(R1) for more detailed information; the new guidance is currently being developed (Q14, specifically targeting the analytical method development), but it has not been released as of yet. In addition to the final product characterization, in-process controls (IPCs) are frequently implemented during the production process, but these are not subject to regulation/monitoring by the regulatory agencies.

Once developed, the analytical method must be closely monitored vis-à-vis its performance throughout the life cycle. Performance of various assays is matrix-dependent, and it can be affected by, for example, change in the raw material sources or consumables (such as filters). Other sources of variation are changes of the reagents and standards and transitioning to new analytical instrumentation. Personnel changes (for analyses carried out on-site) or a business decision to switch to a new CRO

(for outsourced analytical work) may also be a source of variation. Method validation provides representative snapshots of the analytical method performance over time and a quantitative measure of variability. The latter can be used to determine whether the variability is acceptable to avoid situations when the assay results are within the specifications range, but the assay itself fails to meet the validity criteria.

In addition to playing a critical role in addressing regulatory issues, MS is also indispensable in many other aspects of the biopharmaceutical products' development, including the discovery stage. The tasks supported by MS can range from identification of therapeutic targets at the early discovery stage via implementing large-scale proteomic and/or metabolomics screens to localizing epitopes within the already identified therapeutic targets and optimizing paratopes within the drug candidates, to name a few. Although MS is a mature analytical technique, it constantly evolves and its capabilities continue to expand at an accelerating pace, enabling ever more sensitive and informative analyses that allow most intimate details of the biopolymer structure to be determined. It is therefore not surprising that MS not only continues to enjoy a steady growth in popularity in the biopharmaceutical arena, but in fact is the most important analytical tool in this field [22]. The purpose of this book is to familiarize the readers who are interested in/involved with the biopharmaceutical sector with the capabilities of modern MS that are relevant for both characterization of the finished products and the discovery/development of novel biologic medicines. No prior experience with MS is expected of the readers; in fact, the structure of the material presentation allows this book to serve as an effective introduction to this field for analytical scientists and biochemists who have little familiarity with the MS-based methods of analysis of biopharmaceutical products. At the same time, the expansive range of topics and the depth of their coverage should also make this book a valuable resource even for the most seasoned MS veterans, as they will be able to find a wealth of valuable information outside of their area of expertise.

References

[1] Knäblein, J. (ed.) Modern Biopharmaceuticals: Design, Development and Optimization, Vols 1–4. (WILEY-VCH Verlag GmbH & Co., Weinheim; 2005).
[2] Wagner, J.G. Biopharmaceutics: absorption aspects. J. Pharm. Sci. 50, 359–387 (1961).
[3] Belgi, A., Hossain, M.A., Tregear, G.W. & Wade, J.D. The chemical synthesis of insulin: from the past to the present. Immunol. Endocr. Metab. Agents Med. Chem. 11, 40–47 (2011).
[4] Walsh, G. Biopharmaceutical benchmarks 2018. Nat. Biotechnol. 36, 1136 (2018).
[5] Fu, L., Suflita, M. & Linhardt, R.J. Bioengineered heparins and heparan sulfates. Adv. Drug Deliv. Rev. 97, 237–249 (2016).
[6] Wirth, T., Parker, N. & Yla-Herttuala, S. History of gene therapy. Gene 525, 162–169 (2013).
[7] Meeks, S.L. & Leissinger, C.A. The evolution of factor VIIa in the treatment of bleeding in haemophilia with inhibitors. Haemophilia: Off. J. World Fed. Hemophilia 25, 911–918 (2019).

[8] Mian, A. & Hill, B.T. Brexucabtagene autoleucel for the treatment of relapsed/refractory mantle cell lymphoma. Expert Opin. Biol. Ther. 1–7 (2021). doi: 10.1080/14712598.2021.1889510.
[9] In Brief: New meningococcal serogroup B vaccination recommendations. Med. Lett. Drugs Ther. 62, 191–192 (2020).
[10] Wolf, J. et al. Development of pandemic vaccines: ERVEBO case study. Vaccines 9 (2021).
[11] Thomas, S.J. & Yoon, I.K. A review of Dengvaxia®: development to deployment. Hum. Vaccin. Immunother. 15, 2295–2314 (2019).
[12] Al-Zaidy, S.A. & Mendell, J.R. From clinical trials to clinical practice: practical considerations for gene replacement therapy in SMA Type 1. Pediatr. Neurol. 100, 3–11 (2019).
[13] Stevens, D., Claborn, M.K., Gildon, B.L., Kessler, T.L. & Walker, C. Onasemnogene abeparvovec-xioi: gene therapy for spinal muscular atrophy. Ann. Pharmacother 54, 1001–1009 (2020).
[14] Smietana, K., Siatkowski, M. & Møller, M. Trends in clinical success rates. Nat. Rev. Drug Discov. 15, 379–380 (2016).
[15] Logunov, D.Y. et al. Safety and immunogenicity of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine in two formulations: two open, non-randomised phase 1/2 studies from Russia. Lancet 396, 887–897 (2020).
[16] Polack, F.P. et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. N. Engl. J. Med. 383, 2603–2615 (2020).
[17] Kaplon, H. & Reichert, J.M. Antibodies to watch in 2021. mAbs 13, 1860476 (2021).
[18] Zhou, Y., Wang, J., Xiao, Y., Wang, T. & Huang, X. The effects of polymorphism on physicochemical properties and pharmacodynamics of solid drugs. Curr. Pharm. Des. 24, 2375–2382 (2018).
[19] Yu, L.X. et al. Understanding pharmaceutical quality by design. AAPS J. 16, 771–783 (2014).
[20] Eon-Duval, A., Broly, H. & Gleixner, R. Quality attributes of recombinant therapeutic proteins: an assessment of impact on safety and efficacy as part of a quality by design development approach. Biotechnol. Prog. 28, 608–622 (2012).
[21] Jackson, P. et al. Using the analytical target profile to drive the analytical method lifecycle. Anal. Chem. 91, 2577–2585 (2019).
[22] Rogstad, S. et al. A retrospective evaluation of the use of mass spectrometry in FDA biologics license applications. J. Am. Soc. Mass Spectrom. 28, 786–794 (2017).

Chapter 2
A brief overview of biological mass spectrometry: concepts and definitions

This chapter provides concise background material on modern mass spectrometry (MS) instrumentation and techniques that will be referred to in the subsequent chapters of this book. The chapter equips the reader with a knowledge base needed for making an informed choice of a particular instrument or technique for a specific task. The important concepts pertaining to biomolecular MS are briefly reviewed, with an emphasis on features that are most relevant to biopharmaceutical analysis.

2.1 Mass measurements: stable isotopes, isotopic distributions, ionic and molecular masses, and mass resolution

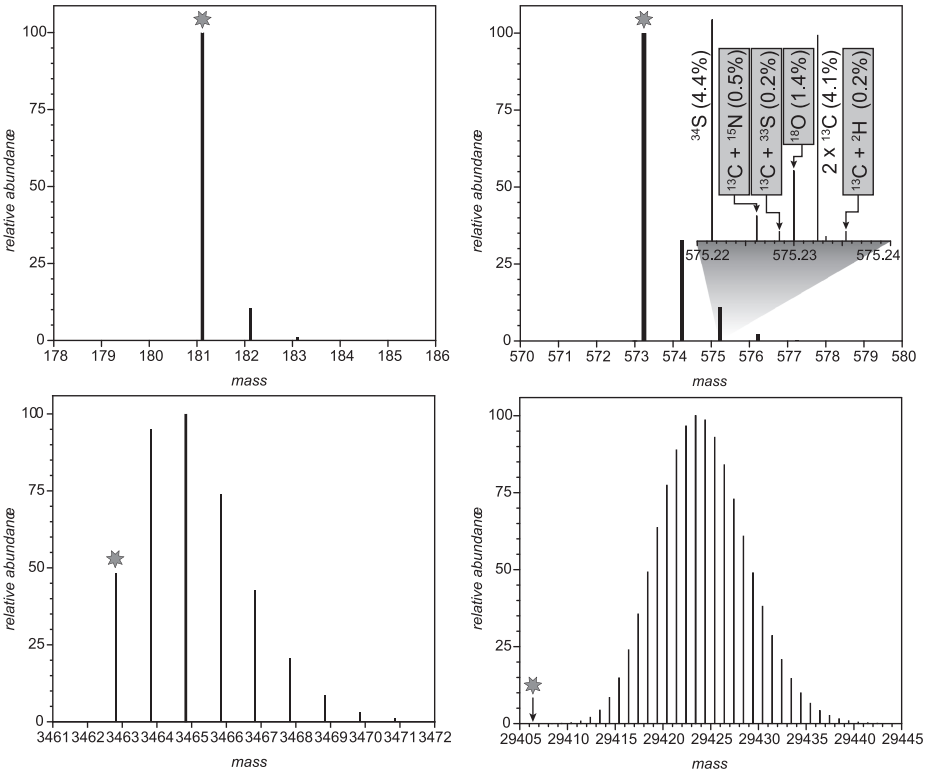
Mass spectrometers are analytical devices that transfer molecules from a gas or condensed phase to vacuum, ionize them and manipulate them using electromagnetic fields to determine abundance of ionic species as a function of their mass-to-charge ratios (m/z). While MS analyses can be performed across a wide range of atomic and molecular objects, we will focus our discussion on biological macromolecules and try to avoid generalizations or presenting the material that may not be directly relevant to the field of biopharmaceutical analysis. A variety of excellent books cover the general aspects of MS [1–4]; in addition, comprehensive resources are available describing applications of this technique in the fields of synthetic small-molecule medicines [5–7] and natural products [8].

There is a great variety of ways to perform MS measurements, all of them based on the fact that behavior of any ion in electromagnetic fields is uniquely determined by its m/z ratio. Therefore, even though it is the ionic mass (m) that is the primary target of most MS measurements, the actual measured characteristic is m/z. Since small molecules usually give rise to ions bearing a single charge, the difference between the ionic mass and the m/z ratio is largely semantic (they are numerically equal to each other). In contrast, macromolecules are frequently represented in MS by multiply charged ions, and the numeric value of the m/z ratio could be a small fraction of the ionic mass. We will consider the multiple charging phenomenon (and discuss various way to calculate the value of z based on the appearance of the ionic signal in MS) in Section 2.2. Since all charge carriers have a physical mass of their own, the ionic mass always differs from that of a neutral molecule even in the case of singly charged species. The extent of this deviation varies significantly across different charge carriers (e.g., the mass ratio of the two most common carriers – a proton and an electron – is $m_p/m_e = 1.83615 \times 10^3$).

https://doi.org/10.1515/9783110546187-002

Up to this point, our discussion was based on an implicit assumption that each ion (as well as the molecule it represents) can be assigned a unique mass. This assumption, however, ignores the fact that the majority of elements have more than one stable isotope, giving rise to a possibility of multiple molecules with identical elemental composition having unequal masses. Most of the elements that are commonly encountered in proteins and other biopolymers have at least two stable isotopes, but the relative abundance of the lightest isotopes of carbon, hydrogen, oxygen, and nitrogen dwarfs that of the heavier isotopes of the same element (e.g., $^{13}C/^{12}C$ natural abundance ratio is 0.011, while $^{15}N/^{14}N$ is less than 0.004). Nevertheless, the large number of atoms comprising a biopolymer molecule dramatically increases the probability of at least one heavier isotope being incorporated into the molecule (Figure 2.1). Even relatively short peptides contain multiple isotopologs (species having identical chemical structures, but differing in their isotopic compositions). The so-called isotopic distributions (abundance distribution of isotopologs plotted as a function of their masses) are usually presented as histograms showing a monoisotopic species M (containing only the lightest isotope of each element) and a series of isotopologs M + 1, M + 2, and so on. This presentation ignores the fact that each of the M + n species is comprised of several unique isotopologs, for example, in addition to a $^{13}C/^{12}C$ substitution, M + 1 species also contains contributions from a $^{15}N/^{14}N$ substitution; the number of unique isotopologs within the M + 2 species would be even more significant (see Figure 2.1). State-of-the-art MS instrumentation allows these isotopologs to be distinguished from each other in some cases, but for the most part in this book we will be using a common approach that does not distinguish isotopologs within each M + n species. An important consequence of this choice is that the monoisotopic species (M) is the only molecular entity that has a well-defined mass (e.g., for methionine enkephalin having a molecular formula $C_{27}H_{35}N_5O_7S$ it would be a sum of $27 \times m_{12C} + 35 \times m_{1H} + 5 \times m_{14N} + 7 \times m_{16O} + m_{34S} = 573.22575$ Da), which is usually referred to as the monoisotopic mass.

The monoisotopic mass values are sometimes rounded to the nearest integer (a value commonly referred to as the nominal mass), but this results in a loss of very important information. For example, the nominal mass of methionine enkephalin considered above (YGGFM) would be numerically equal to that of a range of other compounds with different compositions (e.g., a peptide KAREA would have the same nominal mass of 573, but obviously a different amino acid composition). Compounds having identical nominal masses but a different chemical compositions are called isobaric species or isobars (YGGFM is isobaric to KAREA). The ability to measure the monoisotopic mass with a high degree of precision/accuracy afforded by modern MS allows isobars to be distinguished from each other based on their accurate masses (which are sometimes referred to as exact masses). Such measurements typically require high mass resolution, a measure of the ability of a mass spectrometer to distinguish between ionic signals of two species with close (but not identical) masses. Mass



Y

YGGFM

YGGFMTSEKSQTPLVTLFKNAIIKNAYKKGE

MPRSCCSRSGALLLALLLQASMEVRGWCLESSQCQDLTTESNLLECIRACKPDLSAETPM
FPGNGDEQPLTENPRKYVMHFRWDRFGRRNSSSSGSSGAGQKREDVSAGEDCGPLPEGPE
PRSDGAKPGPREGKRSYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAEAFPLEFKRELTG
QRLREGDGPDGPADDGAGAQADLEHSLLVAAEKKDEGPYRMEHFRWGSPPKDKR**YGGFMT**
SEKSQTPLVTLFKNAIIKNAYKKGE

Figure 2.1: Isotopic distributions of biological molecules of different sizes (an amino acid tyrosine, a short peptide methionine enkephalin incorporating tyrosine, a polypeptide β-endorphin incorporating the methionine enkephalin sequence, and a protein corticotropin incorporating the β-endorphin sequence). In each case, the intensity of the most abundant isotopic peak is set at 100%, and the positions of the monoisotopic peaks are labeled with a star in each panel. The inset in the methionine enkephalin panel shows isotopologs comprising the isotopic peak M + 2. Complete sequences of all biomolecules are shown at the bottom of the figure.

spectra – the plots of ionic abundance as a function of m/z ratio – are continuous graphs (unless purposefully changed to a histogram format to reduce the load on the data processing/handling system), and each molecular species represented by a single

bar in histograms shown in Figure 2.1 gives rise to an ionic signal with a finite width. Mass resolution (R) is defined as

$$R = \frac{M}{\Delta M} \tag{2.1}$$

where M is the m/z value of a particular ion, and ΔM is the width of its signal intensity distribution on m/z scale (usually taken at 50% of the maximum intensity). Since both the numerator and the denominator in (2.1) contain z values, a mathematically equivalent definition of mass resolution would have an $m/\Delta m$ ratio; however, it is important to remember that the mass resolution is frequently a function of m/z (this will be examined in more detail later in this chapter). Accurate mass measurements for any ion cannot be carried out unless the resolution is sufficient to eliminate interferences from other ions with close m/z values. This is illustrated in Figure 2.2, where ionic signals of monoisotopic peaks of methionine enkephalin and two of its isobars are simulated at mass resolution ranging from 2,500 to 60,000. Even if the interferences are absent, determining the accurate ionic mass is a much more challenging task at lower resolution, since localization of a peak apex becomes increasingly difficult as the peak top becomes "flatter," especially in the presence of its inevitable modulation by noise.

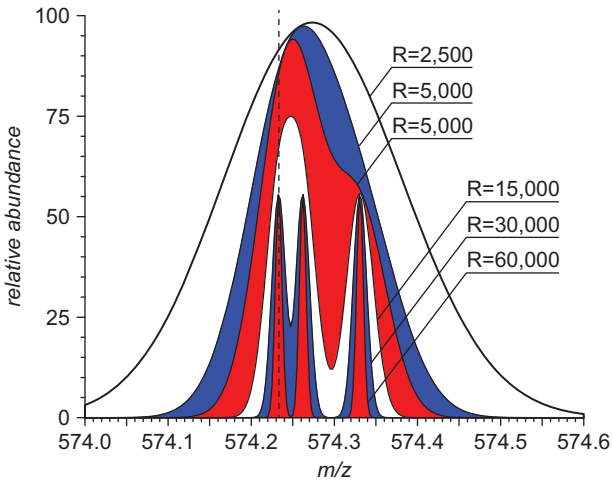


Figure 2.2: Simulated signals for iso-abundant ions representing protonated forms (MH^+) of three isobaric peptides (YGGFM, GWAPGS, and KAREA) at different levels of mass resolution as indicated on the graph. The dotted line indicates the exact mass of methionine enkephalin (574.2330); the exact masses of its two isobars are 574.26258 (GWAPGS) and 574.3307 (KAREA). Only the m/z region corresponding to the monoisotopic peaks is shown for clarity.

The results of accurate mass measurements are typically reported together with the measurement uncertainty; this allows a determination to be made vis-à-vis how many different mass values corresponding to unique compounds fall within the confidence interval. Should the confidence interval be sufficiently narrow to contain only one unique mass value, the accurate mass measurements can be used to establish a unique empirical formula. For example, measuring the mass of the protonated form of methionine enkephalin as 574.2330 with a 0.3 ppm precision would allow a unique empirical formula to be identified whose mass falls within the confidence interval ($C_{27}H_{35}N_6O_7S_1$, which is the molecular formula of the $[M + H]^+$ ion of methionine enkephalin). Expanding the confidence interval would of course result in a progressive increase of the number of putative compounds whose masses fall within the interval (e.g., 2 at 1 ppm and 15 at 10 ppm for the example considered above), although many of the candidate masses can be filtered out as mathematical artifacts, that is, corresponding to chemically unreasonable empirical formulae. It is important to remember that mass resolution is not the only factor defining the mass measurement precision. A significant systematic error can be introduced by poor calibration of the instrument. An interested reader is referred to a tutorial on this subject [9]; a detailed technical study comparing various types of MS instrumentation vis-à-vis their ability to provide accurate mass measurements is also available [10].

As the size of the biopolymer increases, the laws of combinatorics dictate that the relative abundance of the isotopolog comprised of the lightest isotopes of each element decreases precipitously (Figure 2.1). In fact, the abundance of the monoisotopic peak falls to extremely low values even for proteins of a relatively modest size (as illustrated in Figure 2.1 using a 29.5 kDa precursor of β-endorphin as an example). In practice, such low-abundance species evade MS detection, as their signal levels fall below that of noise. Therefore, the notion of the monoisotopic mass is not particularly useful for larger polypeptides and proteins, unless isotopic depletion strategies are employed to generate recombinant proteins that are artificially enriched with ^{12}C, ^{14}N, ^{16}O, and so on [11]. The ionic or molecular mass of larger biopolymers can be reported using the most abundant isotopolog in the signal distribution, which is sometimes referred to as the most abundant mass. The problem with this approach is that in many cases (particularly, for high-molecular-weight macromolecules) it may be difficult to select the most abundant isotopic peak, as the binomial distributions dictating the appearance of the isotopic clusters are wide, and even minor modulations by noise may change the abundance order of peaks at the apex of the distribution. Furthermore, as the macromolecular mass increases, the ability to observe distinct isotopic peaks requires progressively higher levels of mass resolution: as one can see from equation (2.1), regardless of the ionic charge the level of mass separation sufficient for resolving individual isotopic peaks is numerically equal to the ionic mass. Since the masses of the majority of modern protein therapeutics exceed 100 kDa, the isotope-level resolution can only be achieved when high-resolution state-of-the-art MS instrumentation is available (these features will be discussed in more detail later in this

chapter). Therefore, a better (and commonly used) approach to assigning the mass of large macromolecular species uses the notion of an average mass, which is simply an abundance-weighted average of masses of all detectable isotopic peaks in the signal distribution. The average mass can be readily obtained even if the ionic signal is not isotopically resolved, in which case it can be estimated with a reasonable degree of precision and accuracy using either a centroid or an apex of the continuum signal (Figure 2.3).

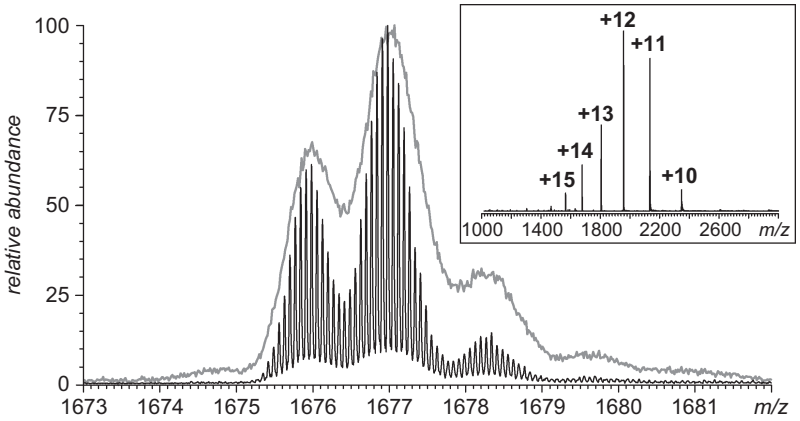


Figure 2.3: A zoomed view of the ionic signals of methylated trypsin at $z = +14$ obtained with a high-resolution mass spectrometer (FT ICR MS, black trace) and a medium-resolution instrument (TOF MS, gray trace).

2.2 Generation of macromolecular ions: matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI)

The advent of matrix-assisted laser desorption/ionization (MALDI) [12, 13] and electrospray ionization (ESI) [14, 15] in late 1980s paved the way toward rapid expansion of MS-based methods of structural analysis into the realm of biological macromolecules. There is a variety of comprehensive resources addressing various aspects of MALDI and ESI in great detail, including several excellent books [16, 17]. This section provides a brief overview of those features of both MALDI and ESI that are particularly relevant for the MS analysis of recombinant proteins and other biological macromolecules that are the primary subject matter of this book.

As the name suggests, MALDI belongs to a family of ionization techniques that produce molecular ions off solid surfaces by combining the desorption and ionization processes. This is typically achieved by embedding the analyte molecules

(e.g., proteins or other biopolymers) in micro-crystals made of the so-called matrix molecules simply by drying the concentrated matrix solutions that incorporate small amounts of the analyte. The most common MALDI matrices are organic acids that exhibit high extinction coefficients in the near-UV region, which allows them to absorb irradiation of common UV lasers (such as a nitrogen laser or an Nd:YAG laser) very effectively. Therefore, irradiation of the analyte molecules co-crystallized with the matrix results in highly selective channeling of the photon energy to the matrix molecules. Consequently, a short laser pulse results in a rapid and localized heating of the surface of the matrix crystal, followed by a near-adiabatic expansion of the material within the affected region of the crystal, leading to the desorption of both the matrix and the analyte molecules off the micro-crystal surface. The matrix molecules assisting this process not only serve as an energy buffer protecting the analyte molecules from the radiation damage, but also enable the analyte ionization by acting as proton donors (or proton acceptors in the negative ion mode). It is the matrix that makes MALDI a soft ionization technique by allowing the analyte fragmentation to be minimized or indeed eliminated (in contrast to the laser desorption/ionization – LDI – a method of ion production that preceded MALDI). At the same time, the use of photons as high-energy "projectiles" allow high-mass analytes to be readily desorbed by enabling highly selective energy channeling to the matrix molecules (in contrast to the fast atom bombardment – FAB – a desorption technique that uses high-velocity ions or neutral atoms as projectiles to desorb the analyte molecules embedded in non-volatile liquid matrices).

MALDI predominantly produces singly charged ions (Figure 2.4), usually via protonation (MH^+) and/or alkali metal ion adduct formation (e.g., MNa^+). For macromolecules exceeding several kilodaltons, lower abundance signals corresponding to multiply charged species (e.g., $[M + nH]^{n+}$ or $[M + nH + (z - n)Na]^{z+}$ can also be present, although the extent of multiple charging remains low compared to the ESI-generated ions of the same biomolecule (vide infra). Both the extent of multiple charging and the abundance of ionic species at $z > 1$ depend on the specific matrix being used [18, 19]. As shown in Figure 2.4, even proteins as small as insulin can exhibit doubly charged species in MALDI MS. Nevertheless, it is the singly charged species that are usually most abundant in MALDI mass spectra, and because of that, the ionic signals of larger biopolymers populate high m/z regions of mass spectra (the upper limit of the m/z region must exceed the numeric value of the biopolymer mass in order for the analyte signal to be detected), necessitating the use of mass analyzers capable of making measurements at extended m/z range (e.g., time-of-flight (TOF) MS, vide infra). Another feature frequently observed in MALDI mass spectra of larger biopolymers is extensive adduct formation with the matrix molecules, as well as other polar species present in solution. This phenomenon leads to asymmetric broadening of the macromolecular ion peak and limits the accuracy of mass measurements. Adduct formation can also involve two or more analyte molecules, giving rise

to non-covalent dimeric and trimeric analyte ions in MALDI mass spectra; these should not be confused with the physiologically relevant non-covalent assemblies observed in the so-called native MS experiments (which will be discussed in detail in Chapter 4).

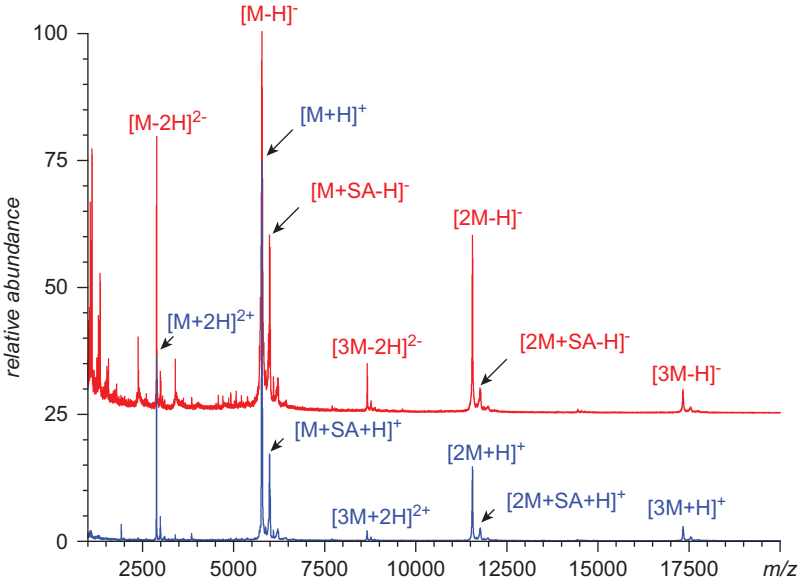


Figure 2.4: MALDI mass spectra of insulin acquired in the positive (blue trace) and negative (red) ion modes. "M" in peaks labels represents the molecular (neutral) species of insulin, "SA" represents the matrix (sinapinic acid) adducts, and "H" represents a proton. The abundance of non-specific dimers and trimers populating the mass spectra at $m/z > 6{,}000$ can be reduced by lowering the protein concentration in the sample. No protein ion peaks are isotopically resolved. Data courtesy of Lingxiao Chaihu (Nanjing Normal University)

Although the majority of MALDI MS measurements are carried out in the positive ion mode, generation of negative ions is also possible (usually via deprotonation or anion adduct formation, e.g., $[M\text{-}H]^-$ or $[M + Cl]^-$), which can be particularly appealing when working with acidic analytes. Furthermore, fragmentation pathways of many biopolymers depend on the polarity of the parent ion undergoing dissociation; therefore, a judicious choice of the ionic polarity may help achieve specific objectives in tandem MS (MS/MS) measurements, which will be discussed in more detail later in this chapter. Lasers used in all commercial MALDI mass spectrometers produce radiation in short pulses, which unfortunately makes this ionization method incompatible with several types of mass analyzers, including quadrupole MS and double-sector MS (this will be discussed in more detail later in this chapter). However, most other analyzers commonly used in MS are compatible with the pulsed mode of ion production and, therefore, can be readily interfaced with a MALDI ionization source.

ESI is another commonly used method of generating biomolecular ions. Unlike MALDI, it desorbs and ionizes the analyte molecules directly from a solution, rather than from a solid state. The analyte solution is typically supplied to the ESI source at relatively low flow rates (ranging from tens of nL/min to hundreds of μL/min) through a fine (usually metallic) capillary. A high electrostatic potential applied to the tip of the capillary creates peculiar instability conditions resulting in formation of a fine jet of charged liquid. Disintegration of this jet produces charged aerosol particles that become unstable following evaporation of a fraction of solvent sufficient to increase the droplet surface charge above the so-called Rayleigh limit (a condition when the electrostatic repulsion overcomes the cohesive action of the surface tension). Disintegration of the droplets upon their reaching the Rayleigh limit gives rise to smaller (progeny) charged droplets, which in turn start shrinking due to solvent evaporation, eventually becoming unstable as well. Multiple repeats of this process on progressively smaller scales eventually liberate the analyte molecules from the droplets, generating ions that may accommodate multiple charges depending on their physical size [20]. Although the multiple charging phenomenon is not unique to ESI (vide supra), its extent is significantly higher for the ESI-generated macromolecular ions compared to those produced by MALDI. This is illustrated in Figure 2.5, which shows ESI mass spectra of insulin displaying the most abundant ionic signal at $z = 4$, alongside prominent peaks at $z = 5$ and 3 (compare this with the appearance of a MALDI mass spectrum of insulin shown in Figure 2.4).

Multiple charging of biopolymers is beneficial in that it results in localization of the MS signal within a relatively narrow m/z range (500–4,000 under typical ESI conditions even for biopolymers exceeding 100 kDa).[1] This allows all types of mass analyzers to be interfaced with ESI sources (unlike MALDI, which can only be used with TOF MS when applied to biopolymers exceeding 10 kDa). On a flip side, the multiple charging phenomenon may complicate the data analysis, as the analyte ions' masses are no longer numerically equal to their respective m/z values, and the relationship between the latter and the mass of the analyte molecule in its electrically neutral (m_O) form is given by

$$\frac{m}{z} = \frac{m_O + z \cdot m_{H^+}}{z} \tag{2.2}$$

if the ions are produced by multiple protonation. Since ubiquitous alkali metal cations can also participate in multiple charging of the analyte molecules (see the bottom-left panel in Figure 2.5), a more universal formula would take into account the possibility of Na^+ and K^+ adduct formation:

1 The exception is the so-called native MS, where the use of non-denaturing solvents give rise to ions with relatively low number of charges at higher m/z values (this technique will be considered in Chapter 4).

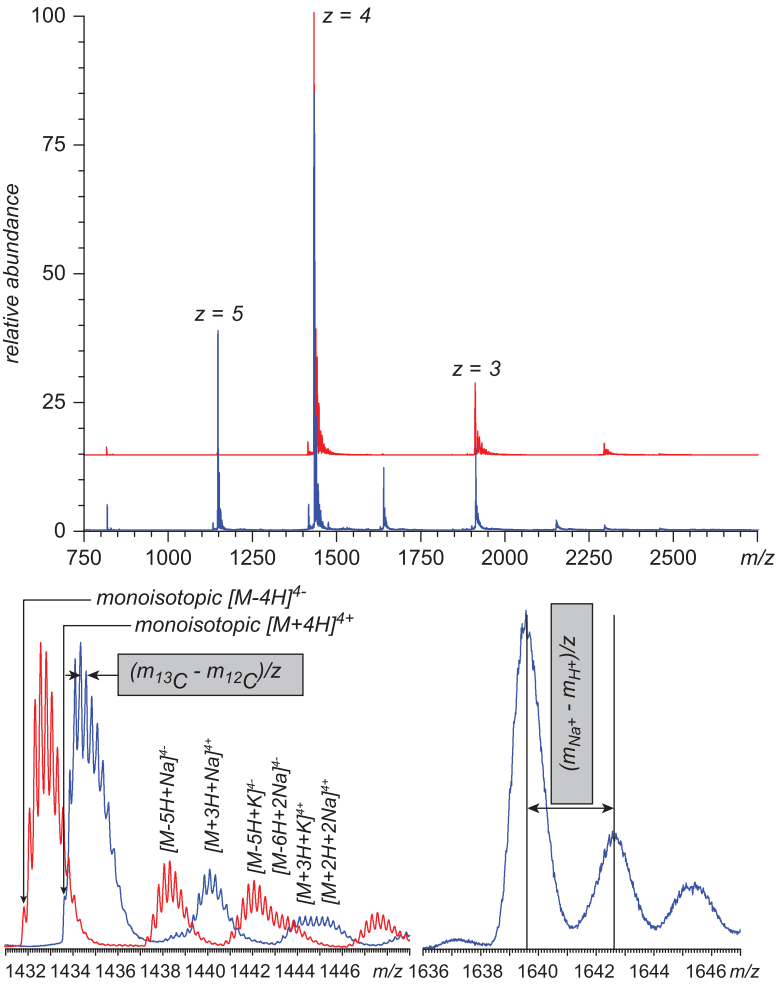


Figure 2.5: ESI mass spectra of insulin acquired in the positive (blue trace) and negative (red) ion modes indicating the number of charges for each protein ion. The diagrams at the bottom of the figure show zoomed views of the ionic signal representing the charge states +4 and −4 of insulin (*left*) and +7 of the non-specific insulin dimer (*right*).

$$\frac{m}{z} = \frac{m_{\mathrm{O}} + (z - n - k) \cdot m_{\mathrm{H}^+} + n \cdot m_{\mathrm{Na}^+} + k \cdot m_{\mathrm{K}^+}}{z} \tag{2.3}$$

If the biopolymer size is sufficiently small to enable resolution of isotopic peaks, the ionic charge can be readily calculated as an inverse of the distance between any two adjacent peaks in the isotopic distribution (see the bottom-left panel in Figure 2.5).

If the isotope-level resolution cannot be achieved (either the protein is too large, or the resolving power of the mass analyzer is too low), the ionic charge state in many cases can be estimated using the m/z distance between the ionic peak representing the

fully protonated species (usually the most abundant for a given charge state) and the next peak in the same charge state cluster (usually corresponding to the single Na^+ adduct), as shown in the bottom-right panel of Figure 2.5 for the insulin dimer. A more universal way of assigning the charge states within the distribution of multiply charged ions of the same species takes advantage of the fact that the adjacent peaks in such distributions are related to each other by a gain of one charge unit ($\Delta z = 1$) and one mass unit ($\Delta m = m_{H+}$). This allows a system of linear equations to be constructed based on (2.2), which have only two unknowns (m_O and the lowest charge state in the distribution), while the number of equations is typically much larger than two (it is determined by the number of detectable charge states in the distribution). Such redundancy is actually beneficial, as it allows the precision of the mass calculation to be improved. All modern mass spectrometers equipped with ESI sources provide an option of automatic deconvolution of the ESI data (the MS data conversion from the m/z scale to the m_O scale), which may be particularly useful if the mass spectra contain ionic signals representing multiple analytes.

As is the case with MALDI MS, ESI is not restricted to generating ions only in the positive ion mode. Usually, any given biopolymer can give rise to both polycations (in the positive ion mode) and polyanions (in the negative ion mode). In most cases, the extent of multiple charging of the polyanionic species is very similar to that of polycations (Figure 2.5), although high-pI species tend to produce more abundant signals in the positive ion mode, while their low-pI counterparts typically fare better in the negative ion mode. The charge state assignment in the negative ion mode is done in the same way as discussed above for polycations; the only difference that equations (2.2) and (2.3) would have to use negative z-values in the numerator and their absolute values $|z|$ in the denominator. Formation of the anionic species usually proceeds via (multiple) deprotonation.

An important distinction of ESI from MALDI is its ability to generate ions continuously, which makes it compatible with all types of mass analyzers (if necessary, the continuously produced ions can be accumulated using various trapping devices followed by their pulsed introduction into the mass analyzer). The continuous mode of ion production, as well as the ability of ESI to generate ions off solutions that are continuously infused into the ion source, makes it also compatible with a range of front-end separation techniques, such as liquid chromatography and capillary electrophoresis (many of these techniques will be discussed in more detail in Chapter 3). However, an important caveat is that ESI works well only with a rather limited range of acids, salts and buffers, strongly preferring weak and volatile electrolytes (such as acetic acid or formic acid). Unfortunately, this prevents the use of popular ion pairing reagents in reversed-phase HPLC, such as trifluoroacetic acid (which is usually substituted with formic acid), as well as NaCl salt gradients in ion exchange chromatography. Both water and volatile organic co-solvents can be used in ESI, while polar low-volatility co-solvents (such as dimethyl sulfoxide) are detrimental for the ESI performance. Ionic strength can be varied within a wide range (from sub-

mM levels to 1 M and even greater), but the choice of salts that can be used to obtain the desired level of the ionic strength is once again limited to volatile electrolytes, such as ammonium acetate and ammonium bicarbonate.

An important modification of ESI is the so-called nanospray (which sometimes is also referred to as nano-electrospray ionization or nano-ESI). As the name suggests, nanospray operates at low flow rates (typically ranging from tens to hundreds of nL/min). The reduced flow rates result in generation of very small charged droplets, whose large surface-to-volume ratio results in accelerated droplet evaporation and significantly improved ionization efficiency [21, 22]. Additional advantages include lower sample consumption, higher tolerance to salts and buffers, more uniform response factors, and improved stability of the spray without the help of the nebulizing gas [23, 24]. Nanospray can be used both in combination with narrow-bore and capillary LC systems that operate at flow rates below 1 μL/min, and in the static mode [25]. The latter uses a low-volume (1–2 μL) metal-coated glass capillary[2] into which the analyte solution is loaded prior to the MS measurements and then pulled out by electrohydrodynamic forces upon application of a high electrostatic potential. This allows a stable signal from an extremely low sample volume to be generated for a very long time period (from tens of minutes to several hours).

2.3 Ion fragmentation: tandem mass spectrometry

Both MALDI and ESI are soft ionization methods in that they enable damage-free desorption/ionization of intact macromolecular ions. However, it is possible to choose ionization conditions that would favor production of abundant fragment ions alongside the intact molecular ions. For example, elevation of the laser power significantly above the optimal level required for the molecular ion production in MALDI readily induces dissociation of even fairly large proteins, giving rise to fragment ions in MALDI spectra that are referred to as the products of in-source decay (ISD) [26]. Likewise, increasing the kinetic energy of ions in the ESI interface region beyond that required for optimal ion desolvation induces fragmentation of molecular ions, a process known as in-source collision-induced dissociation [27]. While the in-source dissociation of molecular ions provides an effective way to obtain structural information on biomolecules in homogeneous samples using inexpensive (single-stage) mass spectrometers [28], the data interpretation becomes increasingly complicated when ions representing multiple molecular species dissociate simultaneously. The interference problem can be avoided if a specific molecular ion is selected prior to its activation, an approach frequently referred to as tandem MS or simply MS/MS [29].

2 Use of or uncoated capillaries with a fine wire insert is also possible.

Ions can be activated in MS/MS experiments using a variety of approaches, including ionic collisions with neutral atoms or molecules (the so-called collision-activated dissociation, or CAD [30]), as well as photon absorption and ionic reactions with charged particles of opposite polarity (e.g., electron-capture dissociation, or ECD involving polycation interactions with low-energy electrons [31]). Tandem MS analysis of biopolymer ions frequently produces valuable structural information, as illustrated in Figure 2.6 using a proteolytic peptide derived from a monoclonal antibody as an example. In the case of peptides, it is frequently possible to extract the amino acid sequence information from the tandem mass spectra, as each ion fragmentation technique favors fission of specific sets of covalent bonds within the polypeptide chain, thereby giving rise to ladders of peaks that are spaced by mass increments that uniquely identify a particular amino acid residue (with the exception of isomeric leucine/isoleucine, which have identical masses). The tandem mass spectrum shown in Figure 2.6 was obtained using low-energy CAD (in which typical collision energies range from tens to hundreds of electronvolts), which strongly favors fission of amide bonds, giving rise to either *b*- or *y*-type of fragment ions (the nomenclature is explained in Figure 2.7). In contrast, high-energy CAD (1–10 keV) produces preferentially *a*- and *x*-type of fragment ions (fission of the C–C bonds), while the electron-based ionization techniques (the already mentioned ECD and its sister technique, electron transfer dissociation or ETD) lead to generation of *c*- and *z*-ions due to fission of the amide nitrogen-C_α bonds in the peptide backbone. The photon-induced fragmentation pathways are also energy dependent: the classical infrared multiphoton dissociation (IRMPD) [32] typically results in fragment ion patterns that are very similar to those generated by low-energy CAD (i.e., mostly *b*- or *y*-ions), while the more recently introduced ultraviolet photodissociation (UVPD) [33] is capable of generating fragments that are typical of high-energy CAD (*a*- and *x*-type). Both high-energy CAD and UVPD also produce abundant fragment ions corresponding to the side chain cleavages, such as *v*- and *w*-ions in Figure 2.7 (a feature typically absent from the fragmentation patterns obtained with low-energy CAD, IRMPD and the electron-based dissociation techniques). This provides an opportunity to differentiate between isomeric leucine and isoleucine residues, which remain indistinguishable from one another if the tandem mass spectrum is populated solely by fragment ions derived from the peptide backbone dissociation. Certain side chain cleavages may also facilitate characterization of disulfide-containing and disulfide-linked polypeptides [34, 35], although such fragments also result in spectral crowding. Another feature contributing to the spectral crowding in tandem MS is the presence of the so-called internal fragments (i.e., dissociation products that result from fission of two covalent bonds within the peptide ion backbone – see the lower m/z range in Figure 2.6). Interpretation of such crowded mass spectra is usually assisted by various automated algorithms; most commercial instruments providing tandem MS capabilities have pre-installed software to accomplish these tasks.

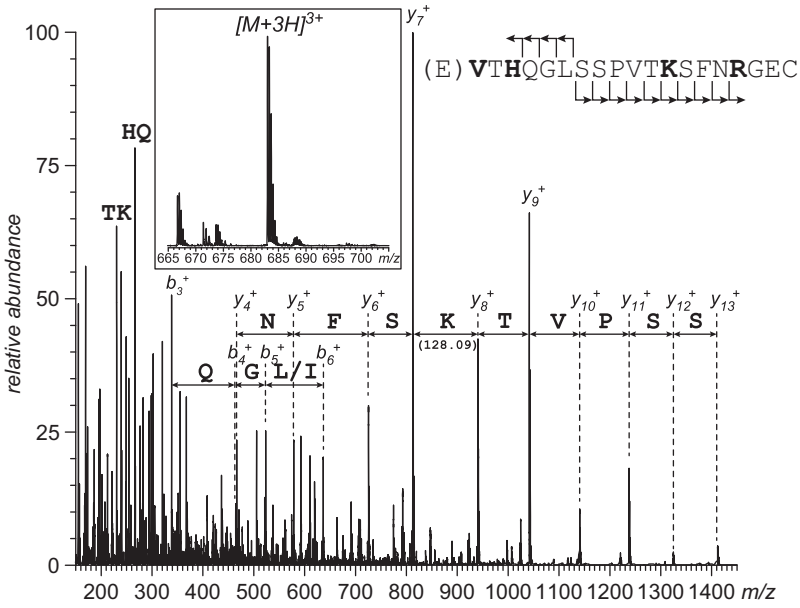


Figure 2.6: CAD products of a peptide ion generated by ESI of a peptic digest of a monoclonal antibody. The intact precursor ion signal (MS1) is shown in the inset.

A commonly used approach to protein structure elucidation using tandem MS combines proteolytic fragmentation in solution followed by MS/MS analyses of peptides, as discussed/illustrated in the preceding paragraphs. However, many modern high-resolution mass spectrometers allow the proteolytic step to be bypassed and the protein fragmentation to be carried out entirely in the gas phase. This approach – termed top-down MS [36] – is rapidly gaining popularity in a variety of applications in the biopharmaceutical analysis field, several of which will be discussed in detail in Chapters 3 and 4. Lastly, all of the discussion so far was focused on the analytical utility of fragment ions derived from positive peptide and protein ions. Since both ESI and MALDI are capable of producing ions in both positive and negative ion modes, tandem MS can be applied to anionic species as well [37]. However, the CAD pathways of peptide and protein anions are usually more complex and frequently feature abundant fragments produced by the side chain losses; while this feature is generally detrimental vis-à-vis sequencing tasks, it may facilitate disulfide mapping work [38, 39]. Anions are obviously less suited for structural interrogation using ECD and ETD due to the anion/electron repulsion; nevertheless, higher energy electrons can interact with anionic species forming the basis of an ion fragmentation technique known as electron detachment dissociation (EDD) [40]. EDD has been shown to be capable of providing structurally diagnostic fragments for both polypeptides and intact proteins [41, 42], although facile side chain fragmentation [43] remains a major factor limiting the utility of this technique [41].

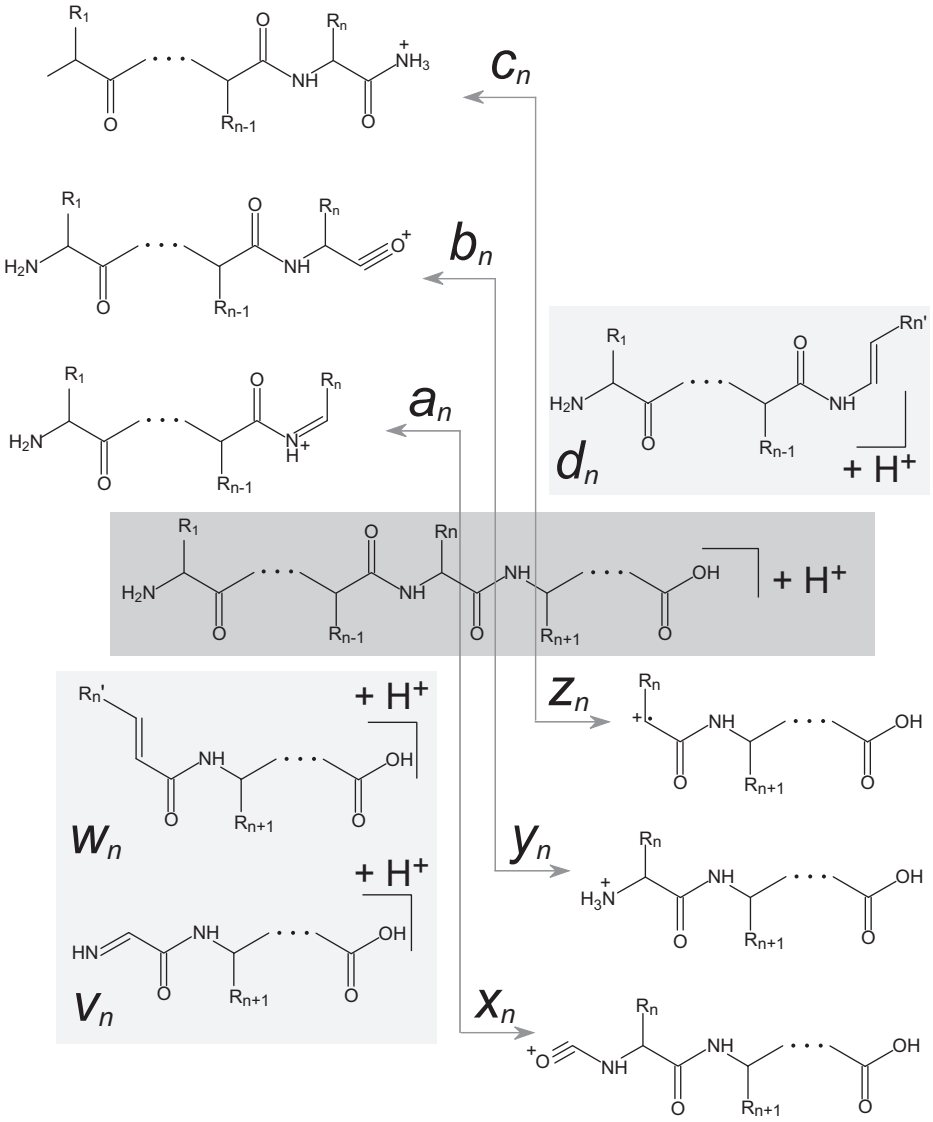


Figure 2.7: Biemann's nomenclature of peptide ion fragments [44]. Fragment ions shown in gray boxes correspond to either complete or partial loss of the side chains and are usually observed only in high-energy collision-activated dissociation.

Anionic species are a much more popular choice when it comes to both MS and MS/MS analysis of other biopolymers, such as oligonucleotides and oligosaccharides. Gas-phase dissociation of these two classes of macromolecules also produces structurally diagnostic fragments (Figure 2.8), although the size of oligonucleotides and oligosaccharides for which meaningful MS/MS data can be generated tend to

be significantly lower compared to polypeptides and proteins. Some examples of using tandem MS as a means of structural interrogation of glycan chains derived from protein therapeutics and oligonucleotide-based medicines will be considered in Chapters 3 and 8, respectively.

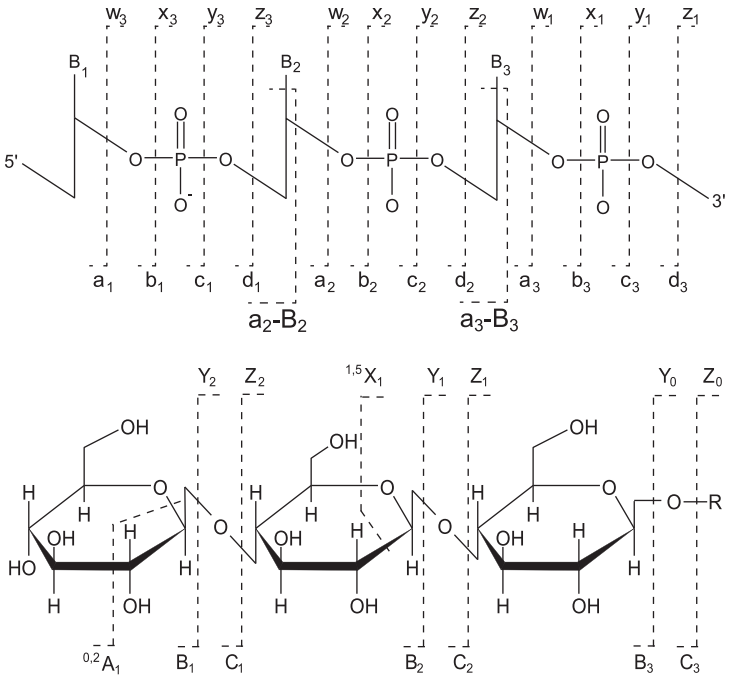


Figure 2.8: Top: McLuckey's nomenclature of oligonucleotide ion fragmentation [45]. Note that formation of a-ions is accompanied by a loss of the base adjacent to the phosphodiester backbone cleavage site. Bottom: The Costello-Domon nomenclature of linear oligosaccharide ion fragmentation [46].

2.4 Ion manipulation in electromagnetic fields: mass analyzers

The ability of mass spectrometers to sort ions according to their m/z ratios is based on the unique way electric and magnetic fields influence the ionic motion in vacuum, as reflected in the mathematical expression of the Lorenz force:

$$m\ddot{\vec{r}} = ze\left(\vec{E} + \left[\dot{\vec{r}} \times \vec{B}\right]\right) \tag{2.4}$$

Although this expression suggests that there are countless ways in which ions can be manipulated with the purpose of determining the m/z ratio, the number of approaches commonly used in modern MS is limited to several scenarios. These include the use of

static electric and magnetic fields (a classical double-sector MS), a combination of electrostatic and electrodynamic fields (quadrupole filters and ion traps), electrostatic fields (time-of-flight MS), and a static magnetic field (ion cyclotron resonance MS). A detailed discussion of all of these types of mass analyzers can be found in many excellent general texts on MS [1–4] and will not be repeated here. Instead, our discussion will be focused on practical issues that frequently need to be considered when a specific instrument has to be selected for a particular analytical task. Therefore, we will discuss several mass analyzers that are most popular in biopharmaceutical applications considering relevant analytical figures of merit, their suitability for accomplishing a set of common tasks (e.g., tandem MS measurements) and their compatibility with common front-end separation techniques.

2.4.1 Quadrupole mass filters and triple-quadrupole MS

A quadrupole mass filter, or quadrupole MS [47, 48], uses a combination of static and harmonic (radio-frequency) electric fields to manipulate ions and acquire mass spectra. It is the most affordable mass analyzer among those that are commonly used in the field. It is also relatively rugged and easy to operate, which makes it very affordable in terms of maintenance costs as well. It can be readily interfaced with an ESI source, but is incompatible with pulsed ionization techniques (such as MALDI). Quadrupole mass filters offer a low resolution (typically a mass-unit resolution) that is constant across the entire m/z range. The latter is usually rather narrow (most commercial instruments offer an m/z range that extends only up to 2,000 m/z units for basic models and up to 6,000 for advanced models). However, changing the frequency of the dynamic component of the electric field allows this limit to be extended to at least 20,000. The latter, however, requires a hardware change (a radiofrequency generator), an option that is not always provided. The data acquisition rates of modern quadrupole mass filters are fairly high, making this instrument compatible with a range of front-end separation techniques. Unfortunately, a single quadrupole mass filter does not allow tandem MS measurements to be carried out. This, as well as the low mass resolution offered by quadrupole MS are two major disadvantages of this mass analyzer.

The lack of the tandem capabilities is addressed in a configuration combining three quadrupoles, of which two act as mass filters, and one (located in the middle of the assembly) serves as an ion guide and a collision cell. This configuration, frequently referred to as a triple quadrupole MS, allows a range of elegant tandem MS measurements to be carried out (including a neutral loss scan and a precursor ion scan, in addition to the common product ion scans). It also enables highly selective measurements where a specific fragmentation reaction is monitored (a technique referred to as a selected reaction monitoring, or SRM). In many applications, SRM provides a very high level of selectivity, enabling very sensitive measurements of low-abundance

compounds in complex biological matrices. Modern triple quadrupole MS instruments are capable of monitoring several reactions simultaneously, a technique known as multiple reaction monitoring, or MRM. The selectivity and sensitivity of SRM- and MRM-based analyses make them very popular as a means of quantitation; such applications are considered in more detail in Chapter 7.

2.4.2 Quadrupole (3-D) and linear ion traps

Quadrupole (3-D) ion trap MS [49] is a mass analyzer that uses physical principles of ion manipulation that are very similar to those employed in quadrupole mass filters, but the ions are retained within the trapping device up until their ejection/detection. The analytical figures of merit of the quadrupole traps are similar to those offered by the quadrupole filters (including modest mass resolution and limited m/z range). A significant advantage of the quadrupole ion traps is their ability to carry out tandem MS measurements. Unlike the "tandem-in-space" scheme employed by the triple quadrupole instruments (where precursor ion isolation, its activation/fragmentation and the subsequent analysis of the fragment ions are all carried out in separate physical compartments of the mass spectrometer), ion traps use the so-called tandem-in-time approach. In this scheme, all stages of ion manipulation in MS/MS measurements are carried out within the single physical compartment, thereby allowing multi-stage tandem MS measurements to be implemented (which are frequently referred to as MS^n – with the superscript "n" indicating the number of fragmentation stages – as opposed to MS/MS or MS^2). MS^n measurements can be particularly useful in situations when the presence of a labile group within a biomolecular ion directs its fragmentation toward a single (lowest energy) dissociation channel, thereby limiting the number of structurally diagnostic fragment ions in MS^2. In this case, the additional activation of such fragment ions (which are already devoid of the labile functionality) may generate more structurally informative dissociation patterns in MS^3. Conversely, the MS^n functionality can also be useful when dealing with over-crowded fragment ion mass spectra by enabling structural interrogation of individual fragment ions produced in MS^2. Although the current technology enables implementation of MS^n with a high number of fragmentation stages, most practical applications do not benefit from extending it beyond MS^3.

While the ability to keep ions inside the small volume of the ion trap is the key to successful implementation of the multi-stage "tandem-in-time" measurements, it is also the source of one of the most significant limitations of this type of mass analyzers. Accumulating a large number of ions inside the small volume of the ion trap gives rise to significant electrostatic interactions within this ensemble and distorts the quadrupolar electric field configuration. This phenomenon, known as a space-charge effect, can result in distortions of the mass spectra and appearance of multiple artifacts. Most modern instruments are equipped with ion counters that allow

the operator to adjust the data acquisition parameters such that the mass spectral quality is not compromised by the space-charge effect.

Ion traps are pulsed mass analyzers and therefore, can be readily interfaced with a MALDI source. Unfortunately, the limited m/z range offered by most ions traps means that MALDI-generated protein ions of even modest size remain off limits; at the same time, lower mass ions (e.g., peptides) can be effectively trapped and analyzed both in MS1 and MS^n modes. Interfacing an ion trap MS with an ESI source relies on external ion accumulation devices that allow the continuously generated ions to be collected externally for a certain period of time followed by their pulsed injection into the mass analyzer. MS^n analyses of the trapped ions can utilize a range of ion activation techniques, with CAD and ETD being the two most popular choices which are available with most commercial instruments of this type. Data acquisition rates are high in both MS and MS^n modes, making ion traps ideally suited for LC-MS and LC-MS/MS applications. A more recent addition to this class of mass analyzers are the linear ion traps, which are essentially quadrupole filters converted to trapping devices by creating a trapping potential well to prevent the ion escape at either end of the quadrupole assembly [4]. Compared to the classical 3-D traps, linear ion traps are more robust and provide better sensitivity and higher mass resolution.

2.4.3 Time-of-flight MS

The time-of-flight (or simply TOF) MS is based on one of the oldest concepts in MS, which nonetheless remains very popular and commercially successful due to continuous technological improvements [50]. Its ability to distinguish among ions with different m/z values is based on the fact that any ion accelerated within a uniform electrostatic field travels in the field-free region at a velocity that is uniquely determined by its m/z ratio. The m/z ratio also dictates the length of the time period required for any given ion to traverse the distance separating the ionization source (where the ions are created and accelerated) and the detector. In order to achieve optimal performance of TOF MS, a range of technological innovations had been introduced over the past several decades, including an ion mirror (presently known as a reflectron) [51] and delayed ion extraction from the ionization source [52]. A detailed discussion of the technical aspects of modern TOF MS is beyond the scope of this book, but an interested reader can find a wealth of information in an excellent text by R.J. Cotter [50].

The TOF mass analyzers are ideally suited for pulsed ionization techniques, such as MALDI, but can also be readily interfaced with ESI using external accumulation of ions followed by their pulsed acceleration. TOF MS are unique among all commercially available mass analyzers in that they offer unlimited m/z range, making them an ideal choice for the analysis of MALDI-generated protein ions, as well as the so-called native ESI MS applications, which also produce ions at high m/z

values (and will be considered in detail in Chapter 4). Unlike the ion trap MS considered above, TOF MS does not suffer from the space-charge effect, which makes it superior vis-à-vis the dynamic range that can be achieved in MS measurements. TOF analyzers equipped with reflectrons and using delayed ion extraction to compensate for the initial velocity distribution of ions feature excellent mass resolution (which can approach 50,000 in some commercially available instruments); however, it places much more stringent requirements on the vacuum compared to the quadrupole mass filters and ion traps considered earlier. Unfortunately, it is always reflected in the instrument price and the maintenance costs.

Tandem MS measurement can be carried out using a single TOF mass analyzer with a reflectron, but it requires that multiple measurements be taken in order to acquire a single MS^2 spectrum. A much more effective way to carry out tandem MS measurements is offered by a TOF/TOF configuration, where a collision cell is placed between the two mass analyzers, one of which is used for selecting the precursor ion (by gating out all other ions upon their arrival at the gate), and the second analyzer is used to determine the m/z values of the fragment ions produced upon collisional activation of the precursor ion. More options are provided by the so-called hybrid mass spectrometers, where a single TOF analyzer works in combination with e.g., a front-end quadrupole filter. The data acquisition rates provided by commercial TOF MS instruments are high, making them an excellent choice for on-line measurements in both LC-MS and LC-MS/MS schemes.

2.4.4 Fourier transform MS: ion cyclotron resonance and Orbitrap mass analyzers

The Fourier transform ion cyclotron resonance (FT ICR) MS, which is frequently referred to simply as FTMS, is a mass analyzer that determines ionic m/z ratios by measuring the frequencies of circular motion of ions in a uniform and static magnetic field [53]. This is accomplished by measuring the image current induced on the receiver plates of the ICR cell by the ions followed by Fourier transformation of the time-domain signal to the frequency domain; each frequency corresponds to a unique m/z value. Since frequency is a physical parameter that can be measured with the utmost precision/accuracy, FT ICR MS is unrivaled in its ability to provide high-resolution, high mass accuracy data. Resolving power exceeding 100,000 can be routinely achieved with most commercial instruments, and specially designed experiments enable measurement at resolving power of several million at m/z below 1,000. Unfortunately, the resolution is inversely proportional to m/z. The unique feature of FTMS is the non-destructive ion detection process, which allows multiple measurements to be taken on the same ionic population over a period of time, thereby enabling MS^n experiments. Collisional activation, IRMPD and ECD can be used to induce ion dissociation, and this repertoire can be further expanded in hybrid instruments using a variety of front-end analyzers/trapping devices in combination with the ICR cell. The

pulsed nature of the FT ICR MS measurements makes it ideally suited for interfacing with a MALDI source, but the m/z range of commercial instruments rarely extends beyond 10,000, keeping most proteins beyond the reach of MALDI/FT ICR MS. Ions produced by ESI are usually accumulated in an external trapping reservoir, followed by their pulsed introduction into the ICR cell (usually via a front-end quadrupole filter). Since the signal resolution in the frequency domain is determined by the length of the signal transient in the time-domain, ultra-high-resolution measurements may require fairly long data acquisition steps (up to several seconds). However, routine measurements are usually carried out using much shorter transients (<100 ms, which is still sufficient to provide mass resolution vastly superior to that of TOFMS), making FT ICR MS compatible with on-line LC-MS and LC-MS/MS measurements. Since all ions are detected simultaneously, FT ICR mass analyzer has a duty cycle of 100%, but the actual duty cycle of the entire mass spectrometer may be lower due to the fact that some ions produced by the external ion source may not be injected into the ICR cell in order to control the size of the ionic population. Indeed, as an ion trapping device, FT ICR MS is susceptible to the space-charge effect, although to a less significant extent compared to the ion trap types considered in the preceding sections, and proper controls are usually required in order to avoid the overpopulation of the ICR cell by ions.

The two factors that make the most significant contributions to the instrument price and the cost of its maintenance are the high vacuum requirements (which is needed to avoid collisional de-synchronization of the ionic precession – the so-called collisional damping – leading to shorter transients and lower resolution) and the strong uniform magnetic fields (7 T and higher in modern commercial instruments), which are maintained by liquid-helium cooled superconducting magnets. The need to have a strong magnetic field is eliminated in another mass analyzer that makes m/z measurements based on the frequencies of periodic ionic motions, Orbitrap MS [54]. The latter also records the signal in the form of the ion-induced image current on the receiver plates, which is then Fourier-transformed to the frequency domain and, therefore, Orbitrap MS is sometimes also referred to as an FTMS [55]. As the m/z values are extracted from frequency measurements, Orbitrap is also capable of achieving very high resolving power (although not as high as FT ICR MS). Unlike the ICR cell, Orbitrap itself is not used for ion activation/fragmentation; instead, such tasks are accomplished using front-end analyzers, which are now an integral part of the hybrid Orbitrap MS instruments [55].

2.4.5 Hybrid mass spectrometers

Most modern MS instruments combine two or more different mass analyzers, which makes them truly universal by allowing the strengths of the individual units to be combined and their shortcomings to be effectively addressed. For example, a quadrupole

mass filter is frequently used as an integral part of TOF MS [56] or FT ICR MS [57], where it can serve as both an ion guide and a front-end mass analyzer, enabling precursor ion selection and its activation outside of the higher end mass analyzer. This eliminates the need to introduce collision gases within the instrument compartments that must be maintained at high vacuum to achieve high-resolution measurements. Ion traps, especially linear ones, are another popular element of the hybrid mass spectrometers enabling effective accumulation and transmission of ions in addition to the precursor ion selection/activation [58]. Several commercial instruments, such as the high-end Orbitrap MS configurations, incorporate multiple analyzers, enabling highly complicated ion manipulation workflows [55]. Lastly, many modern instruments incorporate ion mobility analyzers [59] in addition to mass analyzers. Ion mobility (IM) measurements provide an orthogonal dimension to biomolecular analysis by enabling separation of ions according to their gas-phase mobility (which is dictated by a combination of ionic charge z, its mass m, and collisional cross section σ). The latter frequently allows relatively small isomeric ions to be physically separated from one another despite their having identical masses, essentially proving a separation step that is orthogonal to the commonly used front-end separations (such as reversed-phase LC) [60].

References

[1] Watson, J.T. Introduction to Mass Spectrometry, Edn. 3rd. (Lippincott-Raven, Philadelphia; 1997).
[2] Dass, C. Fundamentals of Contemporary Mass Spectrometry. (John Wiley & Sons, Hoboken, N.J.; 2007).
[3] De Hoffmann, E. & Stroobant, V. Mass Spectrometry: Principles and Applications. (J. Wiley, Chichester, West Sussex, England; Hoboken, NJ; 2007).
[4] Gross, J.R.H. Mass Spectrometry: A Textbook, Edn. 3-d. (Springer International Publishing, Cham, Switzerland; 2017).
[5] Gillespie, T.A. & Winger, B.E. Mass spectrometry for small molecule pharmaceutical product development: a review. Mass Spectrom. Rev. 30, 479–490 (2011).
[6] Korfmacher, W. A. Mass Spectrometry for Drug Discovery and Drug Development. (John Wiley & Sons, Inc., Hoboken, NJ; 2013).
[7] Lee, M.S. & Zhu, M. (eds.) Mass Spectrometry in Drug Metabolism and Disposition: Basic Principles and Applications. (John Wiley & Sons, Inc.; 2011).
[8] Bouslimani, A., Sanchez, L.M., Garg, N. & Dorrestein, P.C. Mass spectrometry of natural products: current, emerging and future technologies. Nat. Prod. Rep. 31, 718–729 (2014).
[9] Bristow, A.W. Accurate mass measurement for the determination of elemental formula–a tutorial. Mass Spectrom. Rev. 25, 99–111 (2006).
[10] Bristow, A.W.T. & Webb, K.S. Intercomparison study on accurate mass measurement of small molecules in mass spectrometry. J. Am. Soc. Mass Spectrom. 14, 1086–1098 (2003).
[11] Marshall, A.G. et al. Protein molecular mass to 1 Da by ^{13}C, ^{15}N double-depletion and FT ICR mass spectrometry. J. Am. Chem. Soc. 119, 433–434 (1997).

[12] Tanaka, K., Ido, Y., Akita, S., Yoshida, Y. & Yoshida, T. Detection of High Mass Molecules by Laser Desorption time-of-Flight Mass Spectrometry. In 2-d Japan-China Joint Symp. Mass Spectrom. 185–188 (Bando Press, Osaka, Japan; 1987).
[13] Karas, M., Bachmann, D., Bahr, U. & Hillenkamp, F. Matrix-assisted ultraviolet-laser desorption of nonvolatile compounds. Int. J. Mass Spectrom. Ion Proc. 78, 53–68 (1987).
[14] Alexandrov, M.L. et al. Ion extraction from solutions at atmospheric pressure - a method of mass spectrometric analysis of bioorganic substances. Dokl. Acad. Nauk SSSR 277, 379–383 (1984).
[15] Yamashita, M. & Fenn, J.B. Electrospray ion source. Another variation on the free-jet theme. J. Phys. Chem. 88, 4451–4459 (1984).
[16] Cole, R.B. Electrospray and MALDI Mass Spectrometry: Fundamentals, Instrumentation, Practicalities, and Biological Applications, Edn. 2nd. (Wiley, Hoboken, N.J.; 2010).
[17] Hillenkamp, F. & Peter-Katalinic, J. MALDI MS: A Practical Guide to Instrumentation, Methods, and Applications, Edn. 2nd. (Wiley-VCH Verlag GmbH, Weinheim, Germany; 2014).
[18] Trimpin, S. et al. Extending the laserspray ionization concept to produce highly charged ions at high vacuum on a time-of-flight mass analyzer. Anal. Chem. 83, 5469–5475 (2011).
[19] Trimpin S. “Magic” Ionization Mass Spectrometry. J. Am. Soc. Mass Spectrom. 27, 4-21 (2016).
[20] Kaltashov, I.A. & Abzalimov, R.R. Do ionic charges in ESI MS provide useful information on macromolecular structure? J. Am. Soc. Mass Spectrom. 19, 1239–1246 (2008).
[21] Gale, D.C. & Smith, R.D. Small volume and low flow-rate electrospray ionization mass spectrometry of aqueous samples. Rapid Commun. Mass Spectrom. 7, 1017–1021 (1993).
[22] Wilm, M.S. & Mann, M. Electrospray and Taylor-Cone theory, Dole’s beam of macromolecules at last? Int. J. Mass Spectrom. Ion Proc. 136, 167–180 (1994).
[23] Juraschek, R., Dülcks, T. & Karas, M. Nanoelectrospray - more than just a minimized-flow electrospray ionization source. J. Am. Soc. Mass Spectrom. 10, 300–308 (1999).
[24] Leney, A.C. & Heck, A.J. Native mass spectrometry: what is in the name? J. Am. Soc. Mass Spectrom. 28, 5–13 (2017).
[25] Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. Anal. Chem. 68, 1–8 (1996).
[26] Hardouin, J. Protein sequence information by matrix-assisted laser desorption/ionization in-source decay mass spectrometry. Mass Spectrom. Rev. 26, 672–682 (2007).
[27] Williams, J.D., Flanagan, M., Lopez, L., Fischer, S. & Miller, L.A. Using accurate mass electrospray ionization-time-of-flight mass spectrometry with in-source collision-induced dissociation to sequence peptide mixtures. J. Chromatogr. A 1020, 11–26 (2003).
[28] Parcher, J.F., Wang, M., Chittiboyina, A.G. & Khan, I.A. In-source collision-induced dissociation (IS-CID): applications, issues and structure elucidation with single-stage mass analyzers. Drug Test. Anal. 10, 28–36 (2018).
[29] Busch, K.L., Glish, G.L. & McLuckey, S.A. Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry. (VCH Publishers, Inc., New York; 1988).
[30] Hunt, D.F., Buko, A.M., Ballard, J.M., Shabanowitz, J. & Giordani, A.B. Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. Biomed. Mass Spectrom. 8, 397–408 (1981).
[31] Zubarev, R.A., Kelleher, N.L. & McLafferty, F.W. Electron capture dissociation of multiply charged protein cations. J. Am. Chem. Soc. 120, 3265–3266 (1998).
[32] Little, D.P., Speir, J.P., Senko, M.W., O’Connor, P.B. & McLafferty, F.W. Infrared multiphoton dissociation of large multiply charged ions for biomolecule sequencing. Anal. Chem. 66, 2809–2815 (1994).

[33] Thompson, M.S., Cui, W. & Reilly, J.P. Fragmentation of singly charged peptide ions by photodissociation at λ=157 nm. Angew. Chem. Int. Ed. 43, 4791–4794 (2004).
[34] Zubarev, R.A. Reactions of polypeptide ions with electrons in the gas phase. Mass Spectrom. Rev. 22, 57–77 (2003).
[35] Fung, Y.M.E., Kjeldsen, F., Silivra, O.A., Chan, T.W.D. & Zubarev, R.A. Facile disulfide bond cleavage in gaseous peptide and protein cations by ultraviolet photodissociation at 157 nm. Angew. Chem. Int. Ed. 44, 6399–6403 (2005).
[36] Sze, S.K., Ge, Y., Oh, H. & McLafferty, F.W. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. Proc. Natl. Acad. Sci. U.S.A. 99, 1774–1779 (2002).
[37] Bowie, J.H., Brinkworth, C.S. & Dua, S. Collision-induced fragmentations of the $(M-H)^-$ parent anions of underivatized peptides: an aid to structure determination and some unusual negative ion cleavages. Mass Spectrom. Rev. 21, 87–107 (2002).
[38] Bilusich, D., Brinkworth, C.S. & Bowie, J.H. Negative ion mass spectra of Cys-containing peptides. The characteristic Cys gamma backbone cleavage: a joint experimental and theoretical study. Rapid Commun. Mass Spectrom. 18, 544–552 (2004).
[39] Bilusich, D. et al. Direct identification of intramolecular disulfide links in peptides using negative ion electrospray mass spectra of underivatised peptides. A joint experimental and theoretical study. Rapid Commun. Mass Spectrom. 19, 3063–3074 (2005).
[40] Budnik, B.A., Haselmann, K.F. & Zubarev, R.A. Electron detachment dissociation of peptide di-anions: an electron-hole recombination phenomenon. Chem. Phys. Lett. 342, 299–302 (2001).
[41] Ganisl, B. et al. Electron detachment dissociation for top-down mass spectrometry of acidic proteins. Chemistry (Weinheim an der Bergstrasse, Germany) 17, 4460–4469 (2011).
[42] Song, H. & Håkansson, K. Electron detachment dissociation and negative ion infrared multiphoton dissociation of electrosprayed intact proteins. Anal. Chem. 84, 871–876 (2012).
[43] Kjeldsen, F. et al. C alpha-C backbone fragmentation dominates in electron detachment dissociation of gas-phase polypeptide polyanions. Chemistry (Weinheim an der Bergstrasse, Germany) 11, 1803–1812 (2005).
[44] Biemann, K. Appendix 5. Nomenclature for peptide fragment ions (positive ions). Meth. Enzymol. 193, 886–887 (1990).
[45] McLuckey, S.A., Van Berkel, G.J. & Glish, G.L. Tandem mass spectrometry of small, multiply charged oligonucleotides. J. Am. Soc. Mass Spectrom. 3, 60–70 (1992).
[46] Domon, B. & Costello, C.E. A systematic nomenclature for carbohydrate fragmentations in fab-ms ms spectra of glycoconjugates. Glycoconj. J. 5, 397–409 (1988).
[47] Leary, J.J. & Schmidt, R.L. Quadrupole mass spectrometers: an intuitive look at the math. J. Chem. Educ. 73, 1142–1145 (1996).
[48] Dawson, P. H. Quadrupole Mass Spectrometry and Its Applications. (Elsevier Scientific Pub. Co., Amsterdam; 1976).
[49] March, R.E. An introduction to quadrupole ion trap mass spectrometry. J. Mass Spectrom. 32, 351–369 (1997).
[50] Cotter, R.J. Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research. (American Chemical Society, Washington, DC; 1997).
[51] Mamyrin, B.A. Time-of-flight mass spectrometry (concepts, achievements, and prospects). Int. J. Mass Spectrom. 206, 251–266 (2001).
[52] Cotter, R.J. The new time-of-flight mass spectrometry. Anal. Chem. 71, 445A–451A (1999).
[53] Marshall, A.G., Hendrickson, C.L. & Jackson, G.S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrom. Rev. 17, 1–35 (1998).
[54] Hu, Q. et al. The Orbitrap: a new mass spectrometer. J. Mass Spectrom. 40, 430–443 (2005).

[55] Eliuk, S. & Makarov, A. Evolution of orbitrap mass spectrometry instrumentation. Annu. Rev. Anal. Chem. 8, 61–80 (2015).
[56] Chernushevich, I.V., Loboda, A.V. & Thomson, B.A. An introduction to quadrupole–time-of-flight mass spectrometry. J. Mass Spectrom. 36, 849–865 (2001).
[57] Kaplan, D.A. et al. Electron transfer dissociation in the hexapole collision cell of a hybrid quadrupole-hexapole Fourier transform ion cyclotron resonance mass spectrometer. Rapid Commun. Mass Spectrom. 22, 271–278 (2008).
[58] Makarov, A. et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. Anal. Chem. 78, 2113–2120 (2006).
[59] Bohrer, B.C., Mererbloom, S.I., Koeniger, S.L., Hilderbrand, A.E. & Clemmer, D.E. Biomolecule analysis by ion mobility spectrometry. Ann. Rev. Anal. Chem. 1, 293–327 (2008).
[60] D'Atri, V. et al. Adding a new separation dimension to MS and LC-MS: what is the utility of ion mobility spectrometry? J. Sep. Sci. 41, 20–67 (2018).

Chapter 3
Characterization of covalent structure of protein therapeutics

Mass spectrometry (MS) has been an indispensable tool for covalent structure analysis of recombinant proteins from the early days of biotechnology, enabling the analysis of both amino acid sequence [1] and post-translational modifications (PTMs), including tasks as challenging as glycan analysis [2] and disulfide mapping [3]. Extensive characterization of various aspects of the covalent structure of biopharmaceutical products remains the primary task of MS in the field of biotechnology, with no other bioanalytical technique capable of providing structural information at the same level of detail and sensitivity.

3.1 Amino acid sequence analysis and characterization of sequence variant types

3.1.1 Confirmation of the amino acid sequence of protein biopharmaceuticals by intact mass measurements

Confirmation of the primary sequence of protein biopharmaceuticals to be consistent with the cDNA sequence is one of the most critical tests routinely performed during the drug development process. In many cases, intact mass measurement provides the most straightforward way of addressing this task by verifying that the polypeptide mass is consistent with the presumed sequence [4]. While both ESI and MALDI are capable of producing intact polypeptide ions, the latter generates mostly singly charged species (for which the m/z values are numerically equal to the polypeptide chain mass), and the ionic signal appears at high m/z values (thousands of m/z units for smaller protein therapeutics, and over a hundred thousand m/z units for mAbs and related protein therapeutics). While such signals can be readily detected using TOF mass analyzers, the resolution afforded by such measurements is low and the mass accuracy in most cases is inadequate vis-à-vis matching the measured mass and the one calculated based on the protein amino acid sequence. Extensive adduct formation is another problem that complicates intact mass measurement for MALDI-generated protein ions. In contrast, ESI offers a unique advantage of producing multiply charged ions that typically fall within a low-m/z range (frequently below m/z 2,000, as shown in Figure 3.1 for a recombinant form of human platelet factor 4, *rh*PF4). A variety of mass analyzers can be employed to record ionic signals in this m/z range with resolution that is sufficient to enable reasonably accurate mass measurements. Another advantage of ESI is its compatibility with front-end separation techniques (both LC and CE), which simplifies the detection of low-abundance sequence variants and proteoforms affected by PTMs (to be discussed in more detail in the following sections of this chapter).

https://doi.org/10.1515/9783110546187-003

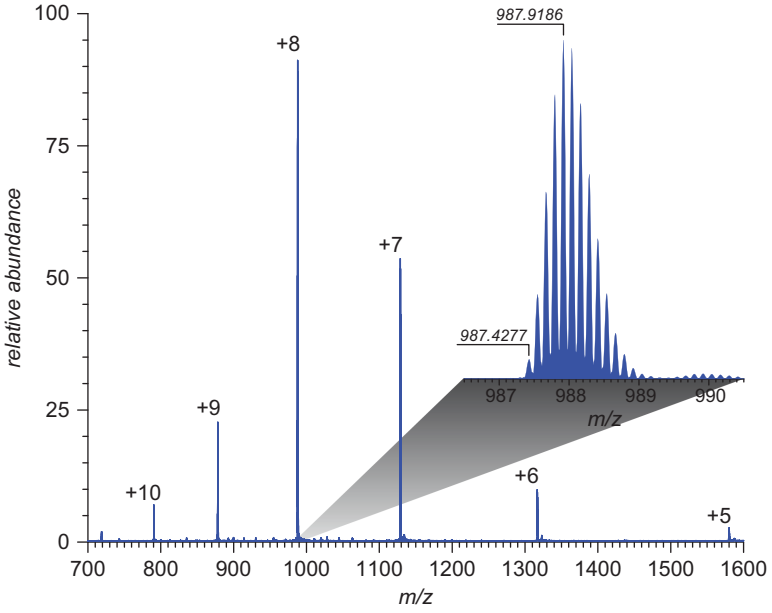


Figure 3.1: Intact mass measurement of recombinant human platelet factor 4. Data courtesy of Dr. Cedric E. Bobst (UMass-Amherst) and Prof. Ishac Nazy (McMaster Univ.).

For smaller recombinant proteins, it is possible to achieve isotope-level resolution (as shown in Figure 3.1 for an 8 kDa polypeptide) using either FT ICR MS or Orbitrap mass analyzers (and in some cases higher end TOF analyzers). The ability to acquire isotopically resolved ionic signal allows monoisotopic masses to be readily determined, which boosts the accuracy of the intact mass measurements (accuracy in the low ppm levels are routinely achieved in such measurements).

Unfortunately, intact mass measurements at the isotopic resolution level are impractical for the majority of protein therapeutics. Even though isotopically resolved ionic signals can be obtained for proteins as large as mAbs using FT ICR MS, the monoisotopic peaks fall below the detection limit, allowing only the average mass to be extracted from such mass spectra [5]. Therefore, such measurements do not provide significant advantages over those carried out with mass spectrometers that do not offer such an impressive resolution (as illustrated in Figure 3.2). The average mass values can be measured with high accuracy (within 10 ppm) for proteins exceeding 100 kDa given proper calibration of the instrument. A notable feature of the mass spectrum presented in Figure 3.2 (shared by the majority of biopharmaceutical products) is the presence of multiple proteoforms (e.g., mAb glycoforms), which obviously complicates the efforts to determine the cumulative mass of polypeptides that comprise a single immunoglobulin molecule. This problem can often be circumvented by carrying out intact mass measurements following

protein deglycosylation [6] (to be discussed in more detail in Section 3.2), although it is not always a trivial task to achieve complete deglycosylation, particularly for proteins containing O-linked glycans.

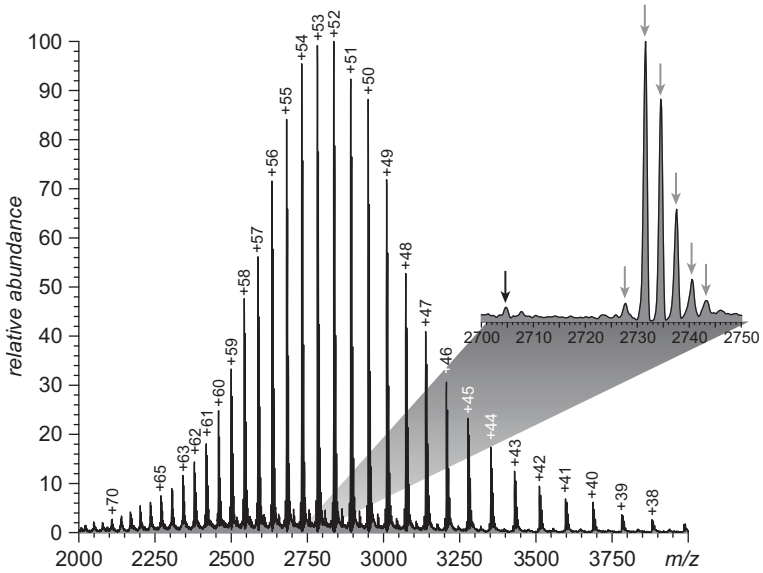


Figure 3.2: Intact mass measurement of a recombinant antibody (IgG1 subtype). The inset shows a zoomed view of protein ions at charge state +54 (the black arrow indicates the protein species lacking a glycan on one of the heavy chains, and the gray arrows indicate various glycoforms). Data courtesy of Dr. Jake W. Pawlowski (Amgen, Cambridge, MA).

Since a significant fraction of large protein therapeutics (such as mAbs) consists of multiple polypeptide chains interconnected by disulfide bonds, confidence of the intact mass analysis can be boosted by measuring the masses of individual polypeptides following the disulfide bond reduction. Mass measurements can also be facilitated by using limited proteolysis, an approach commonly termed "middle-up analysis" (which is distinct from the peptide mapping to be discussed in the subsequent section of this chapter). In the specific case of mAbs, several proteolytic enzymes can be used to dissect the IgG molecules to a small number of fragments whose masses are sufficiently low to afford higher accuracy measurements (e.g., digestion of IgG1 or IgG4 with IdeS to generate a glycan-free $F(ab')_2$ fragment at ca. 100 kDa and two identical Fc/2 fragments at ca. 25 kDa each carrying a single glycan chain [7]). Despite some limitations, intact mass analysis is becoming increasingly popular due to its simplicity and the modest requirement for time/resources that are needed for its completion. This makes intact mass measurements particularly well suited for applications that rely on rapid analyses, such as supporting the upstream process development [8].

3.1.2 Confirmation of the amino acid sequence of protein therapeutics by peptide mapping analysis

The intact mass analysis discussed in the previous section is a convenient way to verify that the recombinant protein mass is consistent with the cDNA sequence. However, in many cases (particularly for larger proteins), the mass accuracy afforded by such measurements is insufficient for confident verification. More definitive sequence confirmation is frequently achieved by peptide mapping analysis, with a protein molecule being enzymatically hydrolyzed into small peptide fragments followed by their mass measurements using LC-MS. The proteolytic step usually employs a protease with known specificity (although non-specific proteases can also be used). The masses of proteolytic fragments can be determined with high accuracy using LC-MS, and this set of numbers can be matched with the predicted peptide masses that are based on the cDNA sequence and the specificity of the protease. The low masses of proteolytic fragments (compared to the intact protein mass) allow isotopically resolved data to be obtained even when relatively inexpensive MS instrumentation is used; utilization of higher resolution mass spectrometers enables mass accuracy in the sub-ppm range. This is illustrated in Figure 3.3 for a tryptic digest of β-thyroglobulin analyzed with a high-resolution LC-MS (an Orbitrap-type mass analyzer). If needed, the amino acid sequence of each peptide fragment can also be determined by means of LC-MS/MS, since most modern mass spectrometers used for such work have tandem capabilities.

Trypsin (a serine protease that specifically cleaves the amide bond at the carboxylic side of lysine and arginine residues) remains the most popular choice for peptide mapping analysis due to its excellent specificity and the high rate of occurrence of Lys and Arg residues in proteins. The majority of tryptic peptides produce abundant ions in ESI, since they have at least two basic sites (the N-terminal primary amine and the C-terminal Lys or Arg residue), which readily accommodate protons during the ionization process. Many commercially available trypsin reagents are chemically modified by reductive methylation at Lys residues, a treatment that greatly reduces the autolysis, thus leading to cleaner and simpler tryptic maps of the target protein. The disadvantages of using trypsin in the peptide mapping work include generation of very short peptide fragments off Lys- and Arg-rich polypeptides (or polypeptide segments). Such peptides typically are not retained by reversed-phase LC and, therefore, escape detection by LC-MS. In addition, large and hydrophobic tryptic peptides can also cause problems, as they normally exhibit poor recovery both during tryptic digestion and during LC separation. Furthermore, the slightly basic conditions commonly applied in tryptic digestion can also lead to undesired artifacts, including asparagine deamidation and disulfide bond scrambling (to be discussed in more detail in Section 3.3.1). Lastly, some low-pI proteins may not be amenable to peptide mapping using trypsin due to the paucity of basic amino acid residues.

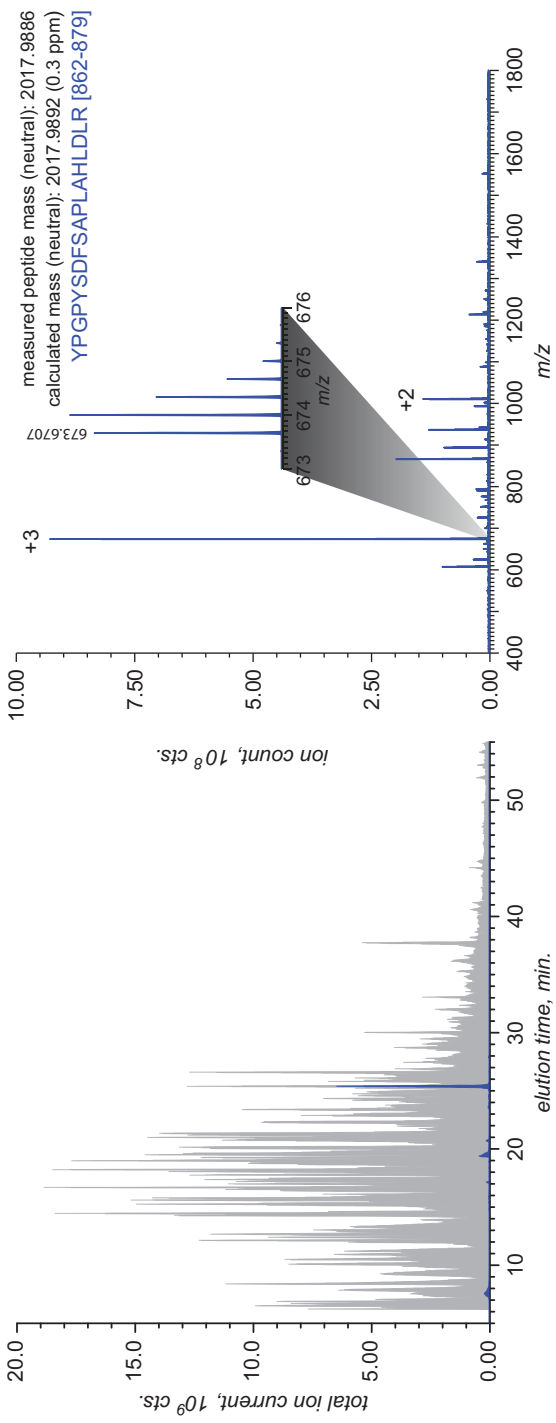


Figure 3.3: TIC of a tryptic digest of reduced and alkylated β-thyroglobulin (gray) and XIC of a +3 charge state of the peptide fragment [862–879], YPGPYSDFSAPLAHLDLR (blue) drawn to scale (***left panel***). The online mass spectrum of the peptide is shown in the right-hand side panel (inset shows the isotopic peaks of the MH_3^{3+} ionic species). Data courtesy of Dr. Chendi Niu (Astra-Zeneca)

Trypsin digestion alone allows extensive sequence coverage of most recombinant proteins to be achieved, although some sequence gaps are to be expected. This can be particularly problematic in the case of peptide mapping of mAbs if the missed sequence segments belong to the critical complementarity-determining regions (CDRs). Achieving full sequence coverage of protein drug molecules is highly desirable and might be expected from the regulatory agencies. In scenarios where trypsin digestion alone cannot provide sufficient sequence coverage, a common strategy to achieve higher (or indeed complete) coverage is to perform a secondary enzymatic digestion using a protease with different specificity [9, 10]. There is a variety of commercially available proteases with specificities orthogonal to that of trypsin, some of which are listed in Table 3.1. Producing two peptide maps using proteases with different specificities not only allows the sequence gaps to be filled, but also provides a means of establishing the correct order of the peptide fragments within the polypeptide chain. Although such a capability is rarely needed when working with recombinant proteins whose sequences are (largely) known, this feature is extremely useful in applications related to de novo sequencing [11].

Table 3.1: Commonly used proteases in peptide mapping analysis of protein therapeutics.

Protease	Family	Specificity	Optimal pH	Characteristics to consider
Asp-N	Metalloprotease	N-Terminal of Asp and Glu (slower)	8	The specificity is orthogonal to trypsin Generates longer peptides with more missed cleavages Most frequently used in combination with trypsin to achieve 100% sequence coverage
Arg-C	Cysteine protease	C-Terminal side of Arg and Lys (slower)	8	Useful to cover Lys-rich segment that might produce short/hydrophilic peptides by trypsin that are difficult to detect
Chymotrypsin	Serine protease	C-Terminal side of Phe, Tyr, Trp, Leu and Met	8	The specificity is highly orthogonal to trypsin Often generates a large number of peptides due to varying efficiency toward different residues Frequently used in combination with trypsin to achieve 100% sequence coverage

Table 3.1 (continued)

Protease	Family	Specificity	Optimal pH	Characteristics to consider
Glu-C	Serine protease	C-Terminal of Glu at pH 4 C-Terminal of Glu and Asp at pH 8	4 or 8	The specificity is orthogonal to trypsin Generates longer peptides with more missed cleavages Exhibited specificity is sensitive to digestion conditions
Lys-C	Serine protease	C-Terminal of Lys	8	The specificity is not orthogonal to trypsin Retains activity under harsh denaturation conditions (such as 8 M urea) Ideal for proteins that are resistant to trypsin digestion or exhibit poor solubility
Lys-N	Metalloprotease	N-Terminal of Lys	8	More resistant to denaturants than trypsin Useful for de novo sequencing

One protease that became very popular in the peptide mapping applications in recent years is Lys-C, which (as the name suggests) catalyzes the cleavage of the amide bond at the carboxyl side of lysine residues. It is particularly well-suited for proteins that are either proteolytically resistant or exhibiting poor solubility in common digestion buffers, owing to its retained activity under strongly denaturing conditions such as 8 M urea. The resilience of Lys-C vis-à-vis acid denaturation can also be used in applications that require peptide mapping to be carried out below the neutral pH to eliminate deamidation artifacts [12]. An important point to consider when using Lys-C for peptide mapping is that the proteases originating from different bacterial hosts (e.g., *Pseudomonas* or *Lysobacter*) exhibit some variation in their cleavage preferences. While this feature might be exploited to improve the sequence coverage in proteomics studies, it also emphasizes the importance of using the protease from the same source to achieve reproducible peptide maps.

While the majority of peptide mapping work in the biopharmaceutical arena relies on reversed-phase LC-MS, CE coupled with online MS detection (CE-MS) is another attractive option that steadily gains popularity. Unlike LC-MS, CE-MS frequently allows a complete sequence coverage of recombinant proteins to be achieved with a single enzymatic digestion [13]. This is due to the nature of CE separation being solely based on the charge-to-size ratios of the analytes, which allows both small/hydrophilic peptides and large/hydrophobic peptides to be efficiently recovered and separated for subsequent MS analysis. Importantly, recent studies have shown that CE-MS is sufficiently robust to allow method transfer across multiple sites [14], making it

well suited for demanding applications in biopharmaceutical analysis. Several alternative avenues to maximize the sequence coverage in peptide mapping of therapeutic proteins are actively explored, including a tandem use of reversed-phase LC and CE with MS detection [15], as well as coupling two columns with different stationary phases. For instance, a porous graphitic carbon (PGC) column can be coupled with a C18 column and used for LC-MS-based peptide mapping analysis, allowing full sequence coverage of therapeutic mAbs to be readily achieved using trypsin digestion alone in a single integrated LC-MS run. Taking advantage of the unique retention characteristic of PGC column (e.g., charge-induced dipole and π–π interactions), small and hydrophilic tryptic peptides that elute from the C18 column with the solvent front can be effectively retained on the PGC column during the loading step (Figure 3.4a). Subsequently, application of two tandem switching valves enables the sequential elution of tryptic peptides from the C18 column and the PGC column for MS detection (Figure 3.4b and c). As a result, short tryptic peptides (e.g., consisting of two to four amino acid residues) could be successfully retained, eluted, and confirmed by both MS and MS/MS analyses.

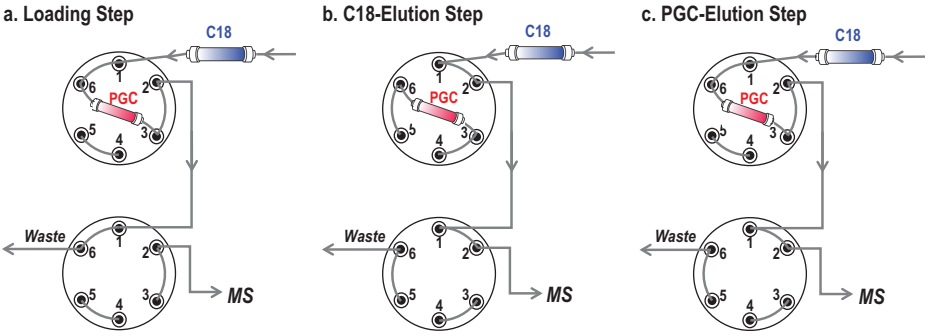


Figure 3.4: Strategy of coupling PGC column with C18 column to achieve 100% sequence coverage from peptide mapping analysis of therapeutic mAbs using trypsin digestion alone in a single integrated LC-MS run. Adapted with permission from Liu et al. [173]

De novo protein sequencing is rarely required in the analysis of therapeutic proteins; instead, peptide mapping analysis is frequently used to reveal low levels of unintended amino acid substitutions, also known as sequence variants [16, 17]. The occurrence of sequence variants in proteins produced by living organisms is inevitable due to the intrinsic error rates associated with each biosynthesis step, including DNA replication (10^{-11} to 10^{-8} per base pair), RNA transcription (10^{-6} to 10^{-4} per base pair), and protein translation (10^{-5} to 10^{-4} per codon). Because the biosynthesis process operates with such high fidelity, the naturally occurring sequence variants in proteins are normally present at very low levels at any given amino acid sites (<0.1%) [18]. In contrast, during recombinant protein production, the occurrence of high-level

sequence variants is more frequently observed and can be attributed to different causes, including high levels of genetic variations in subclones from a selected parent cell line [19], the usage of a non-optimal codon [20] and the starvation of certain amino acids during cell culture [21]. As high levels of sequence variants often indicate a non-optimized process, it is essential to identify and quantitate potential sequence variants early on during upstream development to support the clone selection and cell culture process optimization.

Currently, LC-MS/MS based peptide mapping analysis is the method of choice for sequence variant analysis. The main challenges associated with the sequence variant analysis are the need to detect low-level variant peptides on the background of very abundant peptides having "correct" sequences and the subsequent data analysis. To improve the LC-MS detection of low-level sequence variants, various strategies can be used. For example, application of ultra-performance LC (UPLC) separation as well as a longer gradient during LC-MS analysis might lead to improved detection of variant peptides, as it reduces the possibility of the variant peptides co-eluting with the unmodified peptides. Co-elution not only leads to ionization suppression of the low-abundance ions during the ESI process, but also reduces the likelihood of having them selected for MS^2 fragmentation in a typical data-dependent acquisition (DDA) method. Therefore, the MS acquisition methods should also be optimized to favor the detection of low-abundance variant peptides. For example, the dynamic exclusion function in a DDA method can be employed with a relatively long interval setting so that the instrument will avoid repeatedly fragmenting the same ions representing the abundant unmodified peptides and instead focus on the low-abundance variant peptides. In addition, some intelligent data acquisition strategies developed for proteomics studies might also allow the sequence variant analysis to be improved [22]. For example, an iterative MS/MS sampling strategy may help improve the number of identified sequence variants from consecutive LC-MS/MS analysis of the same digest. During the analysis, the identified precursor ions from the previous run are automatically added to the exclusion list for the next run so that the instrument can be directed to focus on ions that were previously not selected for MS/MS fragmentation due to their low abundance [23].

Once the data acquisition has been completed, identification of sequence variants can be achieved using various data-mining algorithms, such as Byonic [24], BiopharmaFinder [25], and Mascot Error Tolerance Search [26]. Although the identification can be performed automatically, database search for sequence variants often generates high levels of false positives. The reason for this is that during the search, a large number of potential amino acid substitutions, resulting from a single-base change between two codons (substitutions from multiple-base mutation can also be searched, but their occurrences are much lower), will all be treated as dynamic modifications, thus leading to a very large search space. Moreover, the MS/MS data quality for a variant peptide is typically suboptimal due to its low abundance, which can sometimes lead to ambiguity in data interpretation. Finally, the presence of various endogenous

or sample processing-induced PTMs can also lead to misinterpretation of sequence variants if they happen to exhibit the same mass changes [17]. For instance, oxidation of Met residue (addition of an oxygen atom) is isobaric to the Phe/Tyr substitution, while the Cys persulfide modification (addition of a sulfur atom) is isobaric to the Val/Met substitution. Differentiation between a PTM and an amino acid substitution is usually enabled by MS/MS measurements. However, the differentiation becomes impossible if the modification occurs at the same residue as amino acid substitution. For example, the Cys/Ala substitution is frequently observed during reduced peptide mapping analysis of protein samples if TCEP is used as the reductant. However, this particular amino acid substitution is most likely a result of TCEP-induced chemical degradation, which is well documented in the literature [27]. For these reasons, manual verification of the data processing software-reported sequence variants is almost always required, which of course limits the throughput of this analysis. Several different strategies can be employed for results verification. For example, a retention time change of a peptide, as a result of amino acid substitution, can be confidently predicted using an established model [28], which can then be compared to the observed retention time change. This strategy is particularly useful to remove false positives that result from in-source modifications (e.g., oxidation, adduct formation, and neutral losses of H_2O and NH_3) during the ESI process, as such artifacts will exhibit the identical retention time as the unmodified peptides. Moreover, an alternative strategy takes advantage of the highly specific enzymatic digestion to facilitate the confirmation of amino acid substitutions [17]. For instance, confirmation of Gly to Asp substitution can be facilitated by Asp-N digestion, which specifically cleaves at the N-terminal side of the newly formed Asp residue. Furthermore, for amino acid substitution of Cys residues that are participating in disulfide bond formation, verification can be readily achieved by performing the peptide mapping analysis under non-reducing conditions. In this case, a real substitution of Cys should lead to the formation of a novel peptide free of the disulfide bond linkage. Finally, for sequence variant peptides that cannot be confirmed by MS/MS data, the suspected peptide sequence can be synthesized and analyzed in parallel with the protein digests, and the obtained retention times can be compared for highly confident confirmation. Sequence variant analysis remains still a low-throughput assay that is time-consuming and requires significant experience from the analysts to ensure accurate data interpretation. Further developments in this area are highly desirable to make this analysis more automated and less error-prone.

One notable disadvantage of the peptide mapping analysis as a means of assessing the amino acid sequence of protein molecules is the loss of the peptide connectivity information following the enzymatic digestion. Therefore, the primary sequence of a protein cannot be truly confirmed, even if full sequence coverage is achieved. To establish full peptide connectivity, a second protease with orthogonal specificity should be employed to generate another set of peptides that hopefully contain all cleavage sites from the first enzymatic digestion (see Section 3.1.2). To achieve this,

the combination of proteases should be carefully selected using *in silico* digestion tools. For example, if a Lys-Asp motif is present in a protein sequence, the combination of trypsin and Asp-N digestions will not reveal how peptides are connected at this site, as both proteases lead to the same cleavage site. Furthermore, peptide mapping method requires rather involved sample preparation, lengthy data acquisition and complicated data analysis. Although the advances in data processing software have significantly reduced the time and efforts required in data analysis, peptide mapping is still considered a low-throughput assay. Even though other protein quality attributes in addition to primary sequence can be simultaneously analyzed by peptide mapping, faster and simpler approaches are still highly desirable for tasks demanding high-throughput and fast turnaround.

3.1.3 Confirmation of the primary sequence of protein therapeutics by top-down MS analysis

One elegant way to circumvent the problems associated with the peptide mapping work relies on fragmentation of protein ions in the gas phase without using proteolytic degradation in solution. At least in some cases, this can be accomplished by using the so-called top-down or middle-down analyses, in which the ESI-generated intact protein ions (or those representing large protein fragments) undergo dissociation in the gas phase, and the resulting fragment ions are subjected to MS analysis [29]. Similar to the peptide mapping analysis, the amino acid sequence information can be directly extracted from the top-down analysis. However, unlike the peptide mapping analysis, where a platform method might work for many different molecules, significant efforts might be required to develop a molecule-specific top-down method in order to achieve adequate sequence coverage. In general, obstacles that prevent extensive sequence coverage in top-down measurements include large protein size, the presence of disulfide bonds and the presence of glycosylation. Fortunately, those obstacles can be effectively tackled by various strategies. For instance, mAb molecules are normally considered difficult targets for top-down analysis, because they are large (ca. 150 kDa) and contain multiple disulfide bonds. Simple reduction of disulfide bonds in a mAb molecule liberates light (~25 kDa) and heavy chains (~50 kDa), and extensive (but almost never complete) sequence coverage can be readily achieved by the subsequent top-down analysis [30–33]. This approach, frequently referred to as the "middle-down analysis", is steadily gaining popularity in mAb characterization. In addition to reducing the size of the precursor ion, the reduction step also eliminates the internal disulfide linkages, which would otherwise result in sequence gaps (no structurally diagnostic fragment ions are generated by fission of an amide bond within the polypeptide segment flanked by two cysteine residues that are connected via a disulfide bond).

Introduction of the disulfide reduction/cysteine protection steps in the top-down MS workflow results in a notable increase of the time required to complete the

experiment. Furthermore, a significant number of proteins exhibit poor solubility characteristics following reduction of their disulfide bonds. A solution to these problems is offered by a recently introduced technique which combines in-line protein processing using the so-called cross-path reactive chromatography (XP-RC) platform with online top-down MS [34]. The XP-RC platform (depicted schematically in Figure 3.5) takes advantage of vastly different elution rates of small-molecule reagents used for disulfide reduction and large proteins, such as mAbs in SEC. If the protein is injected onto the column (point ***1*** in Figure 3.5) with some delay with respect to the reagent plug injection (point ***0***), then at some instance their trajectories inside the column will intersect (point ***2***), resulting in reduction of the disulfide bonds within the protein. Following the disulfide reduction, the protein subunits will be separated based on their physical size, and elute separately (points ***3*** through ***5***). The mobile phase does not contain any components of the reagent plug up until the time point corresponding to the total permeation limit (which occurs only after all polypeptides have been eluted off the column). Therefore, all small-molecule reagents and denaturants placed in the reagent plug to facilitate the disulfide bond reduction do not interfere with the online MS detection and characterization of the reduced components of the protein [34]. This allows the sequence information to be determined by fragmenting ions representing intact polypeptides, as illustrated in Figure 3.5 for a light chain of a mAb molecule.

Middle-down analysis of the primary structure of protein therapeutics can also be carried out using enzymes that generate a small number of large-size polypeptides. In addition to IdeS (already mentioned in Section 3.1.1) which digests all subclasses of human IgG to generate a single glycan-free $F(ab')_2$ fragment at ca. 100 kDa and two identical Fc/2 fragments at ca. 25 kDa each [7] and related endopeptidases IdeE, IdeE2, IdeZ, and IdeZ2 [35], several other proteases are actively explored (Figure 3.6). One particularly promising enzyme is gingipain K (also known as Lys-gingipain or PrtP proteinase, EC 3.4.22.47), a cysteine protease from *P. gingivalis*, which is now commercially available in the recombinant form. The recombinant form of gingipain K is reported to hydrolyze human immunoglobulins of the IgG1 subclass at a single digestion site above the hinge region (KSCDK/THTCPPCP) generating a homogenous pool of intact Fab and dimeric Fc fragments [36]. Although the fragment ion sets obtained in middle-down analyses relying on IdeS and gingipain K digestions are partially redundant, the parallel use of the two proteases allows the extent of mAb sequence coverage to be significantly extended, pushing the light chain coverage to 65% and the heavy chain to 50% [37]. Lys-C, a serine protease already discussed in the preceding section in the context of the peptide mapping work, can also be employed in the middle-down workflow. Under near-native conditions, it preferentially cleaves the IgG1 molecules after a Lys residue right above the hinge region disulfide bonds (the gingipain K cleavage site), leading to the formation of two identical Fab fragments (ca. 50 kDa each) and a dimeric Fc fragment (ca. 50 kDa). This reaction can be completed on a very short timescale, as

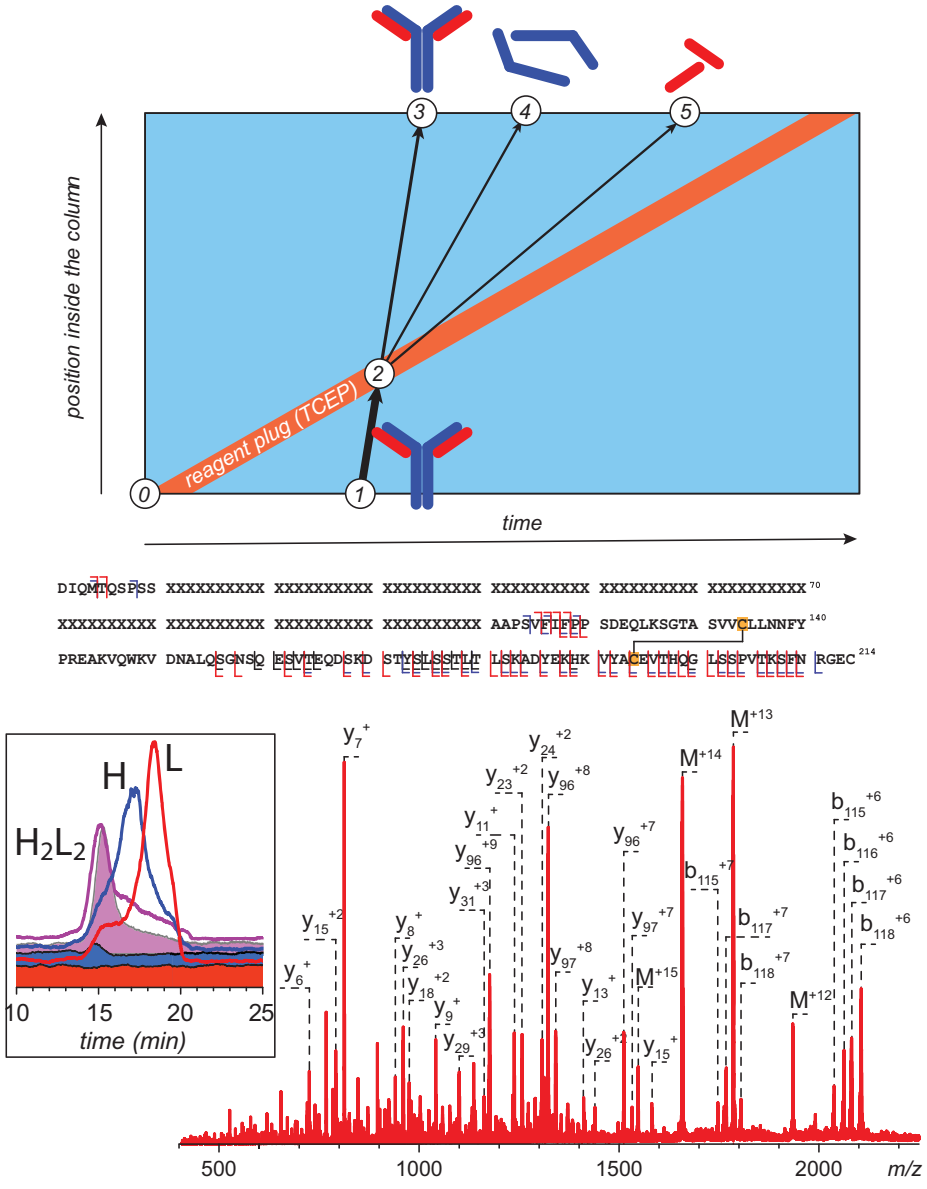


Figure 3.5: Top: Schematic representation of XP-RC using a 2-D depiction of the chromatographic process. The numerals on the diagram indicate injection of the low-molecular weight reagent plug (0), injection of the protein (1), chemical reaction between the protein and the reagent (2), and elution of the unreacted protein species (3) and the products of the chemical reaction (4 and 5). Bottom: XP-RC MS/MS of mAb showing CAD mass spectrum and fragmentation pattern of the L-chain produced upon the in-line reduction of the intact protein. Amino acid sequence is shown only for the constant region of the L-chain; the vertical lines indicate amide bonds whose cleavage gives rise to the detected *b*- and *y*-ions (red lines correspond to this data set; blue lines correspond

long as a sufficiently high enzyme-to-substrate ratio is used (e.g., 5 min for the E/S ratio of 1/400) [38].

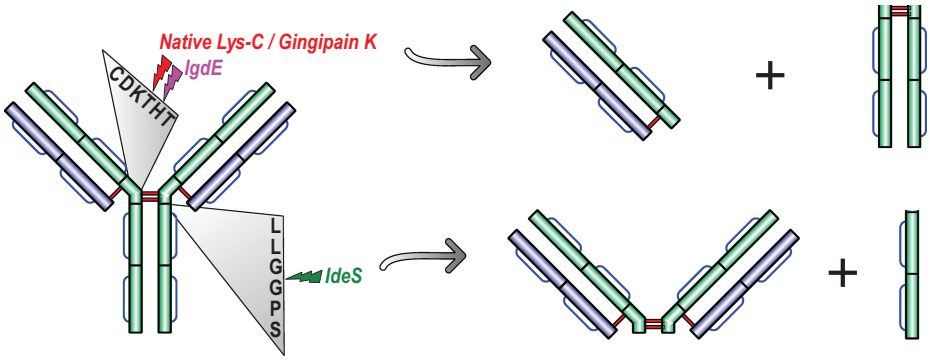


Figure 3.6: Schematic representation of dissecting an IgG1 molecule into specific fragments.

In addition to expanding the repertoire of endopeptidases that can be used in the middle-down analysis of protein therapeutics, there has been a revived interest in chemical cleavage agents that can be employed for selective fission of polypeptide chains. In addition to the now-classic cyanogen bromide (CNBr) that hydrolyses a peptide bond on the C-terminal side of methionine residues, several other selective agents targeting amino acids that have low frequency of occurrence (e.g., 3-bromo-3-methyl-2-(2-nitrophenyl) sulfanylindole, and 2-nitro-5-thiobenzoic acid, targeting Trp and Cys, respectively) show a particular promise [39].

In parallel to continuous improvements/innovations in the upfront sample preparation/manipulation, recent progress in the gas-phase protein/polypeptide ion manipulation techniques has also greatly facilitated the top-down MS analysis of protein therapeutics. While such measurements traditionally relied on collision-activated dissociation (CAD), introduction and development of other gas-phase fragmentation techniques, such as higher energy collisional dissociation (HCD) [40], electron transfer dissociation (ETD) [41], electron capture dissociation (ECD) [42], and ultraviolet photodissociation (UVPD) [43] all contributed to the progress in the top-down protein analysis. It is important to note that the combination of alternative fragmentation techniques often provides the most comprehensive sequence coverage of the protein [44].

Figure 3.5 (continued)
to XP-RC MS/MS measurements carried out without using guanidinium chloride in the reagent plug; and black lines correspond to fragments generated in XP-RC MS using in-source collisional activation). The inset shows selected XICs for several ionic species in XP-RC chromatogram (reference XICs obtained in the absence of the reducing agent in the reagent plug are shown as color-filled curves). Adapted with permission from Pawlowski et al. [34] Copyright 2018 American Chemical Society.

Due to the complexity and structural heterogeneity of most biopharmaceuticals, the top-down MS analysis is usually implemented using online separation (most commonly, reversed-phase LC, although other separation modalities such as ion exchange [32, 45] or CE [46] are also frequently used). Recent experimentation with ion mobility as an orthogonal means of proteoform separation prior to MS analysis led to a conclusion that it can replace liquid-based separations and reduce both the analysis time and the cost [47]. The top-down MS analysis of the protein amino acid sequence remains a dynamic and rapidly evolving field, poised to claim a steadily increasing share of the analytical work in the biopharmaceutical sector (with its contributions extending to tasks beyond those covered in this chapter – other applications of top-down MS will be considered in Chapter 4). For more in-depth information on top-down MS, the readers are referred to a comprehensive review published in *Nature Methods* (and the anticipated follow-up papers) discussing best practices in this field [48].

3.2 Characterization of enzymatic post-translational modifications

Compared to small-molecule drugs, protein biopharmaceuticals are inherently heterogeneous due to their complex chemical structure and the propensity to accommodate multiple enzymatic and chemical post-translational modifications (PTMs). The presence of PTMs, no matter if they arise due to peculiarities of the production process, are introduced by design (the so-called designer PTMs) or induced under stress conditions, might have an important impact on drug safety and efficacy, which need to be closely monitored during the drug development. For example, protein PEGylation, a frequently used designer PTM, has enjoyed a tremendous success in extending the half-lives of protein biopharmaceutical in circulation [49–51]. In contrast, when occurring at the CDR regions of therapeutic mAbs, unexpected PTMs might cause a loss of their function [52]. From the quality control perspective, many PTMs might not be of direct concern to drug safety and efficacy, however, they can significantly contribute to the charge heterogeneity profile of the drug molecule, which is monitored by various assays in QC release to ensure product and process consistency. As a result, those PTMs are still considered important quality attributes that might prevent the drug batches release. In addition, PTMs also contribute significantly to the overall heterogeneity of protein biopharmaceuticals. To date, there are nearly 1,500 entries of protein modifications on UNIMOD (www.unimod.org/, a comprehensive protein modification database), within which a large percentage are unique PTMs affecting specific amino acid side chains. A biopharmaceutical product as complex as mAb contains dozens of sites at which PTMs commonly occur (Figure 3.7).

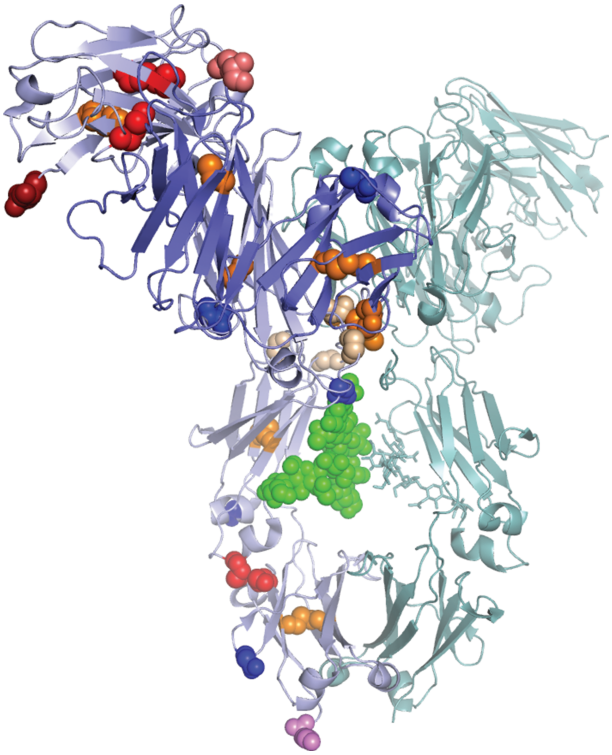


Figure 3.7: Common PTM targets within a mAb molecule shown in a space-fill format: N-glycosylation (green), methionine oxidation (red), tryptophan oxidation in the CDR region (pink), asparagine deamidation (blue), cysteine reduction/conjugation/disulfide scrambling/trisulfide formation (orange), N-terminal glutamine and glutamic acid cyclization to pyroglutamine (brick and salmon), backbone cleavage (wheat), and C-terminal lysine removal (magenta). The PTM sites are shown only within a single pair of light and heavy chains to avoid the clatter.

Because the presence of most PTMs can lead to a measurable mass change, MS has been the analytical tool of choice for both their identification and quantitation. In a typical bottom-up approach (e.g., peptide mapping), the PTM-containing peptides, generated from the enzymatic digestion of protein samples, can often be effectively separated from their unmodified counterparts by reversed-phase LC prior to MS detection. Preliminary identification of a particular PTM is often achieved by the combination of accurate MS1 measurement and the pre-existing knowledge of common PTMs. Subsequently, confirmation and localization of the PTM to a specific amino acid residue can be achieved by MS/MS analysis. Finally, the relative abundance of the PTM can be estimated based on the MS1 responses of both the modified and unmodified peptides. This section is focused on enzymatic PTMs and MS-based methods to characterize them; non-enzymatic and designer PTMs will be considered in Sections 3.3 and 3.4, respectively.

3.2.1 N-Glycosylation

With a very few exceptions, most therapeutic proteins are glycoproteins, i.e., they have at least one N- or O-glycan chain attached to a specific amino acid side chain. Protein glycosylation is an enzymatic process, which is often considered the most complex PTM due to the involvement of a large number of enzymes and mechanisms, and a variety of structures that are produced (Figure 3.8). Properly formed glycosylation is often critical for the biopharmaceutical product's efficacy, stability and safety profiles. For example, therapeutic mAbs contain a conserved N-linked glycosylation site in the Fc region, which is known to impact the antibody effector functions. Antibodies with lower fucosylation exhibit enhanced antibody-dependent cell-mediated cytotoxicity (ADCC) as a result of binding to Fc-γ receptors; antibodies with higher terminal galactose have enhanced complement-dependent cytotoxicity (CDC) as a result of binding to complement protein C1q [53, 54]. The N-linked glycosylation in the Fc region can also impact the pharmacokinetics of antibody-related molecules. For instance, Fc fusion proteins with terminal *N*-acetylglucosamine (GlcNAc) or high mannose glycans were shown to exhibit significantly shortened half-lives comparing to other glycoforms in both Cynomolgus monkeys and humans [55, 56]. In addition, the N-glycans in IgG molecules also contribute to the stability of the CH2 domain, which is frequently the least thermally stable domain within the molecule [57]. Finally, the presence of nonhuman monosaccharide units in glycosylated protein biopharmaceuticals, as a result of producing recombinant glycoproteins in nonhuman mammalian cells, can elicit unwanted anti-drug antibody (ADA) responses in patients and also promote drug clearance [58, 59]. In addition to considerations related to drug safety and efficacy, the batch-to-batch consistency in glycosylation profile can demonstrate a well-controlled process, while inconsistent glycosylation profile might indicate changes in the process. Finally, development of the so-called biosimilars requires even more efforts from the copycat to mimic the glycosylation profile of the originator, in order to establish the analytical similarity between the two (see Chapter 5 for more detail). Therefore, it is becoming increasingly important to achieve in-depth and comprehensive characterization of glycosylation profiles in protein biopharmaceuticals.

Protein glycosylation most frequently occurs on asparagine residues (N-glycans) within the Asn-X-Thr/Ser (where X could be any amino acid with the exception of proline) consensus motifs, and on serine and threonine residues (O-glycans). In general, two levels of information can be obtained from protein glycosylation analysis: the glycosylation site and occupancy (macro-heterogeneity) and the glycan's structure at each site (micro-heterogeneity). Depending on the purpose of the analysis, protein glycosylation can be characterized using a variety of techniques to achieve either high speed or high sensitivity. In the following paragraphs, a range of strategies used for N-glycosylation analysis will be reviewed in detail.

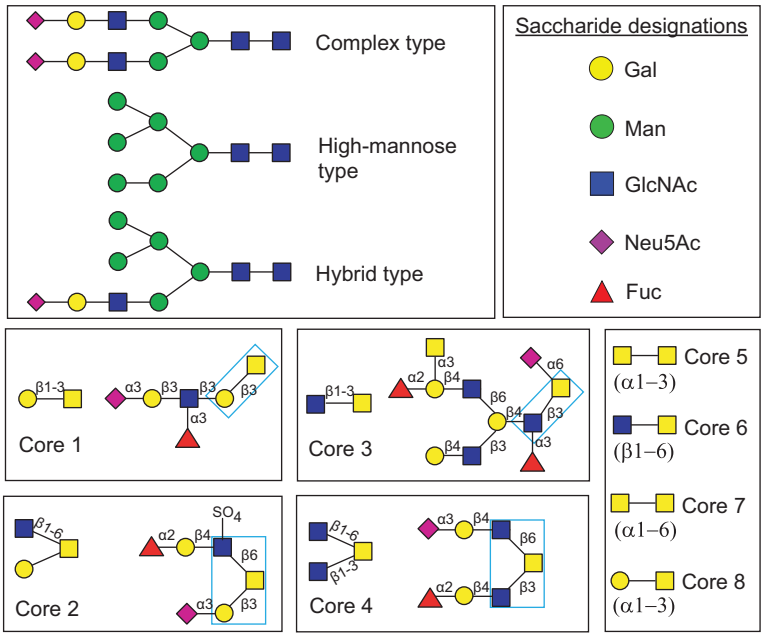


Figure 3.8: Structures of commonly occurring protein N- and O-glycans (top and bottom, respectively). Representative extended O-glycan structures built upon cores 1–4 are shown in the same boxes as their core structures.

Both macro- and micro-heterogeneity of protein glycosylation can lead to increased mass heterogeneity of protein biopharmaceuticals, which often can be directly characterized by intact mass-based approaches. For therapeutic mAbs, each of the two identical N-glycosylation sites in the Fc region can be either vacant or occupied, resulting in mAb molecules with 0–2 N-glycans (macro-heterogeneity). At each site, the glycan's structure can also vary in monosaccharide composition, branching pattern and length (micro-heterogeneity). Fortunately, such mass heterogeneity introduced by N-glycans in a mAb molecule can be readily solved by modern mass spectrometers at the intact protein level (Figure 3.9). The mass differences of 146, 162, and 291 Da observed between the major glycoforms correspond to the masses of fucose, galactose and sialic acid residues, respectively. However, analysis of the intact mAb molecule, particularly without front-end separation, often fails to detect minor glycoforms due to the limited dynamic range in direct MS analysis. In addition, glycoforms with similar molecular weight, such as G0F and G1 ($\Delta M = 16$ Da), can be difficult to resolve if other PTMs are also present (particularly oxidation, which results in the same mass shift). In order to improve the method resolution and sensitivity, the mAb molecule can be dissected into smaller subunits using the strategies discussed in Section 3.1, and then the intact mass analysis can be performed on the Fc/2 fragment (25 kDa). This approach not only reduces the size of

the polypeptide chain, but also limits the number of the glycan chains to not more than one per each polypeptide. Furthermore, analysis of protein glycosylation at intact levels can also be facilitated by front-end separation techniques. The recently demonstrated capabilities of both HILIC [60, 61] and ion exchange (IEX) LC [45] to be interfaced with ESI MS provide exciting opportunities for intact mass analysis of chromatographically resolved glycoforms of glycosylated protein therapeutics.

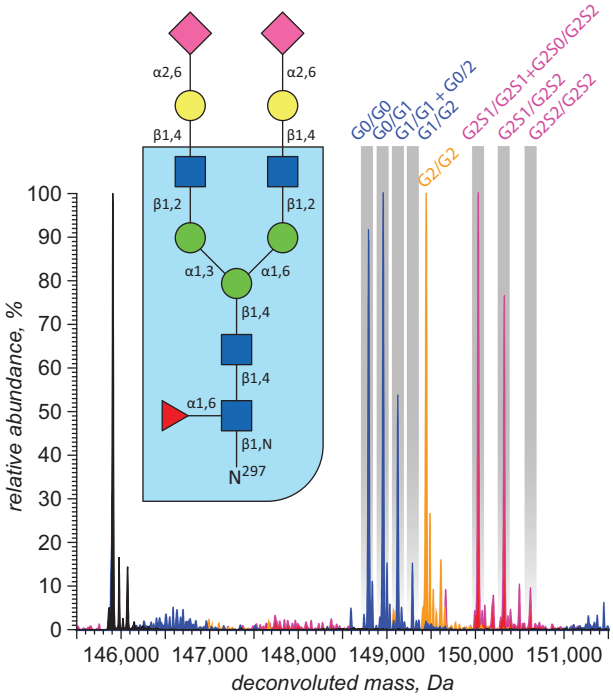


Figure 3.9: Deconvoluted mass spectra of intact (blue), fully deglycosylated (black), hypergalactosylated (yellow), and hypersialylated (magenta) forms of IgG1. The inset shows the structure of the glycan chain. Reproduced from Pawlowski et al. [62] with permission from Elsevier.

Many applications require a detailed characterization of the N-glycan structures that goes beyond establishing the composition or even the connectivity order of individual monosaccharide residues within the chain. For example, there is a difference in sialic acid linkages between endogenous human proteins and those recombinantly expressed in Chinese hamster ovary cell (CHO) cells ([2,6] vs. [2,3]), and confirmation of the correct linkage type is sometimes required in order to address concerns related to a potential immunogenicity of therapeutic proteins. To achieve such an in-depth characterization of N-glycans in protein biopharmaceuticals, released N-glycan analysis is frequently performed, where the glycoprotein is first treated with PNGase F to release the intact oligosaccharides that are subsequently isolated and derivatized for detection.

Historically, capillary electrophoresis coupled with laser-induced fluorescence detection (CE-LIF) was the technique commonly used for carbohydrate analysis due to its high resolving power, short analysis time and exceptional sensitivity. However, this method is highly dependent on using standards to confirm N-glycan identities. As a result, its utility is limited when less common or abnormal N-glycans are present in a biopharmaceutical product. In contrast, MS-based detection methods, particularly the application of tandem MS, have dramatically alleviated the dependence on using standards to elucidate the N-glycan structure. MS-based glycan sequencing is now a mature analytical technology, which is routinely used not only for determining the order of monosaccharide units within linear segments of the glycan chain, but also for deciphering the branching patterns (which is made possible by the so-called cross-ring cleavages within single saccharide residues) [64–66]. One disadvantage of using tandem MS for glycan sequencing is its reliance on chemical derivatization, although the latter can be avoided if the MS analyses are carried out in the negative ion mode [67].

The multiplicity of fragment ions that are usually generated in the course of tandem MS measurements of glycan chains leads to crowded mass spectra, which may be difficult to interpret without specialized data processing algorithms [68]. Furthermore, complete sequence/branching information on complex glycans is almost never achieved using tandem MS alone. Instead, most current protocols rely on availability of a battery of highly specific exoglycosidases that can be used to distinguish positional and/or linkage isomers (Table 3.2), with MS serving as a tool to report a success (or failure) of a specific exoglycosidase (or a cocktail of exoglycosidases) to remove terminal saccharide residues. Recent advances in both hyphenated CE-MS and LC-MS methods have further expanded the role played by MS in released N-glycan analysis [69, 70]. Notably, the excellent resolving power of CE and HILIC methods often allows

Table 3.2: Common analytical exoglycosidases [71] (an up-to-date web-based version is available at the manufacturer's site www.neb.com/tools-and-resources/feature-articles/redesigning-glycosidase-manufacturing-quality–for-pharmaceutical-and-clinical-applications).

Common abbreviation	Enzyme name	Specificity	Origin
NAN1	α2-3 Neuraminidase S	α2-3 Neu5Ac	*Streptococcus pneumoniae*
ABS	α2-3,6,8,9 Neuraminidase A	α2-3,6,8,9 Neu5Ac/Neu5Gc	*Arthrobacter ureafaciens*
SPG	β1-4 Galactosidase S	β1-4 galactose	*Streptococcus pneumoniae*

Table 3.2 (continued)

Common abbreviation	Enzyme name	Specificity	Origin
BTG	β1-3,4 Galactosidase	β1-3,4 galactose	Bovine (testes)
GUH	β-*N*-Acetylglucosaminidase S	β-*N*-Acetylglucosamine	*Streptococcus pneumoniae*
JBM	α1-2,3,6 Mannosidase	α1-2,3,6 mannose	Jackbeans
BKF	α1-2,3,4,6 Fucosidase	α1-2,3,4,6 fucose	Bovine (kidney)
AMF	α1-3,4 Fucosidase	α1-3,4 fucose	Almondmeal
CBG	α1-3,4,6 Galactosidase	α1-3,4,6 galactose	Coffeebeans
OTF	α1-2,4,6 Fucosidase O	α1-2,4,6 fucose	Omnitrophica bacterium
NAN1	α2-3 Neuraminidase S	α2-3 Neu5Ac	*Streptococcus pneumoniae*

structural isomers to be separated from each other, while the structure elucidation can be greatly facilitated by the subsequent tandem MS analysis.

Identification of the N-glycosylation sites is another task that is usually solved using various MS-based approaches. While intact mass measurements are usually sufficient for characterization of protein biopharmaceuticals containing a single glycosylation site, it is less useful for proteins with multiple glycosylation sites, as it cannot reveal site-specific glycosylation information. Additionally, the overall mass heterogeneity of the protein molecule increases dramatically along with the extent of glycosylation, resulting in a wide (and frequently poorly resolved) distribution of ionic masses, making intact mass analysis nearly useless. Although the capabilities of intact mass analysis to handle highly heterogeneous systems continue to expand [72], its applicability remains limited in the realm of extensively glycosylated protein therapeutics. On the other hand, a significant reduction in heterogeneity can be achieved by using bottom-up strategies, where the glycoprotein is first digested into proteolytic peptides. Ideally, a selective protease is used so that each glycopeptide contains a single glycosylation site. LC-MS analysis of the digests can be used to determine the level of glycan occupancy at each site by comparing signals from the glycosylated peptides and from the non-glycosylated peptide. Alternatively, the digests can be further treated with PNGase F, which removes the N-glycans and converts the asparagine at the glycosylation site to aspartic acid. In this case, the glycan occupancy can be quantitated by comparing signals from the native peptide (i.e., containing unmodified asparagine residue) and from the deamidated peptide, although one should be mindful of a likely change in the peptide ionization efficiency due to incorporation of an acidic residue

into its sequence. This strategy can be further enhanced by performing the PNGase F treatment in ^{18}O-enriched water so that each N-glycosylation site will incorporate an ^{18}O atom after conversion, greatly improving the confidence of identification [73].

In addition to overall glycan occupancy, distribution of glycoforms at each individual site can also be evaluated by the bottom-up-based methods. The separation of glycopeptides prior to MS analysis is most frequently achieved by reversed-phase chromatography. Because the hydrophobicity of a glycopeptide is mostly determined by its peptide portion, glycopeptides with different amino acid sequences (e.g., containing different glycosylation sites) can often be effectively separated on a C18 column. However, glycopeptides with the same amino acid sequence but different glycoforms often co-elute on a C18 column. This feature is useful for quick detection and identification of all major glycoforms from the same peptide, but it might also cause less sensitive MS detection of minor glycoforms due to the ionization suppression. Alternatively, the glycopeptides can also be separated by HILIC, where the retention is mostly determined by the glycan portion. As a result, different glycoforms for a given peptide can be readily separated [74]. More recently, the combination of reversed-phase column and porous graphitic carbon column has also been used to retain extremely hydrophilic glycopeptides in order to improve the coverage of the glycosylation sites [75].

Besides the advancements in the upstream separation methods, various MS-based tools are also available for glycopeptide characterization. For purified protein drug samples, highly accurate mass measurement of the glycopeptides from the protein digests might already provide some degree of confidence for identification, which can be further enhanced using the glycopeptide ion fragmentation data from tandem MS analysis. Historically, CAD has been commonly used for glycopeptide ions fragmentation. During the CAD analysis, the glycosidic bonds are preferentially cleaved as they are more labile than amide bonds, leading to sequential losses of monosaccharide units from the termini of the glycan structure (Figure 3.10). This information is frequently useful for elucidating both the composition and the connectivity order of individual monosaccharide units within the glycan chains. For known glycosylation sites, CAD analysis alone is frequently sufficient to identify the glycopeptides and even elucidate the glycan structure (Figure 3.10). However, as CAD analysis of glycopeptides cannot reveal amino acid sequence information, it's less useful to identify unknown glycosylation sites.

For protein biopharmaceuticals, the amino acid sequence information is always known so that one can easily predict the N-glycosylation sites based on the well-known motifs (e.g., NXT/S, *vide supra*). However, it is worth noting that N-glycosylation could also occur on non-consensus sites (including N-glycosylation at NSG, NQV and NEN observed in recombinant antibodies [77]), although typically at much lower levels. Detection of the occurrence of atypical N-glycosylation in protein biopharmaceuticals by routine methods can be challenging due to their low abundance and unpredictability. Fortunately, the recently reported native ion exchange LC-MS

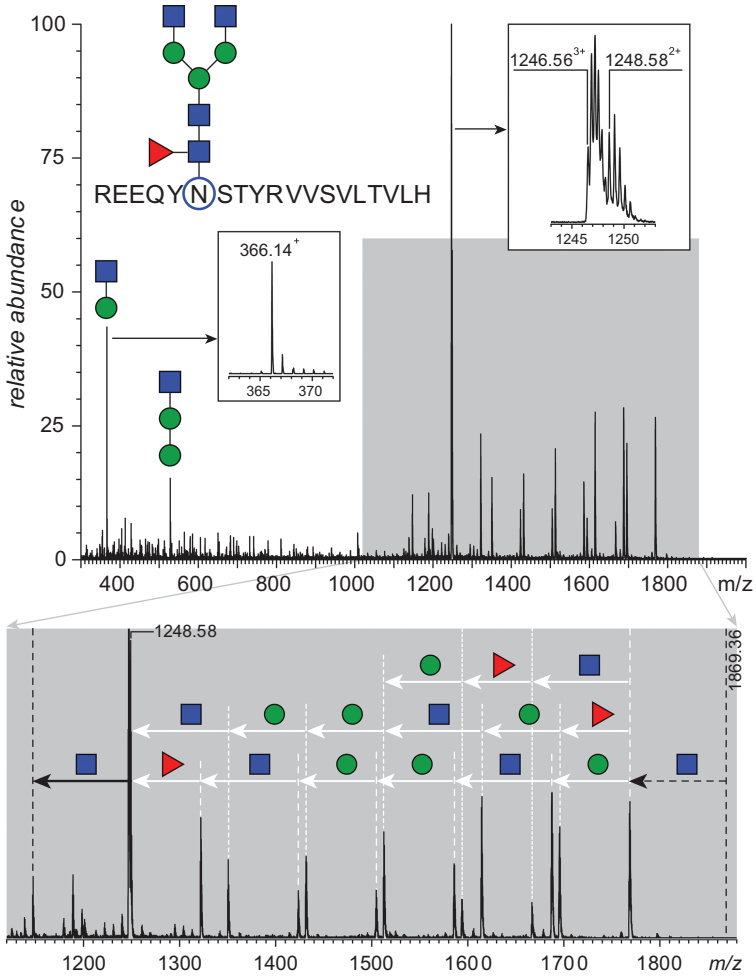


Figure 3.10: Fragment ion (CAD) mass spectrum of a proteolytic peptide fragment of mAb reveals the presence of a carbohydrate chain (whose facile fragmentation gives rise to a series of abundant fragment ions at *m/z* above 1,300), but is not informative vis-à-vis the amino acid sequence (the sequence information can be obtained by using electron-based ion fragmentation techniques, such as ECD or ETD). Data courtesy of Dr. Rinat R. Abzalimov (CUNY Advanced Science Research Center).

method has shown the potential for this task due to its excellent LC resolution and MS sensitivity. For example, application of an online strong cation exchange chromatography-native MS method has successfully detected an atypical glycosylation site in the Fab region (NSG motif) of NISTmAb reference material (Figure 3.11), which is present at extremely low-abundance (< 0.1%). As this N-glycosylated species contains a sialic acid (G3FS), it can be readily separated from the main species by strong cation

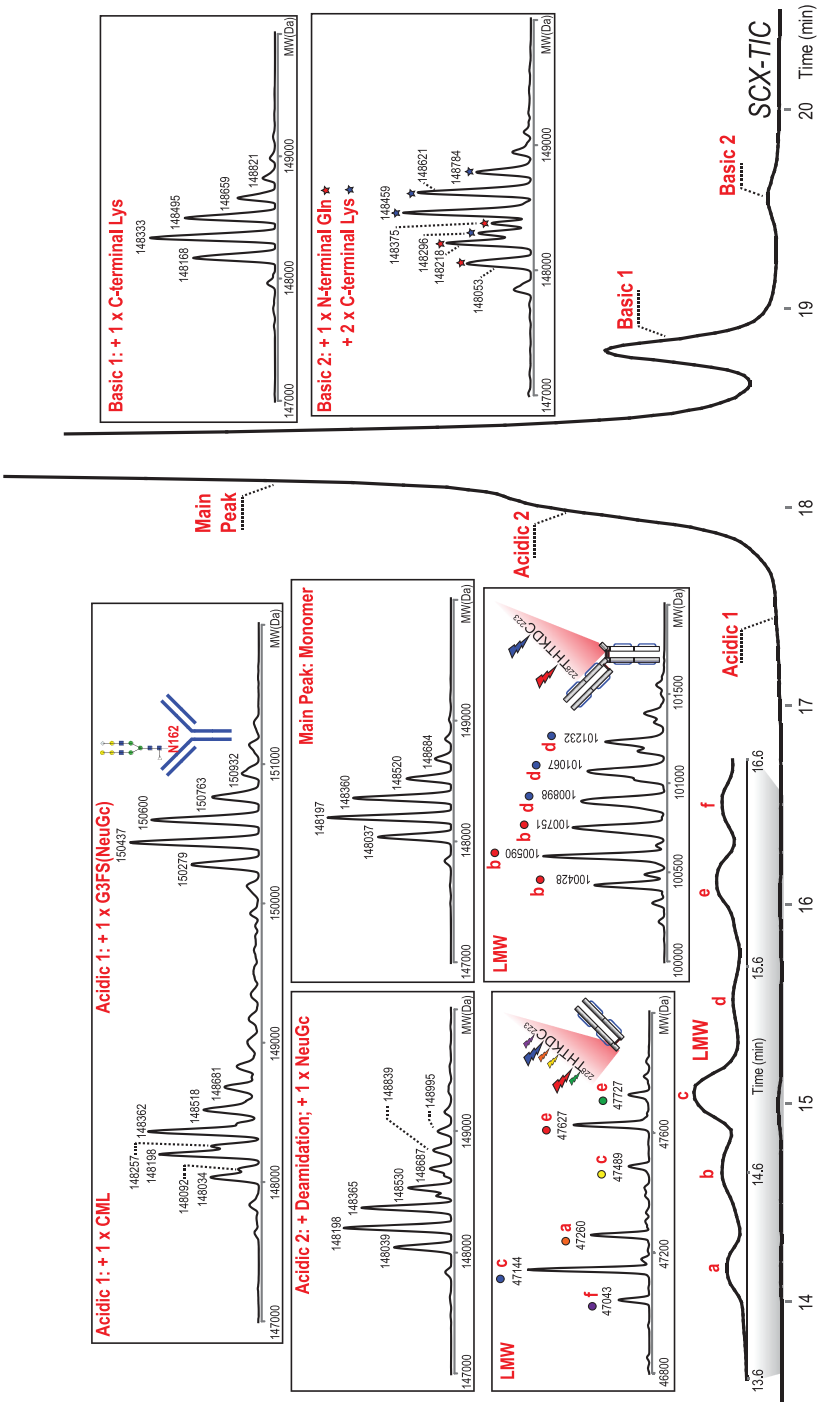


Figure 3.11: Identification of an atypical Fab-glycosylated species (in acidic 1 peak) in NISTmAb by native SCX-MS. Adapted with permission from Yan et al. [78] Copyright 2020 American Chemical Society.

exchange (SCX) LC due to the increased acidity, and therefore be readily detected by subsequent native MS analysis.

3.2.2 O-Glycosylation

In addition to N-glycans decorating the asparagine residues, carbohydrate chains can also be attached to the hydroxyl groups of serine and threonine side chains via an O-glycosidic bond to *N*-acetylgalactosamine residue (the so-called O-glycans, see Figure 3.8, bottom). Analysis of O-glycans has not been as straightforward as the analysis of N-glycans, largely due to the unavailability of enzymes capable of effectively releasing O-glycans. The existing chemical approaches to release O-glycans often involve complicated sample preparation and suffer from insufficient sensitivity, and thus are not ideally suited for routine characterization of protein biopharmaceuticals. By contrast, analysis of O-glycans at glycopeptide levels is more routinely adopted and the methodology is practically identical to the analysis of N-glycopeptides. O-glycopeptide analysis has benefited tremendously from recent advances in both chromatographic separation techniques and tandem MS fragmentation techniques. Just like N-glycans, O-glycans also contribute to the mass heterogeneity of protein biopharmaceuticals, and thus can significantly hinder the MS analysis at the intact protein level. To tackle the increased mass heterogeneity introduced by O-glycans, charge reduction is sometimes an effective technique to reduce spectral crowding due to the overlapping ionic signals at different charge states by shifting the entire charge state envelopes to a higher m/z region. A simple and elegant approach to achieve charge reduction in intact mass analysis is to modify the desolvation gas with the vapor from triethylamine, a commonly used charge stripping reagent [79] (Figure 3.12a). The effectiveness of this approach was demonstrated in a case study where a heavily O-glycosylated protein was successfully characterized. It is evident that, after charge reduction, the previously convoluted mass spectrum can be greatly simplified and readily deconvoluted, revealing both the number and the composition of the attached O-glycans (Figure 3.12b and c).

Overall, MS-based analysis of protein biopharmaceuticals containing high levels of glycans remains an area of active research efforts, as these molecules continue to present significant challenges to analytical scientists, requiring continuous innovation in order to meet the ever – increasing demands in this field.

3.2.3 Disulfide bonds and disulfide scrambling

To achieve desired therapeutic function and safety profile, the protein molecule should not only possess the designed amino acid sequence but also be correctly folded. Disulfide bonds, which are formed by oxidation of a pair of cysteine residues,

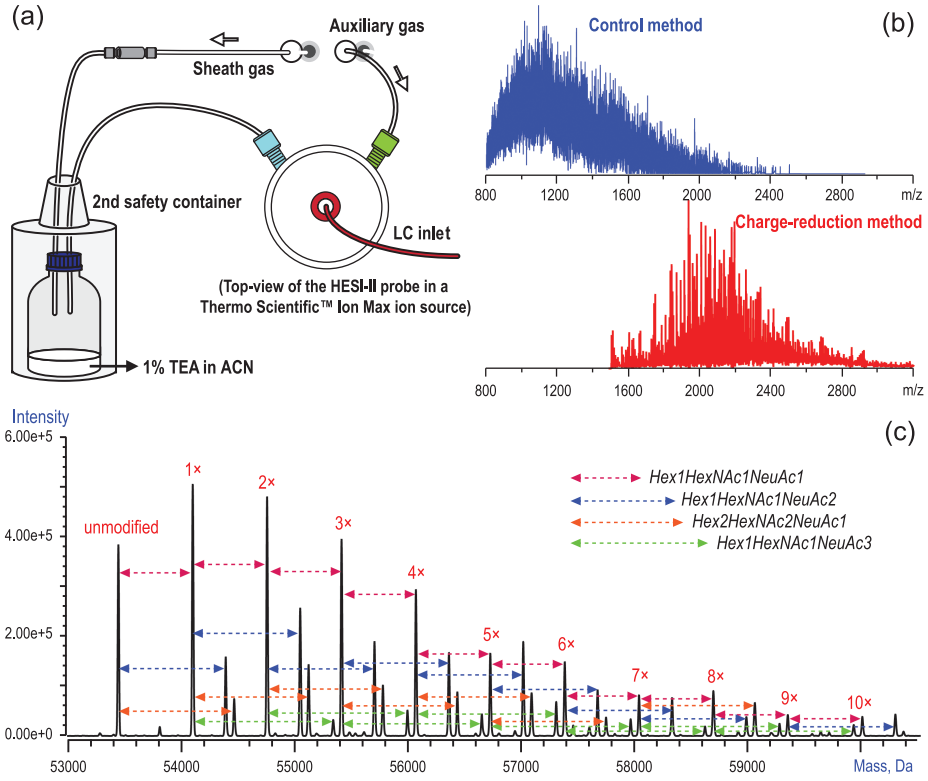


Figure 3.12: Application of the charge reduction strategy to facilitate intact mass analysis of a heavily O-glycosylated protein: (a) representation of modifying desolvation gas with TEA vapor in ESI source; (b) raw mass spectra comparison with and without charge reduction; and (c) deconvoluted mass spectrum under charge reduction condition. Adapted with permission from Wang et al. [79] Copyright 2019 American Chemical Society.

play a critical role in maintaining the protein higher order structure, and thus are critical not only for maintaining the protein drug biological functions, but also for its stability and safety. Indeed, aberrant formation of disulfide bonds has a negative impact on the protein conformational integrity (including both tertiary and quaternary structures), and can promote protein aggregation both directly (via formation of non-native external disulfides) and indirectly (by locking the protein in a non-native conformation). For protein biopharmaceuticals containing disulfide bonds, free cysteine residues, or both disulfides and free cysteines, it is necessary to confirm the redox status of each cysteine residue and the connectivity of each disulfide bond. Various MS-based approaches are now available to decipher the disulfide bond connectivity, most of which involve performing the peptide mapping analysis under non-reducing conditions, and comparing the results of its LC-MS/MS analysis with that for the reduced digests. Ionic species that disappear after the reduction most likely represent the

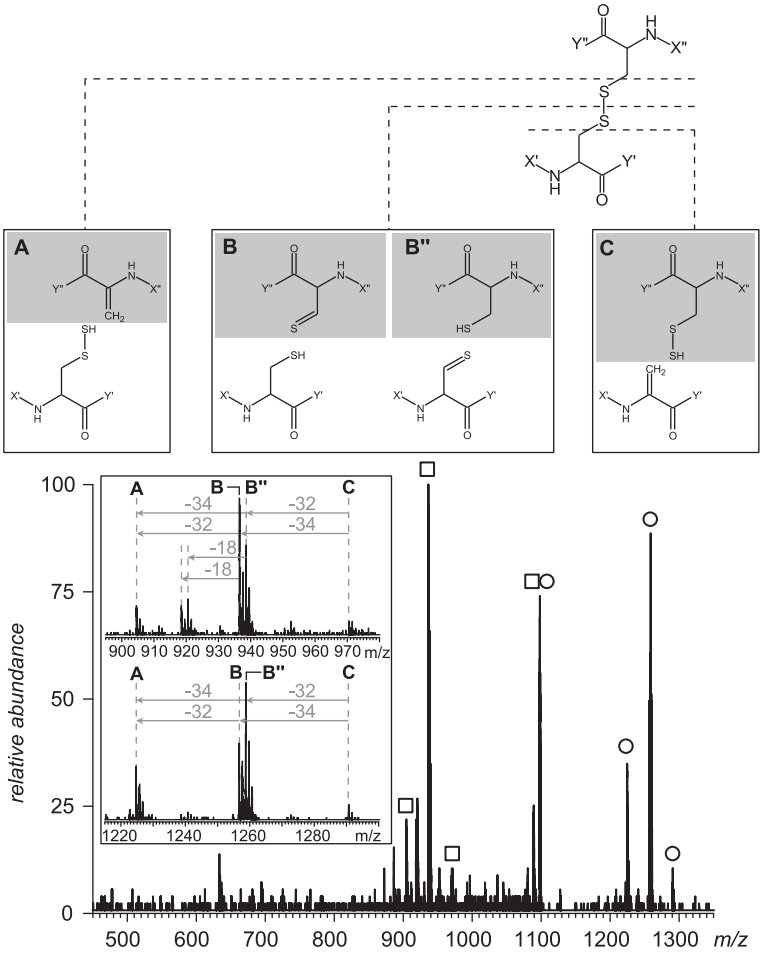


Figure 3.13: Top: A schematic diagram of a disulfide-linked peptide dimer dissociation in the negative ion mode. Bottom: A mass spectrum of fragment anions generated by CAD of a disulfide-linked peptide dimer [8 – 18]/[43 – 50] derived from a tryptic digest of a 37 kDa recombinant protein (N-lobe of human serum transferrin) obtained with a hybrid quadrupole/TOF mass spectrometer (CAD carried out in the RF-only quadrupole). Open squares and circles indicate ions corresponding to the intact peptide monomers produced by dissociation of the disulfide bond in the peptide dimer ion without any backbone cleavages (zoomed views are shown in the insets). Reproduced with permission from Zhang and Kaltashov [80] Copyright 2006 American Chemical Society.

disulfide bond-containing peptides, while the newly formed ions can be assigned to cysteine-containing peptides that are involved in disulfide bonds. Subsequently, the disulfide bond connectivity can be confirmed by correlating the two cysteine-containing peptides with the disulfide bond-containing peptide via a simple relationship:

$$\mathrm{Mass(DSB)} = \mathrm{Mass^{*}(Cys1)} + \mathrm{Mass^{*}(Cys2)} - 2.01565$$

where the masses of the two Cys-containing peptides (*) should be adjusted for the presence of the cysteine-capping reagents if applied. Although effective, this approach can be time-consuming, especially if the protein contains many disulfide bonds, as the number of possible combinations of two cysteine-containing peptides can grow quite significantly. An elegant solution to this problem is to perform online reduction of disulfide bond-containing peptides during the LC-MS analysis, as the correlation between disulfide bond-containing peptide and two Cys-containing peptides can be readily established by the identical retention time. For example, online partial reduction of disulfide bonds during reversed-phase LC-MS analysis can be easily achieved by post-column addition of TCEP, a highly effective reducing agent even at low pH [81, 82]. However, the ionic nature of TCEP will inevitably contribute to the ionization suppression of the peptide analytes during the electrospray process and can also contaminate the mass spectrometer. A promising alternative to the chemical reduction is application of the electrochemical reduction, which can be implemented using ESI-compatible flow cells equipped with electrodes that act as electron donors [83, 84].

As protein biopharmaceuticals are subjected to extensive purification processes, they typically consist of a single protein of high purity. Therefore, identification of disulfide bond-containing peptides can often be achieved using MS1 data alone, particularly when a high-resolution MS instrument is utilized. In addition, many data processing software packages greatly facilitate such work and increase the analysis throughput. To improve the identification, tandem MS methods utilizing various gas-phase fragmentation techniques could also be used. For example, in addition to accurate mass measurement at MS1 level, several matching *b*- and *y*-ions produced by collisional activation of the disulfide bond-containing peptide ions can significantly increase the identification confidence. This is particularly useful for detection and identification of scrambled disulfide bonds that are paired incorrectly and typically present at low levels.

Unlike conventional (positive ion) CAD, which mostly leads to the backbone amide bond fission in the gas phase, collisional activation of negative ions frequently results in facile cleavage of amino acid side chains, including those of cysteine residues without affecting the integrity of the peptide backbone [85].When applied to disulfide-linked peptide dimer ions, negative ion CAD results in separation of the two monomeric peptides by generating highly specific cleavages affecting all three bonds in the linkage segment between the β-carbon atoms of the two cysteine residues (Figure 3.13, top). This leads to appearance of easy-to-recognize signature fragment ions in the negative ion CAD mass spectra, which not only allows the disulfide-linked peptide

dimers (as well as higher oligomers) to be readily distinguished from monomeric precursor ions, but also enables identification of the disulfide-linked dimers based on the masses of their constituent peptide monomers (Figure 3.13, bottom). Since the negative ion CAD results mostly in side chain cleavages, scouting for disulfide-linked peptide dimers can be carried out in a broad-band mode, with wide m/z windows being used for selecting ionic populations for subsequent negative ion CAD (most of the fragment ions remain within the confines of the precursor ion windows, while the disulfide fission products fall outside of such windows) [80]. Other ion fragmentation techniques, such as those relying both on electron-based (ECD and ETD) and UV-photon based (UVPD) ion activation methods, can also be used for disulfide bond characterization, as they enable fission of disulfide bonds in the gas phase [86–88].

The strategies mentioned in the preceding paragraphs may not work very well in the case of proteins (or protein segments) with high density of cysteine residues. Close spacing of cysteine residues in the protein sequence makes it difficult (and in some cases impossible) to select a proteolytic enzyme that produces peptide fragments containing not more than a single cysteine residue. This is illustrated in Figure 3.14, which shows an amino acid sequence of recombinant β-glucocerebrosidase with four cysteine residues located within the 23 residue-long N-terminal segment of the protein (and two of which are separated by a single valine residue). No definitive conclusions vis-à-vis disulfide connectivity patterns can be reached when a single protease is used for peptide mapping of this protein. In fact, the complete separation of all cysteine residues by proteolysis is impossible even when multiple proteases are used (e.g., Lys-C to produce a (1–7)/(8–74) disulfide-linked dimer followed by its deglycosylation and digestion with Asp-N to produce a smaller dimer (1–7)/(8–23) and a trimer (1–7)/(8–18)/(19–23)). However, supplementing the proteolysis with tandem MS measurements provides the data sufficient to establish the disulfide pattern: the presence of distinct fragments produced by the amide bond cleavages within the $Cys^{16}Val^{17}Cys^{18}$ segment in the CAD spectrum of the peptide dimer confirms the presence of Cys^{4}–Cys^{16} and Cys^{18}–Cys^{23} disulfides. Lastly, it should be mentioned that identification of a particular pattern of disulfide connectivity within a therapeutic protein does not necessarily rule out the presence of scrambled disulfide species in the sample. Detection and identification of disulfide-scrambled species in stressed biopharmaceuticals (particularly if they are present at low abundance) presents an elevated challenge. Very frequently disulfide scrambling leads to significant conformational changes, allowing the scrambled species to be separated from the protein with the native disulfide connectivity using SEC. However, scrambling within cysteine-rich segments (similar to that shown in Figure 3.14) may lead to small-scale changes that cannot be resolved by SEC. Detection and identification of the disulfide-scrambled species in such cases is a challenging and demanding task that usually combines peptide mapping and MS/MS analysis, and frequently relies on multiple orthogonal-action proteases and several complementary ion fragmentation techniques.

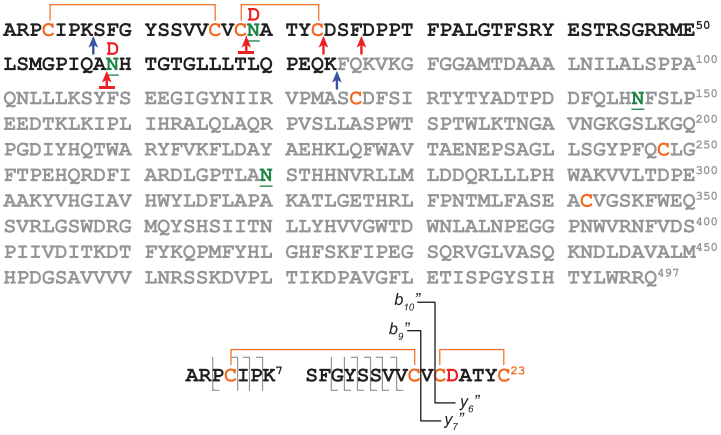


Figure 3.14: Top: the amino acid sequence of β-glucocerebrosidase with the Lys-C and Asp-N cleavage sites in the N-terminal segment of the protein identified with blue and red arrows, respectively. The proteolytic peptide dimer L1L2 (linked by a disulfide bond) is highlighted in bold. The "D" symbols placed above the L1L2 sequence identify the sites of Asn-to-Asp conversion as a result of its deglycosylation with PNGase F. Bottom: the primary structure of the proteolytic peptide fragment produced by digestion of the enzymatically de-glycosylated L1L2 peptide dimer with Asp-N and its CAD fragmentation pattern.

Besides forming disulfide bonds, cysteine residues can also be present in a reduced form, i.e. containing free thiol groups. Unintended free thiols, which are normally involved in disulfide bond formation, can be problematic for protein biopharmaceuticals, as they might lead to compromised structural integrity and biological function. Although the total free thiol levels in protein samples can be readily assessed by some colorimetric assays [89], site-specific quantitation of free thiol frequently relies on MS-based approaches. In particular, a differential alkylation workflow utilizing stable-isotope-labeled alkylation reagent is frequently applied [89]. Briefly, the protein sample is first alkylated using an unlabeled reagent under denaturing conditions so that all free thiols become solvent-accessible and be quantitatively labeled. Subsequently, after complete reduction of disulfide bonds, the newly formed free thiols can be further alkylated with a stable isotope-labeled reagent. Following trypsin digestion and LC-MS analysis, the free thiol level of each Cys residue can be readily calculated based on the isotope profile of the Cys-containing peptide. In addition to several commercially available, isotope-labeled alkylation reagents, preparation of ^{18}O-labeled iodoacetic acid is an easy and affordable option [90].

During the production of protein biopharmaceuticals, the cysteine residues are also susceptible to other post-translational modifications. For example, cysteine side chains can often be modified by cysteinylation and glutathionylation via a thiol-disulfide exchange mechanism, due to the presence of free cysteine and glutathione (GSH) in cell media [91]. Furthermore, free cysteine residues and disulfide bonds are

also prone to conversion to dehydroalanine via the β-elimination mechanism, particularly under elevated temperatures [92]. Such modifications on their own may not be considered critical quality attributes, but they can significantly impact other attributes of protein biopharmaceuticals. For example, those modifications, when occurring within the disulfide bond connecting the light and heavy chains of a therapeutic mAb, can lead to the dissociation of a free light chain from the rest of the mAb molecule (leaving behind a complementary H2L trimer) during purity assessment under denaturing conditions (e.g., SDS-PAGE or CE-SDS), and thus resulting in the elevated levels of lower molecular weight (LMW) impurities. A trisulfide bond, which is formed by insertion of an extra sulfur atom into a disulfide bond, is another peculiar PTM first observed in therapeutic mAbs [93]. The trisulfide bond is not considered one of the critical quality attributes, as it rarely has a negative impact on the drug safety and efficacy. Several studies have shown that trisulfide bonds can be quickly converted back to disulfide bonds during in vivo circulation, and thus have a negligible effect on the PK and PD properties of the protein therapeutic [93]. However, the presence of trisulfide bonds inevitably increases the heterogeneity of already highly complex protein drugs, and thus they are still closely monitored during the drug production. Lastly, the sulfhydryl group of a free cysteine residue present in a therapeutic protein can become a target of oxidation [94], a non-enzymatic PTM that will be considered in detail in Section 3.3.

Some analytical methods developed for mapping disulfide bonds in proteins, such as differential alkylation [95], can also be utilized for probing the higher order structure of proteins (the subject matter of Chapter 4) by evaluating solvent accessibility of individual disulfide bonds. For example, partial reduction followed by differential alkylation can be used to rank the reduction susceptibility of disulfide bonds residing in different domains of an IgG1 molecule [96]. A buried disulfide bond has increased susceptibility to reduction when it becomes more solvent accessible as a result of a local unfolding event. This feature could be utilized to investigate the structural changes of protein biopharmaceuticals under certain stress conditions (e.g., high temperature, agitation and exposure to UV light).

3.2.4 Other enzymatic PTMs

Besides glycosylation and disulfide formation, other enzymatic PTMs are also found in protein biopharmaceuticals. For example, during the production of therapeutic mAbs, certain carboxypeptidases (which are also expressed and secreted into the cell culture media) remove the C-terminal lysine residue from mAb heavy chains [98]. Incomplete processing of a mAb molecule by carboxypeptidases gives rise to sequence heterogeneity within the C-termini of the heavy chains (up to one "extra" lysine residue per chain or up to two "extra" residues per a mAb molecule). Although unprocessed C-terminal lysine is typically not considered a critical quality attribute (it is unlikely to affect the drug safety

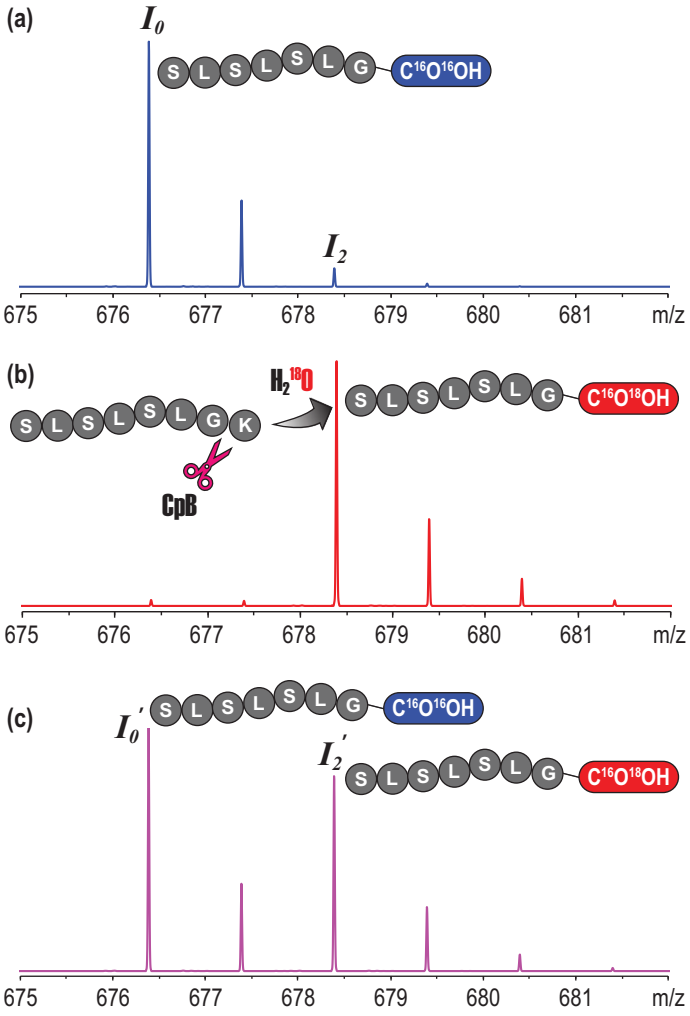


Figure 3.15: MS1 spectra of (a) IgG4 C-terminal peptide without Lys, (b) Lys-containing C-terminal peptide treated with CpB in ^{18}O-enriched buffer, and (c) a 0.9:1.1 mixture of C-terminal peptides with and without Lys treated with CpB in ^{18}O-enriched buffer. Adapted with permission from Wang et al. [97] Copyright 2020 American Chemical Society.

and efficacy), it does significantly impact the charge heterogeneity profile of the drug, which is indicative of the process and product consistency. Moreover, recent reports support the notion of the charge properties of therapeutic mAbs affecting their binding to Fc receptors as well as the tissue distribution and pharmacokinetics, highlighting the importance of well-characterized and controlled charge heterogeneity of mAb products [99–101]. The levels of C-terminal lysine residues in mAb products are frequently quantitated by LC-MS based peptide mapping analysis following tryptic

digestion, as the C-terminal peptides with and without lysine residue have different masses and can often be effectively separated by reversed-phase chromatography. It is generally accepted that the relative abundance of PTMs can be calculated based on the MS responses of both modified and unmodified peptides, assuming their ionization efficiencies are comparable. However, if a PTM significantly changes the peptide ionization efficiency, either directly by changing its charging characteristics or indirectly due to the changes in the solvent condition or co-eluting species as a result of a dramatically altered retention time, MS-based quantitation will be less reliable. In this case, the lysine-containing C-terminal peptide often exhibits significantly higher ionization efficiency compared to the C-terminal peptide without a lysine residue, leading to overestimation of the C-terminal lysine levels. To overcome this problem, one can use an ^{18}O-labeling based strategy [97]. Prior to the LC-MS analysis, the tryptic digests are first treated with carboxypeptidase B in ^{18}O-enriched water to quantitatively convert the lysine-containing C-terminal peptide to C-terminal peptide without Lys and simultaneously label it with one ^{18}O atom (Figure 3.15b). Subsequently, the change in isotope distribution of the resulting C-terminal peptide can be utilized to back-calculate the percentage of the C-terminal lysine (Figure 3.15c).

Hydroxylysine (Hyl) is another example of enzymatic PTM that is less commonly observed in protein biopharmaceuticals. It features a hydroxyl residue attached to the δ-carbon of a lysine residue side chain by an enzyme lysyl hydroxylase that apparently recognizes a Xaa-Lys-Gly consensus sequence [102]. This modification has been identified in several therapeutic proteins, including Activase® (rtPA), a soluble form of CD4 receptor (rCD4), and in therapeutic mAbs [103], presumably due to the presence of endogenous lysyl hydroxylase in cell culture media. Lysine hydroxylation leads to an increase of the monoisotopic mass by 15.9949 Da, which is identical to the mass increase resulting from oxidation of many other amino acid residues (e.g., Met, His, and Trp) or sequence substitutions (e.g., Ala to Ser, and Phe to Tyr). As a result, confirmation of Lys hydroxylation usually requires tandem MS analysis to localize the modification to a specific Lys residue [103].

3.3 Characterization of non-enzymatic PTMs

Above and beyond modifications introduced into biopharmaceutical products by enzymatic processing, PTMs can also arise due to chemical degradation during the lengthy production process (which includes protein expression, purification and formulation), and storage. Unlike the enzymatic PTMs, which require the presence of active enzymes, non-enzymatic PTMs can be introduced into a protein therapeutic by chemical reactions with a variety of molecules that it comes in contact with, including water, air, inorganic and organic salts, nutrients (such as glucose), transition metal ions, sanitizing agents (such as hydrogen peroxide), excipients, and dissolved components of containers. Some of these chemical reactions are spontaneous,

while others can be catalyzed by a range of physical stimuli, such as UV or visible irradiation, mechanical shaking, and temperature spikes. The presence of such PTMs often increases the heterogeneity of a protein therapeutic, and in some cases may affect its safety and efficacy profiles. Therefore, it is important to identify and quantitate non-enzymatic PTMs that are present in a biopharmaceutical product and assess their impact on the drug safety and efficacy. For the past two decades, MS has become an indispensable tool in PTM characterization of protein biopharmaceuticals, with various MS-based strategies and workflows developed to detect and identify a wide range of non-enzymatic PTMs.

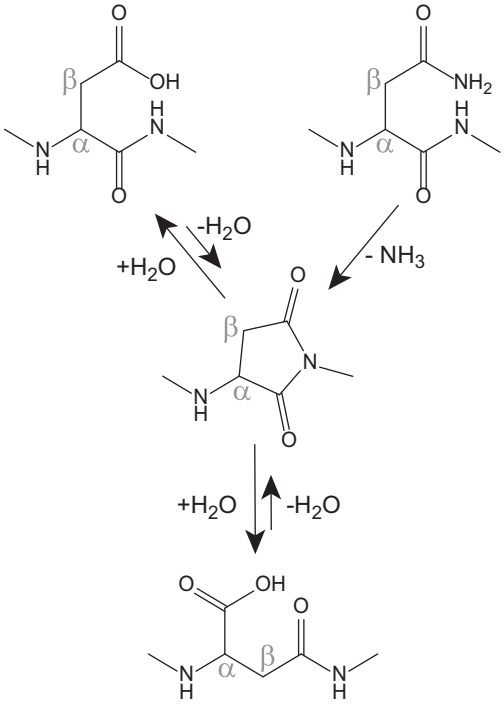


Figure 3.16: Isomerization of aspartic acid (top left) and deamidation of asparagine (top right) to *iso*-aspartic acid (bottom) via a succinimide intermediate (middle). Adapted with permission from Cournoyer et al. [80].

3.3.1 Asparagine deamidation and aspartic acid isomerization

Deamidation of asparagine and isomerization of aspartic acid residues are two of the most commonly occurring non-enzymatic PTMs in protein therapeutics. Under typical production and storage conditions (near-neutral pH), both deamidation and

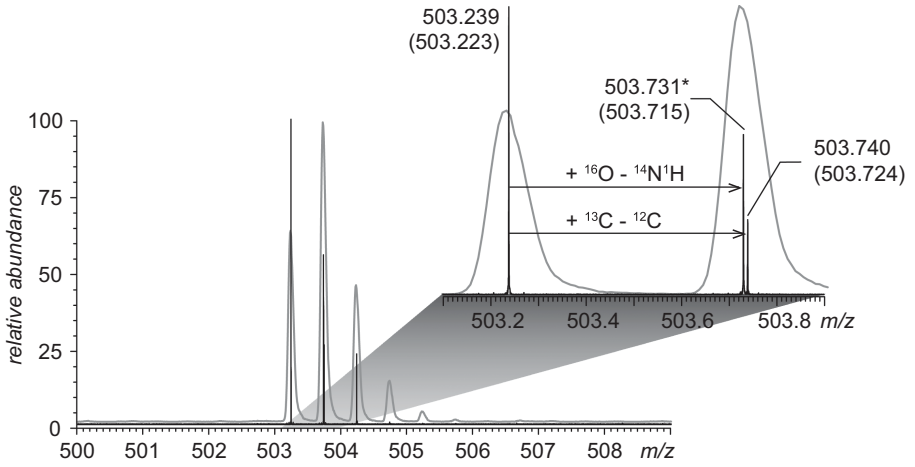


Figure 3.17: ESI mass spectra of a peptide SWANGDEAR acquired with a hybrid quadrupole/TOF MS (gray trace) and an FT ICR MS (black trace). The sample contained both intact and deamidated forms of the peptide. The inset shows a zoomed view of the two lowest mass isotopic peaks. Only FT ICR MS has the resolving power that is sufficient to directly detect the presence of the deamidated form of this peptide (marked with an asterisk).

isomerization proceed through formation of a succinimide intermediate and subsequently result in a mixture of Asp- and *iso*Asp-containing products (Figure 3.16). Asparagine deamidation usually increases the abundance of acidic variants of the drug molecule, as it introduces an additional negative charge. Formation of the *iso*-Asp residue elongates the protein backbone by inserting an extra methylene group, resulting in the β-peptide linkage. In some cases, when occurring at critical locations, asparagine deamidation and aspartic acid isomerization can compromise the efficacy and stability of the drug molecule [104–106], and even trigger undesired immunogenic responses [107].

Asparagine deamidation results in a rather small mass change (a 0.98403 Da increase of the monoisotopic mass), which makes it difficult to detect deamidation products via mass measurements alone even within short peptides (Figure 3.17). No mass change is associated with the aspartic acid isomerization. Therefore, detection and quantitation of these PTMs must necessarily include a chromatographic step, which is usually coupled with online MS or MS/MS detection. The deamidation products within small to mid-size biopharmaceuticals can be detected in many cases using a top-down approach, as illustrated in Figure 3.18 for a stressed form of interferon-β1a. In this example, an online ion exchange LC-MS analysis of the protein sample reveals the presence of two (nearly) isobaric species for each glycoform of the protein, consistent with deamidation of the protein (which is expected to result in a significant change of the elution time, while the *m*/*z* shift magnitude would be too small to allow the presence of the deamidated forms of the protein to be

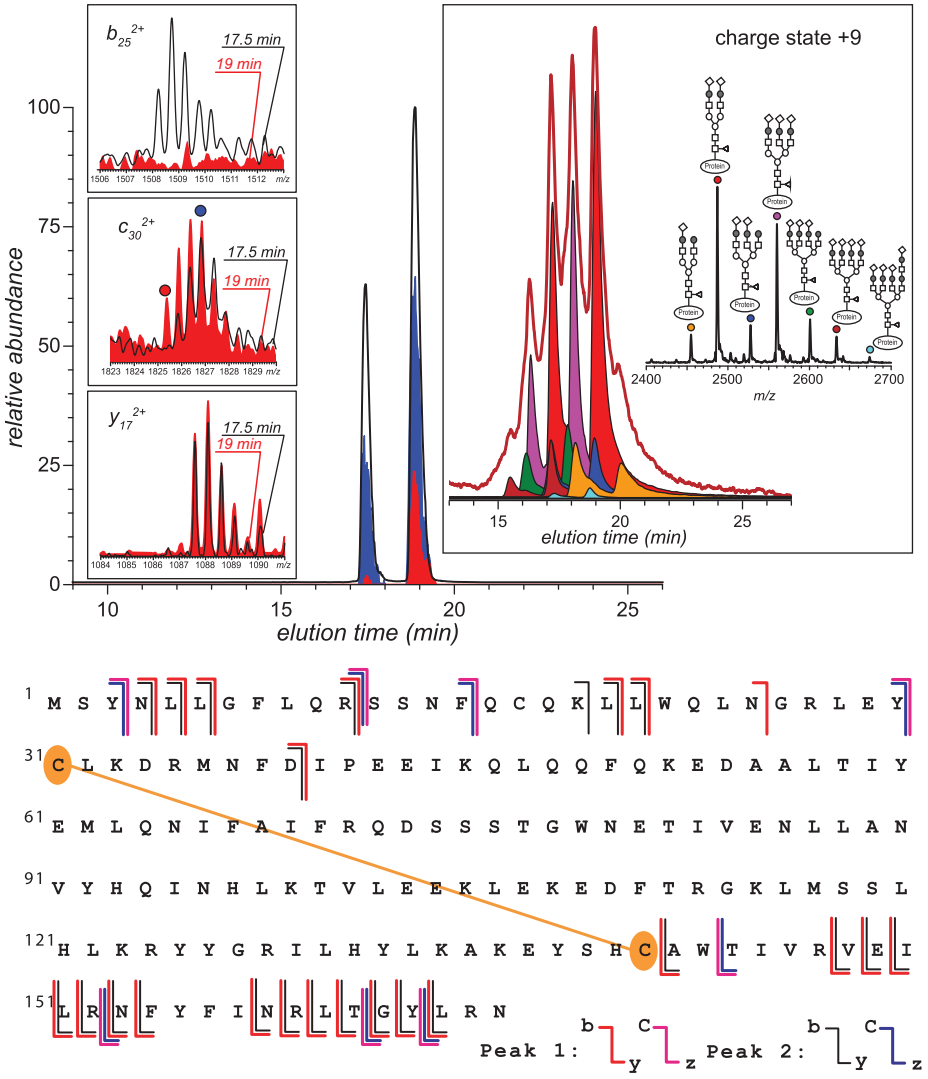


Figure 3.18: Top-down detection and localization of deamidation within stressed interferon-β. Top right: ion exchange LC-MS analysis of stressed protein sample: TIC chromatogram and extracted ion chromatograms of seven different glycoforms identified in the mass spectrum of the protein sample not subjected to separation (shown in inset). Top left: identification and localization of a stress-induced deamidation within the BiNA2 species of interferon-β by ion exchange LC-MS/MS. TIC for all fragment ions produced by ETD and collisional activation of precursor ions corresponding to the + 9 charge state of the BiNA2 species (black trace) and extracted ion chromatograms corresponding to two different isotopic peaks within the c_{30}^{2+} fragment ion (blue, m/z 1,826.9; red, m/z 1,825.4). The insets on the left show isotopic distributions of fragment ions that are observed in both chromatographic peaks (c_{30}^{2+} and y_{17}^{2+}) and unique to the earlier-eluting species (b_{25}^{2+}). The inset on the right shows overall fragmentation patterns for the two chromatographic peaks. Bottom:

detected by MS alone). Online ion exchange LC-MS/MS analysis of the two nearly isobaric species in each glycoform produces nearly identical fragmentation patterns, although some important differences are apparent (Figure 3.18). For example, in the entire series of *b*-ions derived from the thirty amino acid long N-terminal segment of the protein, only fragments produced by cleavage of the backbone amide bonds following Asn-25 show a clear difference in their isotopic distributions indicative of the Asn/Asp conversion (a shift by one mass unit). No differences were detected between the two sets of *y*-ions derived from the protein C-terminal segment. This allows the deamidation site to be confidently identified as Asn-25.

A significant advantage offered by the top-down analysis as a means of identifying deamidation sites within protein therapeutics is the minimal sample workup, which eliminates the possibility of introducing this PTM during the analysis (which remains a distinct possibility when the bottom-up approaches are used, vide infra). However, both the increased mass and the presence of multiple disulfide bonds in larger protein therapeutics (e.g., mAbs) make the top-down approach impractical; therefore, detection, localization, and quantitation of both asparagine deamidation and aspartic acid isomerization are usually achieved by a variety of bottom-up based approaches (e.g., peptide mapping). Trypsin is often used in these studies, although care must be taken to minimize the digestion-induced deamidation or isomerization [108], which obviously alters the protein PTM profile. It is because the conditions (e.g., basic pH and elevated temperature) used for a typical tryptic digestion also facilitate the formation of those two PTMs via the succinimide pathway (the most susceptible degradation sites are Asn and Asp residues followed by Gly or Ser residues). Sample preparation-induced artifacts not only lead to overestimation of those PTMs, but also make quantitative comparison between samples difficult due to a high-level baseline. To reduce the artifact formation, protein digestion should be carried out by trypsin at slightly lower pH (e.g., pH 6.0) [109] or using endoprotease Glu-C at pH 4.5 [110]. Alternatively, strategies have also been developed to distinguish the inherent deamidation from artifacts by performing sample preparation in ^{18}O-enriched water [111]. During such experiments, inherent deamidation only results in a mass increase of 1 Da, while the deamidation artifacts (Asp and *iso*Asp) from sample preparation give rise to a mass increase of 3 Da due to the incorporation of one ^{18}O atom during the hydrolysis of succinimide. Although the 2 Da molecular weight difference is sufficient to differentiate the two deamidated species (the "real" PTM and the artifact), quantitative analysis using this approach can be challenging due to the complication from trypsin-catalyzed ^{18}O-labeling at peptide C-terminal carboxyl groups, resulting

Figure 3.18 (continued)
Amino acid sequence of interferon-β showing all fragment ions detected in the online top-down analysis of the two chromatographically separated BiNA2 species. Adapted with permission from Muneeruddin et al. [45] Copyright 2015 American Chemical Society.

in convoluted isotopic distributions of Asp- and *iso*Asp-containing peptides. One possible solution involves the use of *b*-ions from the MS/MS spectra, as *b*-ions do not contain the ^{18}O-labeled C-terminus, thus greatly simplifying the mass spectra [112]. Besides leading to overestimation of overall deamidation or isomerization levels, sample preparation can also alter the distribution of different degradation products. In particular, succinimide intermediate is often stable within the intact protein, but can be readily hydrolyzed to aspartic or iso-aspartic acid during its trypsin digestion. It has been shown that the succinimide modification within a mAb molecule (originating from Asp isomerization) can be preserved using a high-fidelity trypsin digestion protocol, which involves a low pH (pH 5) denaturation step and a short digestion time (1 h) [113]. However, it is likely that the stability of the succinimide intermediate might vary depending on the higher order structure and the peptide sequences, and therefore optimization of the sample preparation conditions might be required on a case-by-case basis.

Finally, for in-depth characterization of protein deamidation, the Asp/*iso*Asp differentiation in degraded peptides is often required. Historically, this task was approached by both Edman degradation method [115] and some HPLC methods that relied on different retention times of Asp- and *iso*Asp-containing peptides [116]. However, those traditional approaches often suffer from relatively low throughput and less reliable assignment and have gradually been replaced by MS-based approaches. For example, the electron-based fragmentation techniques (electron capture dissociation, ECD, and electron transfer dissociation, ETD) have been widely adopted for differentiation between Asp- and *iso*Asp-containing peptides, where a pair of reporter ions ($c\bullet$ +57 and z −57) that are unique to *iso*Asp-containing peptides and another pair of reporter ions (z −44 and $(M + nH)^{(n-1)+}$ −60) that are unique to Asp-containing peptides can be generated [117–119]. However, this approach requires access to MS instrumentation with ECD or ETD capability and significant efforts might still be required to achieve the extent of fragmentation in MS/MS measurements sufficient for isomeric differentiation. An elegant alternative is to apply an ^{18}O-labeling based strategy to produce deamidated peptide standards followed by LC-MS analysis [114]. Taking advantages of the different deamidation mechanisms, forced deamidation of peptides can be carried out in ^{18}O-enriched solution under both acidic and basic conditions, where isomer-specific mass tags are introduced to ^{18}O-labeled Asp- and *iso*Asp-containing peptides (Figure 3.19). Subsequently, the differentiation can be readily made based on the isotope-labeled mass tags at MS1 level, as well as the retention time profiles.

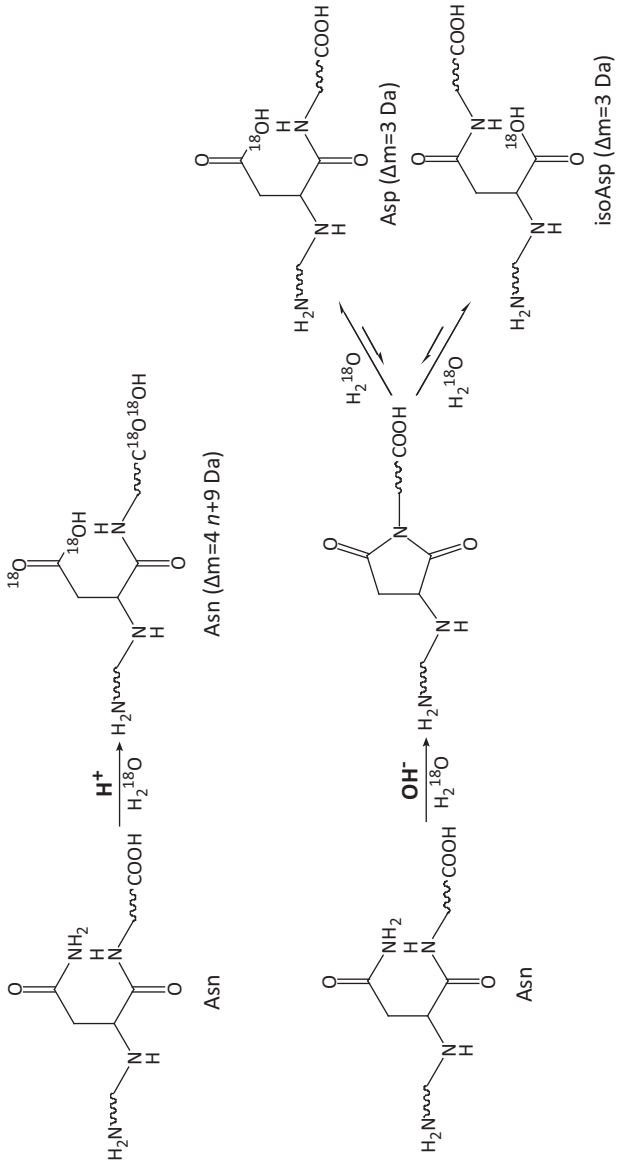


Figure 3.19: Forced deamidation of Asn-containing peptides in ^{18}O-enriched solution under both acidic and basic conditions to introduce isomer-specific mass tags. *n* is the number of acidic residues in the peptides. Adapted with permission from Wang et al. Copyright 2013 American Chemical Society.

3.3.2 Oxidation

Oxidation of protein biopharmaceuticals is another commonly observed non-enzymatic PTM that in some cases may have significant structural and biological consequences. Due to the high reactivity of sulfur atoms and aromatic rings with respect to various oxidants, the side chains of methionine (Met), free cysteine (Cys), and, to a lesser extent, histidine (His), phenylalanine (Phe), tryptophan (Trp), and tyrosine (Tyr) residues are susceptible to oxidation (Figure 3.20). During the drug production and storage, the presence of reactive impurities, such as peroxides and transition metal ions, as well as certain stress conditions can all contribute to accelerated protein oxidation. It is also possible that oxidation of protein therapeutics occurs in vivo (post-administration), as both H_2O_2 and HOCl are produced by the immune system (e.g., upon bacterial infection neutrophils produce H_2O_2 and chloride ions, and an enzyme myeloperoxidase catalyzes the generation of the hypochlorous acid) [120].

Methionine is the most common target of oxidation in protein therapeutics, and the occurrence of this PTM can result in both conformational changes [121] (affecting stability) and alteration of the functional properties. For example, the methionine residues in the Fc regions of mAbs are located in close proximity to the Fc receptor binding sites, and their oxidation has a detrimental effect on the FcRn binding, adversely affecting the lifetime of the antibody in circulation. To better understand the oxidation pathways and the hotspots, protein drugs are often purposely exposed to different stress conditions for extended periods of time (e.g., light, high temperature, hydrogen peroxide, and extreme pH), and then analyzed to identify and quantitate site-specific oxidation. Characterization of oxidized peptides by LC-MS is relatively straightforward, as oxidized peptides typically elute earlier than the native peptides in reversed-phase LC due to the increased hydrophilicity and exhibit readily detectable mass shifts (Figure 3.20).

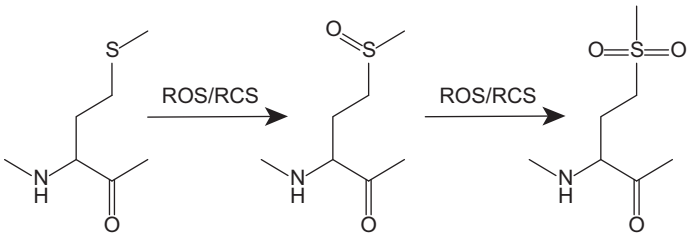


Figure 3.20: Methionine can be oxidized to methionine sulfoxide by reactive oxygen species (ROS) and reactive chlorine species (RCS). High concentrations of these oxidants can further (and irreversibly) oxidize methionine sulfoxide to methionine sulfone. Adapted from Drazik and Winter [120] with permission from Elsevier.

Quantitation of the extent of oxidation can be carried out by comparing the ionic signal abundance for the proteolytic peptides containing unmodified methionine residues

and methionine sulfoxide in LC-MS, although both the hydrophobicity change and the difference in the mobile-phase composition (due to the different elution times for the two forms of the peptide) might contribute to the discrepancy in ionization efficiency. Furthermore, similar to the analysis of deamidation, caution should be taken in order to reduce (or indeed completely eliminate) the sample preparation-induced oxidation artifacts. For instance, it has been reported that such artificial methionine oxidation can be minimized by use of antioxidants (e.g., L-methionine) and chelating agents (e.g., EDTA) during the sample preparation and LC-MS analysis [122]. Alternatively, reliable determination of methionine oxidation can also be achieved using an ^{18}O-labeling assisted strategy, where the protein samples are first fully oxidized using ^{18}O-enriched hydrogen peroxide before tryptic digestion and LC-MS analysis [123]. Complete oxidation of methionine by ^{18}O-enriched hydrogen peroxide prevents further oxidation during sample preparation, thus allowing the quantitation of pre-existing methionine oxidation.

Methionine (and, to a lesser extent, free cysteine) side chains are considered the primary targets for the oxidative damage in protein therapeutics. However, several other amino acid residues can be affected by this PTM type, including tryptophan, tyrosine, phenylalanine [124]; tryptophan oxidation within the CDR regions of mAbs in particular has been attracting significant attention due to its negative impact on the antigen affinity [125]. Methods of detection and localization of these PTMs are very similar to those commonly employed for detecting methionine oxidation and typically include both intact mass measurements and peptide mapping with

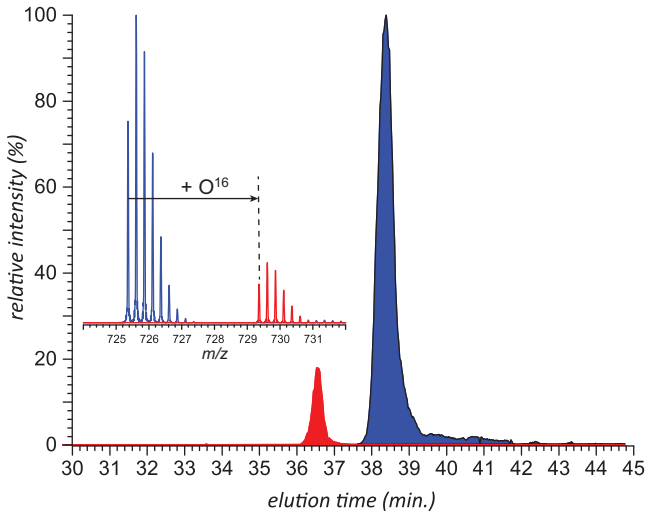


Figure 3.21: Extracted ion chromatograms of a methionine-containing Lys-C peptide fragment derived from the Fc region of a mAb (blue) and its oxidized form (red). The inset shows the mass spectra averaged across each chromatographic peak. Data courtesy of Dr. Jake W. Pawlowski (Amgen, Inc.).

LC-MS/MS [126]. In addition to modifying the amino acid side chains, oxidation can also lead to covalent cross-linking of two residues, giving rise to covalent dimer species if the cross-linked residues reside in two different molecules. For example, it is well known that both the oxidative stress in vitro and the peroxidase-catalyzed mechanism in vivo can promote the covalent cross-linking between two tyrosine (Tyr) residues resulting in the formation of a dityrosine modification [127, 128]. More recently, studies have shown that histidine oxidation can also initiate novel cross-links to other histidine, lysine or cysteine residues via a nucleophilic addition mechanism [129]. Characterization of such cross-links frequently requires a combination of appropriate proteolysis to produce cross-linked peptides and the application of various tandem MS techniques to localize the cross-linking sites [130].

3.3.3 Lysine and N-terminal amine glycation

Biopharmaceutical products are also susceptible to a process known as non-enzymatic glycation, where the reducing sugars present in cell media react with the primary amine groups (e.g., lysine side chains and the N-terminus) of the protein to form Schiff bases and subsequently convert to more stable ketoamine products via Amadori rearrangement (Figure 3.22). Just as other PTMs, when occurring at critical sites, glycation could impact the efficacy, stability and safety profiles of the drug molecule. Moreover, because glycation predominantly modifies the lysine side chains of the protein, it can significantly impact the protein charge heterogeneity as determined by a cation exchange chromatography (CEX) method. Although lysine glycation only leads to a minimal change in the p*K*a of a lysine side chain (p*K*a of 10.6 for primary amine vs. p*K*a of 10.8 for secondary amine), it often significantly reduces the retention of a cationic protein on a CEX column due to the impaired ability of the affected lysine residue to bind to the ligand.

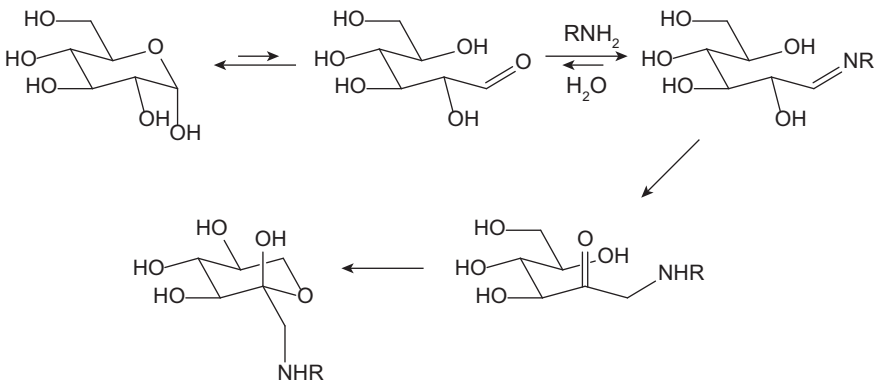


Figure 3.22: A schematic diagram of the primary amine glycation.

Detection, localization, and quantitation of protein glycation present several unique challenges when compared to other non-enzymatic PTMs. First, the mass increase associated with glycation is identical to that caused by extending an *N*- or *O*-glycan chain with a single hexose residue (see Sections 3.2.1 and 3.2.2). Therefore, attempts to detect the presence of glycation in glycoproteins – which constitute the majority of biopharmaceutical products – using intact mass measurements are bound to fail, unless enzymatic deglycosylation is used to remove the N-glycans prior to glycation analysis [132] or middle-up approaches are employed to separate glycan-free protein segments from those containing N- and O-glycosylation sites [131], as illustrated in Figure 3.23 for the assessment of a mAb glycation. In general, confident detection of glycation is possible only at the peptide level (i.e., detection and localization steps must be combined). Second, glycation is often distributed among many lysine residues throughout the entire sequence. As a result, the modification level at any individual site is typically low and may be difficult to detect by conventional peptide mapping methods. Enrichment of glycated proteins and/or peptides with boronate affinity chromatography prior to MS analysis is a common strategy to improve the detectability of low levels of glycation present in protein drugs [133]. The affinity enrichment is made possible by the highly specific and reversible bonding between the tetrahedral anions of boronic acid at alkaline pH and *cis*-1,2-diol groups from the sugar structures in glycated protein and/or peptides.

Glycation is a labile modification under CAD conditions: the conventional methods of collisional activation typically give rise to tandem mass spectra that are dominated by fragment ions produced by neutral losses from the carbohydrate moiety and containing little peptide backbone fragmentation information [134]. This problem can be readily overcome by using electron-based ion fragmentation techniques (ETD or ECD), as illustrated in Figure 3.24 for glycated peptides subjected to ECD. Collisional activation methods can also be used to localize glycation sites within peptide ions following chemical derivatization to enhance the gas-phase stability of the saccharide unit. For example, sodium borohydride ($NaBH_4$) can be used to reduce the double bond present in the Amadori moiety to a single bond, thus greatly stabilizing the carbohydrate structure during the gas-phase fragmentation [134]. Derivatization can be combined with stable isotope labeling to assist in identification of glycated peptides: reduction of glycated peptides with borodeuteride (BD_4^-) results in placing a deuterium atom within the reduced peptide [135]. Comparing the LC-MS datasets for the $NaBD_4$ and $NaBH_4$-treated samples, the 1 Da mass difference can be utilized to identify glycated peptides (Figure 3.25).

Facile identification of potential glycation hotspots within a protein primary sequence can be achieved by forced glycation studies, where the protein drugs are incubated with relatively high concentrations of glucose for an extended period of time followed by LC-MS/MS analysis. This type of study can also be facilitated by using a mixture of regular glucose and $^{13}C_6$-labeled glucose for the incubation [137]. In this case, identification of the glycated peptides is carried out by searching the mass spectra

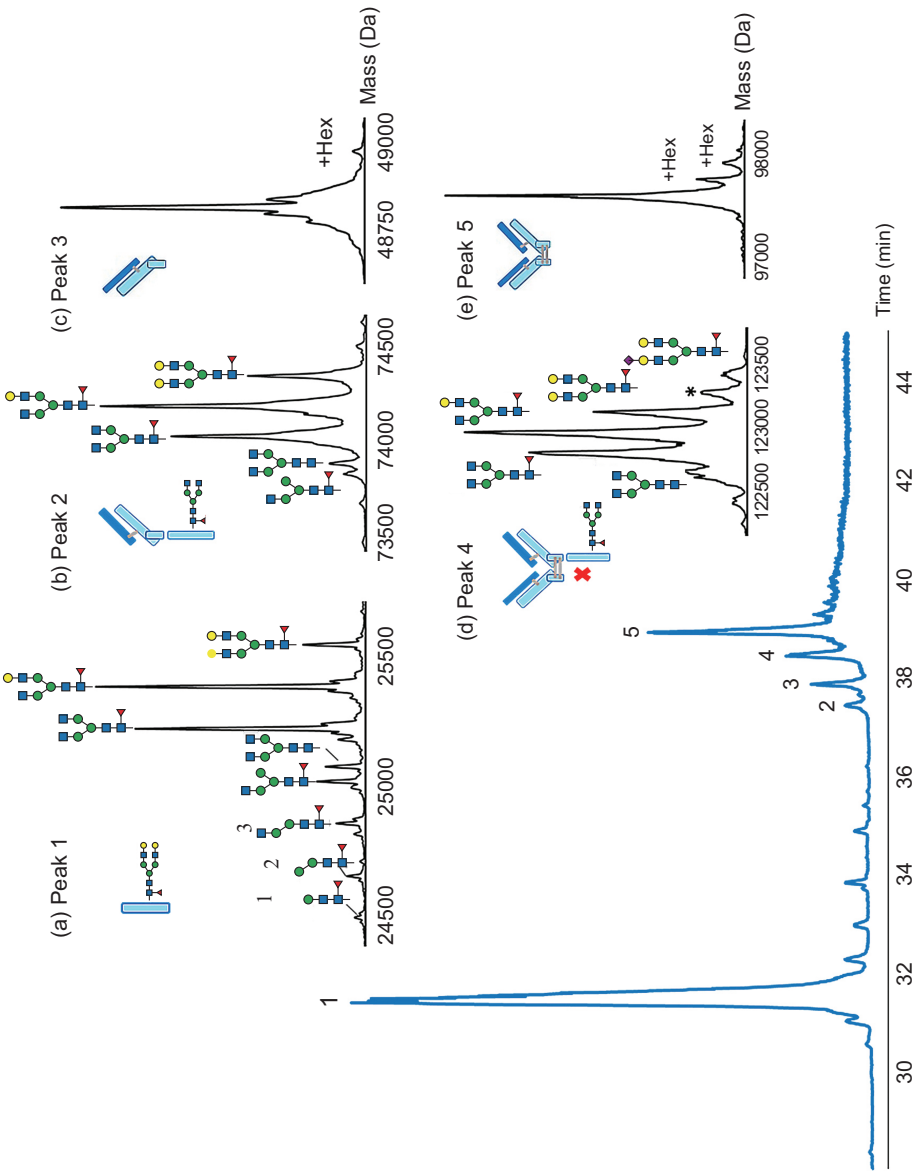


Figure 3.23: Middle-up analysis of mAb glycation using an online IdeS-digestion with an enzyme-immobilized column, followed by reversed-phase LC-MS analysis of the mAb subunits: total ion current chromatogram (the blue trace) and deconvoluted mass spectra of Fc/2 (*a*), half-mAb (*b*), Fab' (*c*), mAb with a missing Fc/2 subunit (*d*), and F(ab')$_2$ (*e*). The MS peaks are labeled with the corresponding N-glycan structures; glycation modifications are represented by (+Hex). Reproduced from Camperi et al. [131] with permission from Elsevier.

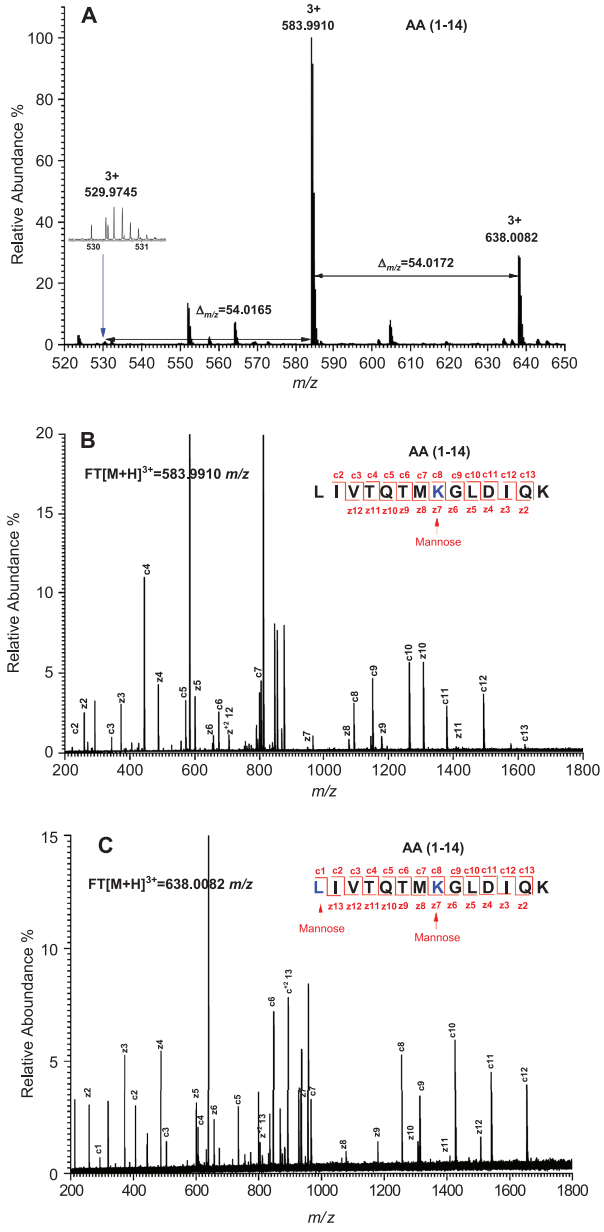


Figure 3.24: Mass spectrum of the peptide 1–14 (A) and ECD MS/MS spectra of the glycated peptide 1–14 with m/z of 583.9910^{3+} (B) and 638.0082^{3+} (C) from the glycated β-lactoglobulin after ultrasound pretreatment at 400 W. The m/z differences are indicated with numbers and arrows. The sequence of the peptide is shown on the top of figure. The determined glycation site (blue bold) is shown by an arrow with a "mannose" label. Reproduced with permission from Yang et al. [136]. Copyright 2017 American Chemical Society.

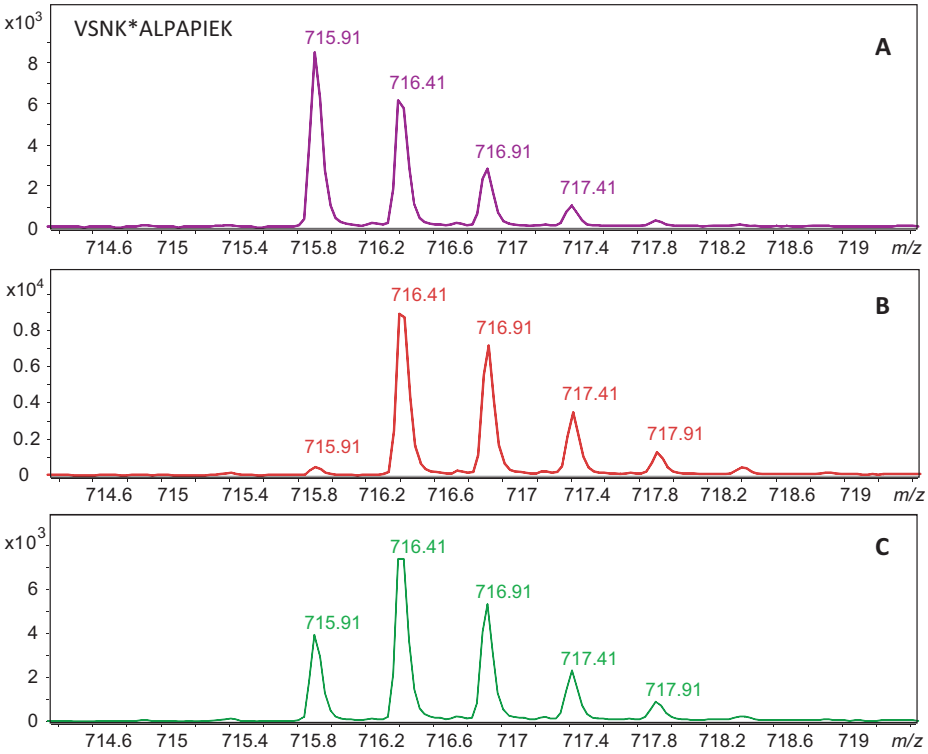


Figure 3.25: Mass spectra of a glycated peptide VSNKALPAPIEK reduced with $NaBH_4$ (*A*), $NaBD_4$ (*B*), and a 1:1 mixture of the samples reduced either with $NaBH_4$ or $NaBD_4$ (*C*). Reproduced from Liu et al. [135] with permission from Elsevier.

for pairs of ions whose monoisotopic masses differ by 6.018 Da. Finally, it is worth noting that protein glycation can also proceed to form a cohort of heterogeneous advanced glycation end products (AGEs) [138]. For example, carboxymethylated lysine (CML) is a well-studied AGE that is mainly formed by the oxidative degradation of glycated lysine residues [139]. The formation of CML introduces an additional carboxylic acid group, thus increasing the levels of acidic variants of the drug molecule. The complexity associated with protein glycation as well as the possibility of AGEs being present has made MS the method of choice for detection and quantitation of those PTMs due to its capability to multiplex the measurements with minimal inter-analyte interference [140].

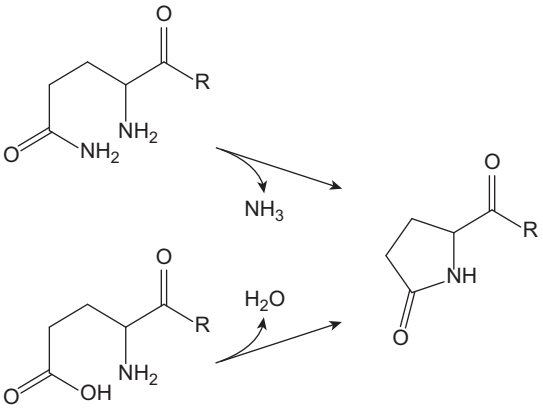


Figure 3.26: Conversion of N-terminal glutamine and glutamic acid residues to pyroglutamate.

3.3.4 N-Terminal glutamate to pyroglutamate conversion

Pyroglutamate (pGlu) is a non-enzymatic PTM that arises due to the spontaneous conversion of an N-terminal glutamine or glutamic acid residue to a pyrrolidinone ring (Figure 3.26). Since the glutamine and/or glutamic acid residues frequently occupy the first position in amino acid sequences of both light and heavy chains of IgG molecules, this PTM is common in mAbs [141], and in some instances near-complete conversion of the N-terminal residues to pyroglutamate may occur [142]. Glutamate conversion to pyroglutamate is slower compared to the glutamine to pyroglutamate conversion, but nevertheless has been observed under near-physiological conditions [143]. Pyroglutamate formation is known to occur endogenously, and therefore usually is not considered to be a critical quality attribute [144]. However, the glutamine-to-pyroglutamate conversion affects the distribution of charge variants, as it reduces the total positive charge of the protein by transforming the primary amine at the protein N-terminus to a γ-lactam nitrogen (the glutamic acid-to-pyroglutamate conversion does not change the charge balance at neutral pH), and is typically included in the list of PTMs that are closely monitored during the production of protein therapeutics. The presence of pyro-glutamate residues within the protein molecules can be revealed by intact mass measurements (by observing a characteristic loss of 17 or 18 Da, depending on the residue being converted to pyroglutamate) [145, 146]. The definitive detection and quantitation of pyroglutamate residues is usually carried out by peptide mapping and LC-MS/MS analysis of the N-terminal peptide [145, 146]. One must be mindful, however, of possible artifacts (including pyroglutamate formation both in-source during MS analysis and in solution just prior to it [147]), which obviously result in overestimating the occurrence of this PTM.

3.3.5 Protein backbone cleavage

The majority of PTMs considered so far concerned chemical modifications of the amino acid side chains; however, the amide bonds of the protein backbone can also be targeted by chemical reactions. Absent proteolytic enzymes, the protein backbone amide linkages remain stable under physiological conditions, but certain sites may become prone to hydrolysis, which can be facilitated by the presence of specific side chains, flexibility of the local structure, action of transition metals and radicals, as well as solution pH and temperature variations [148]. The hinge region of mAbs is known to be particularly labile with respect to the non-enzymatic backbone cleavage [149]. It is also noteworthy that the backbone integrity of recombinant proteins can be compromised by host cell proteases throughout the production process [150]. Backbone fragmentation is a critical quality attribute that must be closely monitored to assess both the purity and the integrity of the protein therapeutic [148]. The occurrence of a backbone cleavage can be readily detected by intact mass measurements [151]; the only caveat is that fission of a backbone amide bond in disulfide-containing proteins does not necessarily result in the physical separation of the two complementary fragments. Indeed, if the products of the backbone cleavage remain connected to each other via thiol–thiol linkage(s), the fission product's monoisotopic mass will exceed that of the intact polypeptide by 18.0106 Da (an equivalent of incorporation of an H_2O molecule upon hydrolysis of the amide bond) and can be easily mistaken for an oxidation event. Therefore, intact mass measurements should always be preceded by disulfide reduction in order to minimize errors in detecting the backbone fragmentation events.

3.4 "Designer" PTMs (chemical conjugation products) and methods of their characterization

As the name implies, designer PTMs are a class of modifications introduced into protein therapeutics at the drug design stage with the purpose of either enhancing their pharmacokinetic profiles (e.g., protein drug PEGylation) or endowing them with novel therapeutic properties (e.g., a cytotoxin conjugation to a carrier protein). Although production of the majority of modern biopharmaceuticals incorporating designer PTMs relies on high-fidelity conjugation protocols, the complexity of the protein covalent structure (and in particular the presence of multiple sites that can be targeted by the conjugation reactions) may result in significant structural heterogeneity exhibited by the final products. This includes not only the presence of dead-ended products (resulting from the hydrolysis of the cross-linking groups) and variations in the total number of the successful conjugations per protein molecule, but also multiple positional isomers that differ from each other by the location of the conjugation sites within the protein sequence. The latter is particularly important to consider when evaluating the ability of the carrier protein to deliver its payload to the

intended site of action, as chemical modifications within the receptor binding interface are likely to be detrimental vis-à-vis receptor recognition by the protein drug.

3.4.1 Protein PEGylation and other protein–polymer conjugates

Conjugation of small-size protein therapeutics with polyethylene glycol chains (PEGylation) is usually done in order to achieve lower rates of renal clearance resulting in increased circulation lifetimes, and at least in some cases to lower incidence of immunogenic response leading to better safety profiles and enhanced resistance to proteolytic degradation [51]. Rapid proliferation of PEGylated biopharmaceuticals poses significant challenges vis-à-vis analytical characterization, as these products are highly complex molecules that exhibit structural heterogeneity at different levels (various protein isoforms may differ from each other by the number of PEG chains attached to a single polypeptide chain and location of the conjugation sites; the PEG chain polydispersity also contributes to the overall product heterogeneity) [152]. SEC allows the isoforms of the first type to be readily resolved in most cases due to the significant differences in their physical size in solution. The extent of heterogeneity associated with the PEG chain polydispersity can be visualized and evaluated by MALDI MS and at least in some cases by ESI MS (vide infra). It is the heterogeneity of the second type (protein–polymer conjugates differing in position of PEGylation sites within the

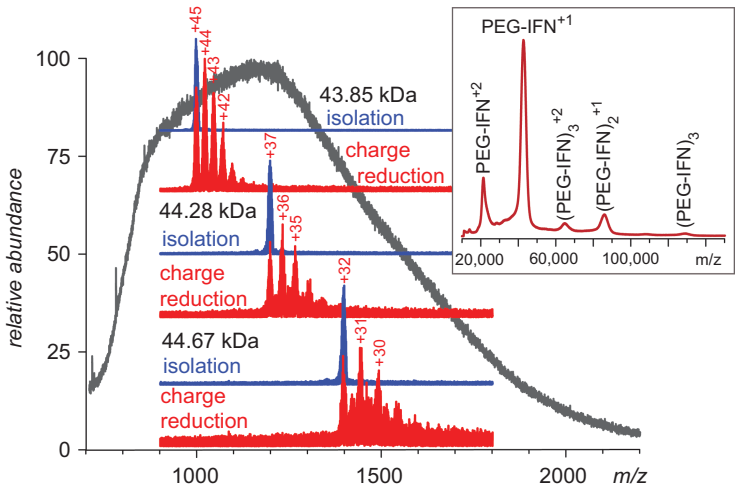


Figure 3.27: ESI mass spectrum of PEGylated interferon-β (PEG-IFNβ, black trace). Isolation of ionic populations within 10 *m*/*z* unit-wide selection windows (blue traces) followed by limited charge reduction induced by reacting with anions give rise to well-defined charge ladders (red traces). The inset shows a MALDI TOF mass spectrum of PEG-IFNβ. Reproduced with permission from Muneeruddin et al. [154]; permission conveyed through Copyright Clearance Center, Inc.

polypeptide sequence) that is usually most demanding in terms of the analytical efforts required for its evaluation. Over the past two decades, a range of MS-based methods of analysis/characterization of PEGylated proteins have been gaining popularity [153], and at present MS is a major tool supporting all stages of development and testing of PEGylated biopharmaceuticals [153].

PEGylation stands apart from other PTMs due to the high degree of polydispersity exhibited by the polymer chains, which gives rise to multiple ionic species with relatively close mass spacing (monoisotopic mass increments 44.0262 Da corresponding to CH_2CH_2O repeat units). Multiple charging in ESI MS results in significantly smaller m/z increments ($44/z$), and the mass spectra are further complicated by the overlap of ionic signals representing different charge states of PEGylated proteins [152]. This problem can be limited by using MALDI MS, which produces mostly singly charged ions [156], but the resolution of TOF analyzers in the high m/z region results in relatively modest mass accuracy. The signal overlap issue in ESI mass spectra can be addressed by adding strong bases (such as trimethylamine, TEA) to the protein solution, which results in charge stripping off the protein ions, pushing their signal to higher m/z and spreading the ionic species further apart from each other [157] (an approach also used in the glycoprotein analysis, as discussed in Section 3.2.2). Complementing ESI MS with ion mobility measurements allows further improvements to be achieved vis-à-vis resolving ionic species that have different masses/compositions but nearly identical m/z values [158]. Application of limited charge reduction to ionic populations within narrow m/z selection windows is another effective way to obtain mass and conjugation stoichiometry information on highly heterogeneous PEGylated biopharmaceutical products [154], as illustrated in Figure 3.27 for PEGylated interferon-β.

Localization of the PEGylation sites within the polypeptide sequence can be achieved using modified peptide mapping approaches, which combine chemical treatment of the polymer chain and proteolysis. The chemical treatment step is introduced into the workflow as a means of removing the PEG chain, and leaving in its place a chemical group that serves as a marker of the PEGylation site, which can be readily localized by examining the proteolytic fragments [159, 160]. In some cases, peptide mapping alone may provide information on the PEGylation sites within protein–polymer conjugates [161, 162], but incorporation of peptide ion fragmentation into the analytical workflow [163] or selective enrichment of PEGylated peptides prior to MS analysis [164] have been shown to be advantageous.

Top-down MS provides an alternative approach that can be used to identify the conjugation sites within PEGylated proteins. Effective schemes utilize a two-stage fragmentation workflow, with the first stage (that can be carried out in-source on the entire ionic population) aiming at eliminating the polymer chain while leaving a small “placeholder tag” at the conjugation site, which is identified/localized during the second fragmentation step (CAD or ECD applied to mass-selected precursor ions with truncated polymer chains) [152, 165]. This procedure is illustrated in Figure 3.28

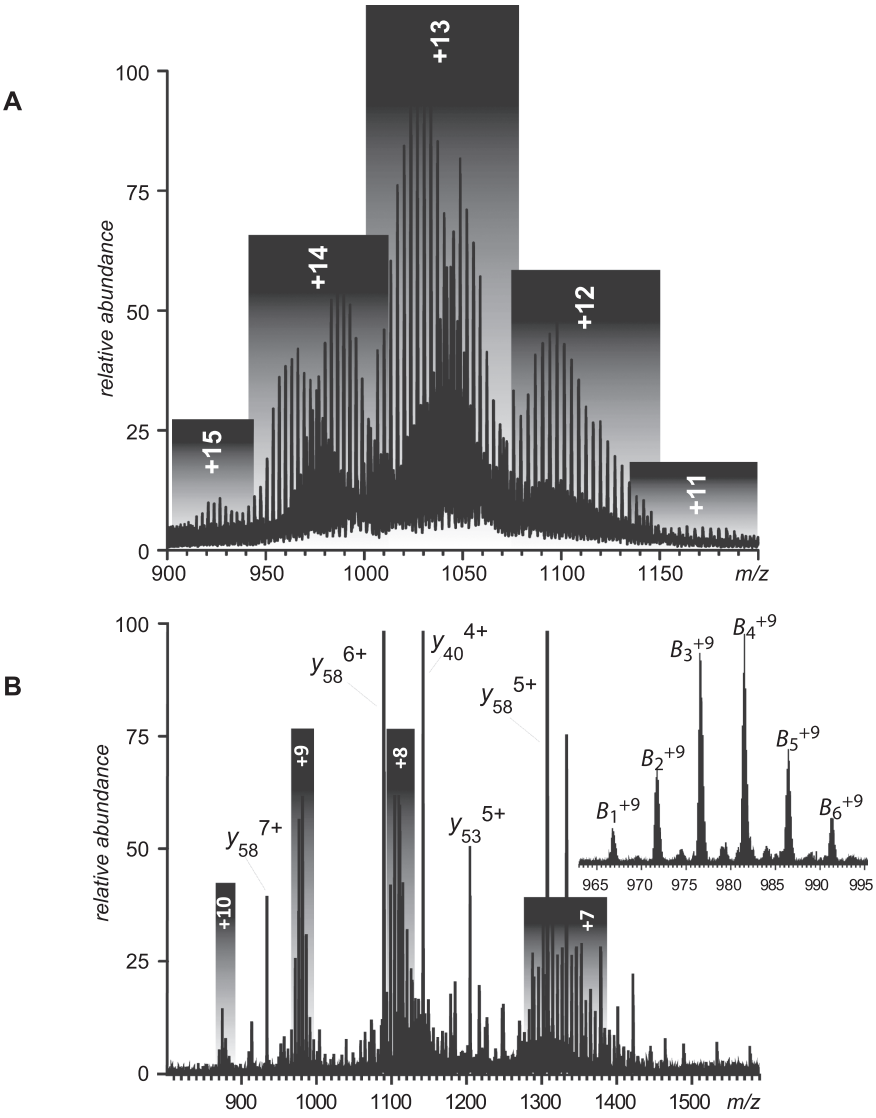


Figure 3.28: Localization of the PEGylation site using top-down MS. (*A*) An ESI mass spectrum (MS1) of the mono-PEGylated ubiquitin fraction. (*B*) Mass spectrum of fragment ions produced by collisional activation of PEGylated ubiquitin ions in the ESI interface region of a hybrid quadrupole/FT ICR MS. Clusters of fragment ions with truncated PEG chains are shaded and their charge states are indicated. The inset shows a zoomed view of one such cluster (charge state +9). (*C*) Fragment ions (second generation) produced by ECD of $B_3$9+, $B_4$9+, and $B_5$9+ ions (shown in panel *B*) in the ICR cell. (*D*) A schematic diagram explaining the designations of fragment ions in panels *B* and *C* (based on nomenclature introduced by Lattimer et al.[155]). Adapted from Abzalimov et al. [152] with permission from Elsevier.

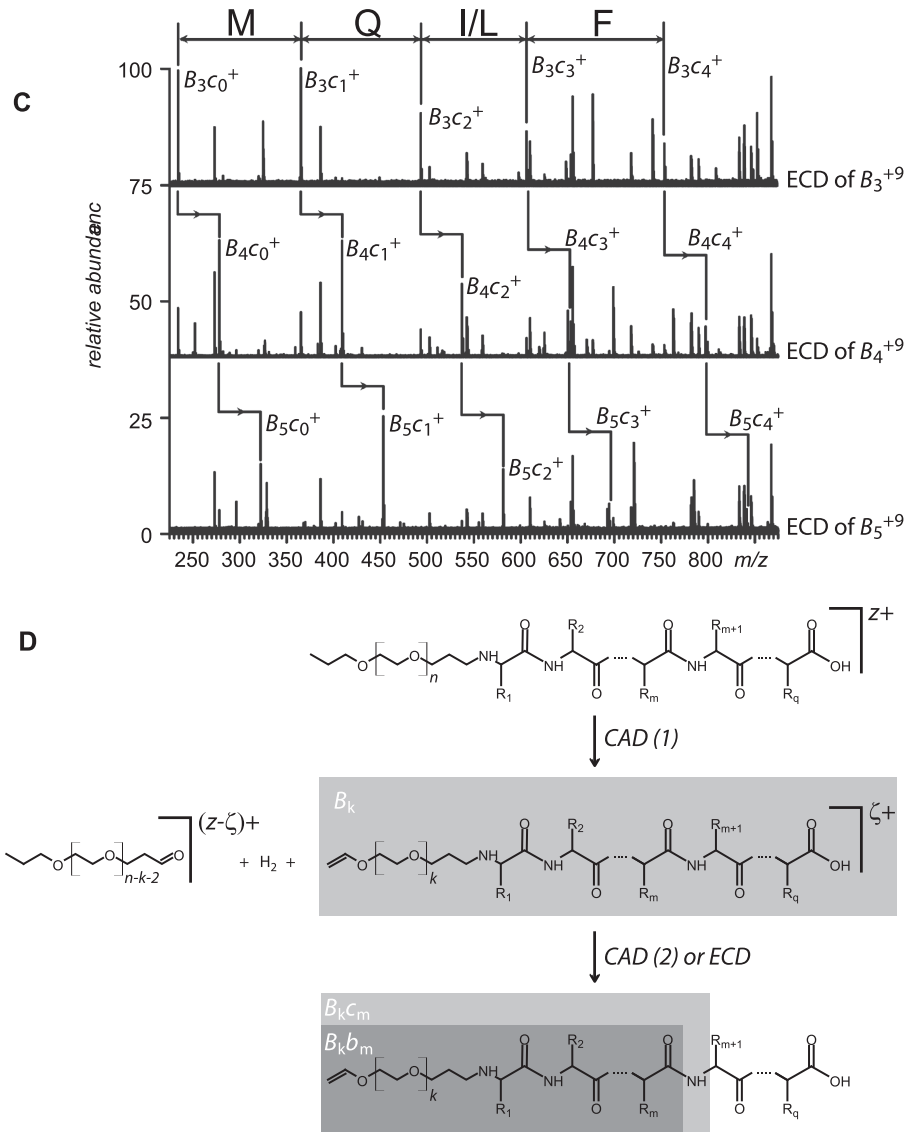


Figure 3.28 (continued)

where the top-down approach is used to identify the N-terminal primary amine as the conjugation site of mono-PEGylated ubiquitin. A combination of online ion exchange LC-MS with the methods of gas-phase chemistry (limited charge reduction and ion fragmentation) allows meaningful analyses to be carried out on complex protein therapeutics that incorporate enzymatic (glycosylation) and non-enzymatic (deamidation) PTMs in addition to PEGylation [154].

3.4.2 Protein–small-molecule drug conjugates

Chemical conjugation of small-molecule drugs to carrier proteins to enable targeted delivery had been considered for a variety of protein vectors, including hemoglobin, haptoglobin, and transferrin, although at present monoclonal antibodies are the only vector used in all approved therapeutics of this type. These biopharmaceutical products, commonly termed antibody–drug conjugates (ADCs), consist of a humanized or fully human mAb conjugated with highly cytotoxic small molecules (payloads) via chemical linkers. The ADC platform allows a potent cytotoxic payload to be delivered selectively to the target (usually neoplastic) cells, thereby improving efficacy and reducing systemic toxicity [166].

A large number of product quality attributes of ADCs relate to the mAb portion of the macromolecule (e.g., antigen binding, Fc function, aggregation, fragmentation, PTMs, higher order structure), while several unique ones describe the distribution of the payload. These include the drug/antibody ratio (DAR), the mass distribution profile (MDP) and the specific conjugation sites. While the first two attributes in this list may seem somewhat redundant (both DAR and MDP determine the cytotoxicity of the ADC product), MDP provides more detailed information by specifying the relative abundances of conjugates with specific stoichiometries, including the cytotoxin-free antibodies. MDP can be readily obtained from intact mass measurements of an ADC product, although the presence of various PTMs (e.g., N-glycosylation in particular) may complicate this task. Therefore, enzymatic deglycosylation is frequently performed prior to the MS analysis in order to simplify the MDP measurement (Figure 3.29). In addition to obtaining the distribution of mAb molecules with different number of payload groups attached, intact mass measurements of ADC provide other important information, such as the extent of antibody modifications with dead-ended linkers (with no payload attached) and the presence of cross-linked species.

Localization of the small-molecule drug conjugation site(s) within ADCs is typically achieved by peptide mapping [167]. While the workflow is very similar to that employed for peptide mapping of mAbs, certain modifications may be required, e.g., to improve the solubility of hydrophobic drug-conjugated peptides in solution, as well as to streamline identification of conjugated peptides by using payload fragments as “signature” ions in LC-MS/MS [168]. In some cases extensive peptide mapping with MS/MS analysis of peptide fragments can be substituted with faster and less labor-intensive middle-down approaches as a means of identifying the conjugation sites within ADCs [169].

Lastly, it should be mentioned that certain conjugation chemistries warrant close investigation of other PTM types considered earlier in this chapter. For example, utilizing thiol chemistry as a means of attaching the payload to the antibody (be it via partial reduction of the native disulfide bonds or engineering-in novel cysteine residues)

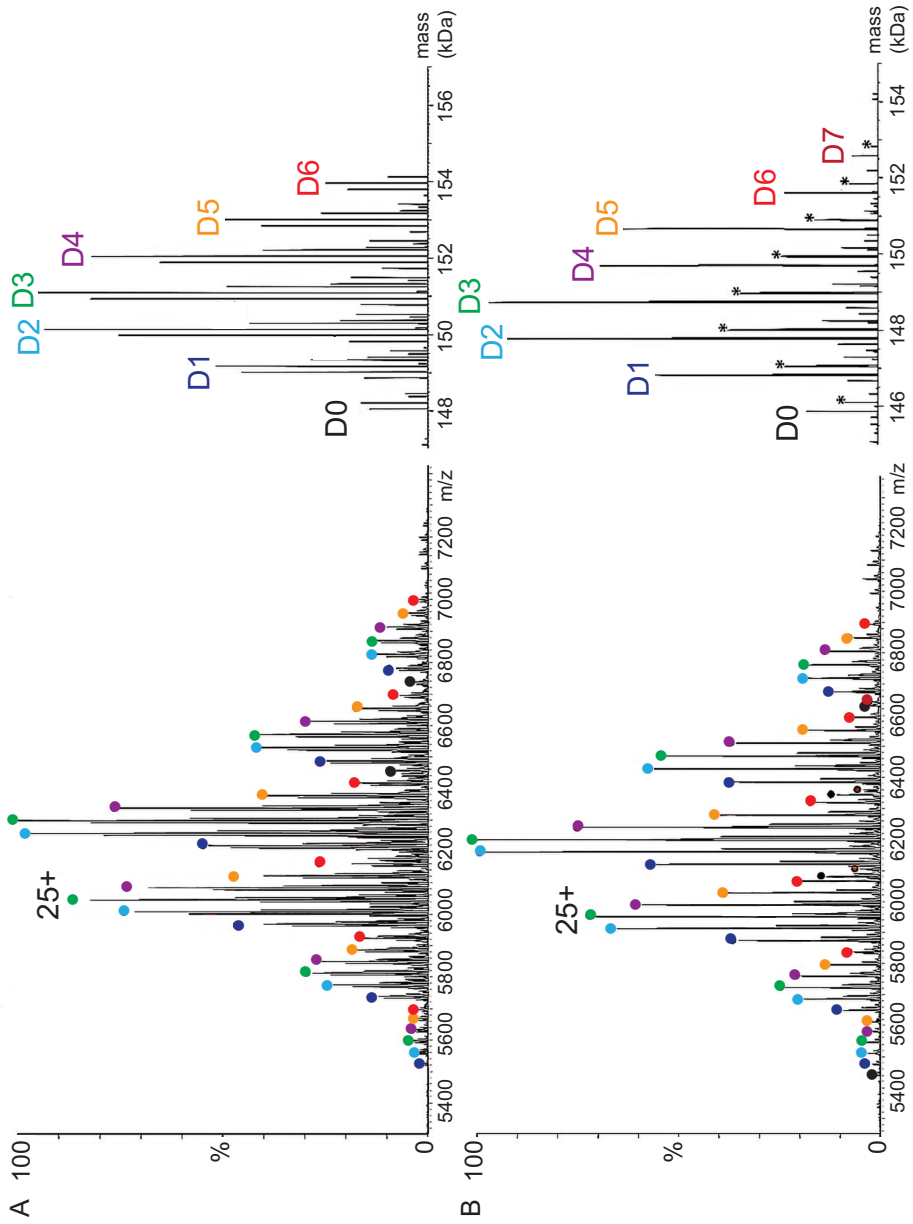


Figure 3.29: Raw (left) and deconvoluted (right) mass spectra obtained under native conditions by direct infusion of trastuzumab emtansine (A) before and (B) after deglycosylation. The asterisks indicate +220 Da linker adducts. Adapted with permission from Marcoux et al. [174]

raises the specter of disulfide scrambling (both internal and external), necessitating the additional scrutiny in the disulfide characterization work.

3.4.3 Protein–protein conjugates

An extension of the protein–small-molecule drug conjugates class includes carrier proteins conjugated to proteins with specific therapeutic properties. Although chemical conjugation is still considered a viable strategy in designing delivery vector/macromolecular payloads systems, the success record in this field remains modest: only a single biopharmaceutical of this class (TransMID, a transferrin/diphtheria toxin conjugate [170]) had progressed as far as stage III clinical trials (but was subsequently withdrawn). All approved and the vast majority of "in-the-pipeline" systems that combine a carrier protein and a payload protein are fusion proteins [171, 172]. Characterization of such biopharmaceutical products should follow the steps outlined in Sections 3.1–3.3.

References

[1] Fukuhara, K., Tsuji, T., Toi, K., Takao, T. & Shimonishi, Y. Verification by mass spectrometry of the primary structure of human interleukin-2. J. Biol. Chem. 260, 10487–10494 (1985).
[2] Sasaki, H., Bothner, B., Dell, A. & Fukuda, M. Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA. J. Biol. Chem. 262, 12059–12076 (1987).
[3] Raschdorf, F., Dahinden, R., Maerki, W., Richter, W.J. & Merryweather, J.P. Location of disulphide bonds in human insulin-like growth factors (IGFs) synthesized by recombinant DNA technology. Biomed. Environ. Mass Spectrom. 16, 3–8 (1988).
[4] Carini, M., Regazzoni, L. & Aldini, G. Mass spectrometric strategies and their applications for molecular mass determination of recombinant therapeutic proteins. Curr. Pharm. Biotechnol. 12, 1548–1557 (2011).
[5] Jin, Y. et al. Comprehensive characterization of monoclonal antibody by Fourier transform ion cyclotron resonance mass spectrometry. MAbs 11, 106–115 (2019).
[6] Mesonzhnik, N.V., Postnikov, P.V., Appolonova, S.A. & Krotov, G.I. Characterization and detection of erythropoietin fc fusion proteins using liquid chromatography–mass spectrometry. J. Proteome Res. 17, 689–697 (2018).
[7] Kirley, T.L., Greis, K.D. & Norman, A.B. Structural characterization of expressed monoclonal antibodies by single sample mass spectral analysis after IdeS proteolysis. Biochem. Biophys. Res. Commun. 477, 363–368 (2016).
[8] Lanter, C., Lev, M., Cao, L. & Loladze, V. Rapid Intact mass based multi-attribute method in support of mAb upstream process development. J. Biotechnol. 314–315, 63–70 (2020).
[9] Giansanti, P., Tsiatsiani, L., Low, T.Y. & Heck, A.J. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. Nat. Protoc. 11, 993–1006 (2016).
[10] Swaney, D.L., Wenger, C.D. & Coon, J.J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. J. Proteome Res. 9, 1323–1329 (2010).

[11] Tran, N.H. et al. Complete De Novo assembly of monoclonal antibody sequences. Sci. Rep. 6, 31730 (2016).
[12] Cao, M. et al. An automated and qualified platform method for site-specific succinimide and deamidation quantitation using low-pH peptide mapping. J. Pharm. Sci. 108, 3540–3549 (2019).
[13] Dada, O.O., Zhao, Y., Jaya, N. & Salas-Solano, O. High-resolution capillary zone electrophoresis with mass spectrometry peptide mapping of therapeutic proteins: peptide recovery and post-translational modification analysis in monoclonal antibodies and antibody-drug conjugates. Anal. Chem. 89, 11236–11242 (2017).
[14] Wenz, C. et al. Interlaboratory study to evaluate the robustness of capillary electrophoresis-mass spectrometry for peptide mapping. J. Sep. Sci. 38, 3262–3270 (2015).
[15] Kumar, R., Shah, R.L. & Rathore, A.S. Harnessing the power of electrophoresis and chromatography: offline coupling of reverse phase liquid chromatography-capillary zone electrophoresis-tandem mass spectrometry for peptide mapping for monoclonal antibodies. J. Chromatogr. A 1620, 460954 (2020).
[16] Yang, Y. et al. Detecting low level sequence variants in recombinant monoclonal antibodies. mAbs 2, 285–298 (2010).
[17] Zhang, Z., Shah, B. & Bondarenko, P.V. G/U and certain wobble position mismatches as possible main causes of amino acid misincorporations. Biochemistry 52, 8165–8176 (2013).
[18] Zhang, A. et al. A general evidence-based sequence variant control limit for recombinant therapeutic protein development. mAbs 12, 1791399 (2020).
[19] Wan, M., Shiau, F.Y., Gordon, W. & Wang, G.Y. Variant antibody identification by peptide mapping. Biotechnol. Bioeng. 62, 485–488 (1999).
[20] Forman, M.D., Stack, R.F., Masters, P.S., Hauer, C.R. & Baxter, S.M. High level, context dependent misincorporation of lysine for arginine in Saccharomyces cerevisiae a1 homeodomain expressed in Escherichia coli. Protein Sci.: Publ. Protein Soc. 7, 500–503 (1998).
[21] Wen, D. et al. Discovery and investigation of misincorporation of serine at asparagine positions in recombinant proteins expressed in Chinese hamster ovary cells. J. Biol. Chem. 284, 32686–32694 (2009).
[22] Hoopmann, M.R., Merrihew, G.E., Von Haller, P.D. & MacCoss, M.J. Post analysis data acquisition for the iterative MS/MS sampling of proteomics mixtures. J. Proteome Res. 8, 1870–1875 (2009).
[23] Zhang, Z. Automated precursor ion exclusion during LC-MS/MS data acquisition for optimal ion identification. J. Am. Soc. Mass Spectrom. 23, 1400–1407 (2012).
[24] Bern, M., Kil, Y.J. & Becker, C. Byonic: advanced peptide and protein identification software. Curr Protoc. Bioinf. Chapter 13, Unit13, 20 (2012).
[25] Zhang, Z. Large-scale identification and quantification of covalent modifications in therapeutic proteins. Anal. Chem. 81, 8354–8364 (2009).
[26] Fenyo, D. & Beavis, R.C. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal. Chem. 75, 768–774 (2003).
[27] Wang, Z., Rejtar, T., Zhou, Z.S. & Karger, B.L. Desulfurization of cysteine-containing peptides resulting from sample preparation for protein characterization by mass spectrometry. Rapid Commun. Mass Spectrom. 24, 267–275 (2010).
[28] Petritis, K. et al. Improved peptide elution time prediction for reversed-phase liquid chromatography-MS by incorporating peptide sequence information. Anal. Chem. 78, 5026–5039 (2006).

[29] Cui, W., Rohrs, H.W. & Gross, M.L. Top-down mass spectrometry: recent developments, applications and perspectives. Analyst 136, 3854–3864 (2011).
[30] Ren, D. et al. Top-down N-terminal sequencing of Immunoglobulin subunits with electrospray ionization time of flight mass spectrometry. Anal. Biochem. 384, 42–48 (2009).
[31] Bondarenko, P.V., Second, T.P., Zabrouskov, V., Makarov, A.A. & Zhang, Z.Q. Mass measurement and top-down HPLC/MS analysis of intact monoclonal antibodies on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. J. Am. Soc. Mass Spectrom. 20, 1415–1424 (2009).
[32] Shi, R.L., Xiao, G., Dillon, T.M., Ricci, M.S. & Bondarenko, P.V. Characterization of therapeutic proteins by cation exchange chromatography-mass spectrometry and top-down analysis. mAbs 12, 1739825 (2020).
[33] He, L. et al. Analysis of monoclonal antibodies in human serum as a model for clinical monoclonal gammopathy by use of 21 Tesla FT-ICR top-down and middle-down MS/MS. J. Am. Soc. Mass Spectrom. 28, 827–838 (2017).
[34] Pawlowski, J.W., Carrick, I. & Kaltashov, I.A. Integration of on-column chemical reactions in protein characterization by liquid chromatography/mass spectrometry: cross-path reactive chromatography. Anal. Chem. 90, 1348–1355 (2018).
[35] Hulting, G. et al. Two novel IgG endopeptidases of Streptococcus equi. FEMS Microbiol. Lett. 298, 44–50 (2009).
[36] Moelleken, J. et al. GingisKHAN™ protease cleavage allows a high-throughput antibody to Fab conversion enabling direct functional assessment during lead identification of human monoclonal and bispecific IgG1 antibodies. MAbs 9, 1076–1087 (2017).
[37] Van Der Burgt, Y.E.M. et al. Structural analysis of monoclonal antibodies by ultrahigh resolution MALDI In-source decay FT-ICR mass spectrometry. Anal. Chem. 91, 2079–2085 (2019).
[38] Gadgil, H.S. et al. Identification of cysteinylation of a free cysteine in the Fab region of a recombinant monoclonal IgG1 antibody using Lys-C limited proteolysis coupled with LC-MS analysis. Anal. Biochem. 355, 165–174 (2006).
[39] Srzentić, K. et al. Chemical-mediated digestion: an alternative realm for middle-down proteomics? J. Proteome Res. 17, 2005–2016 (2018).
[40] Wang, D. et al. Characterization of drug-product-related impurities and variants of a therapeutic monoclonal antibody by higher energy C-trap dissociation mass spectrometry. Anal. Chem. 87, 914–921 (2015).
[41] Fornelli, L. et al. Top-down analysis of immunoglobulin G isotypes 1 and 2 with electron transfer dissociation on a high-field Orbitrap mass spectrometer. J. Proteomics 159, 67–76 (2017).
[42] Mao, Y., Valeja, S.G., Rouse, J.C., Hendrickson, C.L. & Marshall, A.G. Top-down structural analysis of an intact monoclonal antibody by electron capture dissociation-Fourier transform ion cyclotron resonance-mass spectrometry. Anal. Chem. 85, 4239–4246 (2013).
[43] Cotham, V.C. & Brodbelt, J.S. Characterization of therapeutic monoclonal antibodies at the subunit-level using middle-down 193 nm ultraviolet photodissociation. Anal. Chem. 88, 4004–4013 (2016).
[44] Fornelli, L. et al. Accurate sequence analysis of a monoclonal antibody by top-down and middle-down orbitrap mass spectrometry applying multiple ion activation techniques. Anal. Chem. 90, 8421–8429 (2018).
[45] Muneeruddin, K., Nazzaro, M. & Kaltashov, I.A. Characterization of intact protein conjugates and biopharmaceuticals using ion-exchange chromatography with online detection by native electrospray ionization mass spectrometry and top-down tandem mass spectrometry. Anal. Chem. 87, 10138–10145 (2015).

[46] Belov, A.M. et al. Complementary middle-down and intact monoclonal antibody proteoform characterization by capillary zone electrophoresis – mass spectrometry. Electrophoresis 39, 2069–2082 (2018).
[47] Melani, R.D. et al. Direct measurement of light and heavy antibody chains using ion mobility and middle-down mass spectrometry. MAbs 11, 1351–1357 (2019).
[48] Donnelly, D.P. et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. Nat. Methods 16, 587–594 (2019).
[49] Turecek, P.L., Bossard, M.J., Schoetens, F. & Ivens, I.A. PEGylation of biopharmaceuticals: a review of chemistry and nonclinical safety information of approved drugs. J. Pharm. Sci. 105, 460–475 (2016).
[50] Veronese, F.M. & Mero, A. The impact of PEGylation on biological therapies. BioDrugs 22, 315 (2008).
[51] Harris, J.M. & Chess, R.B. Effect of PEGylation on pharmaceuticals. Nat. Rev. Drug Discov. 2, 214–221 (2003).
[52] Dashivets, T. et al. Oxidation in the complementarity-determining regions differentially influences the properties of therapeutic antibodies. MAbs 8, 1525–1535 (2016).
[53] Abes, R. & Teillaud, J.L. Impact of glycosylation on effector functions of therapeutic IgG. Pharmaceuticals (Basel) 3, 146–157 (2010).
[54] Peschke, B., Keller, C.W., Weber, P., Quast, I. & Lunemann, J.D. Fc-galactosylation of human immunoglobulin gamma isotypes improves C1q binding and enhances complement-dependent cytotoxicity. Front Immunol. 8, 646 (2017).
[55] Jones, A.J. et al. Selective clearance of glycoforms of a complex glycoprotein pharmaceutical caused by terminal N-acetylglucosamine is similar in humans and cynomolgus monkeys. Glycobiology 17, 529–540 (2007).
[56] Keck, R. et al. Characterization of a complex glycoprotein whose variable metabolic clearance in humans is dependent on terminal N-acetylglucosamine content. J. Int. Assoc. Biol. Stand. 36, 49–60 (2008).
[57] Zheng, K., Bantog, C. & Bayer, R. The impact of glycosylation on monoclonal antibody conformation and stability. mAbs 3, 568–576 (2011).
[58] Noguchi, A., Mukuria, C.J., Suzuki, E. & Naiki, M. Immunogenicity of N-glycolylneuraminic acid-containing carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. J. Biochem. 117, 59–62 (1995).
[59] Flesher, A.R., Marzowski, J., Wang, W.C. & Raff, H.V. Fluorophore-labeled carbohydrate analysis of immunoglobulin fusion proteins: correlation of oligosaccharide content with in vivo clearance profile. Biotechnol. Bioeng. 46, 399–407 (1995).
[60] Tengattini, S. et al. Hydrophilic interaction liquid chromatography-mass spectrometry as a new tool for the characterization of intact semi-synthetic glycoproteins. Analyt. Chim. Acta. 981, 94–105 (2017).
[61] Domínguez-Vega, E. et al. High-resolution glycoform profiling of intact therapeutic proteins by hydrophilic interaction chromatography-mass spectrometry. Talanta 184, 375–381 (2018).
[62] Pawlowski, J.W. et al. Influence of glycan modification on IgG1 biochemical and biophysical properties. J. Pharm. Biomed. Anal. 151, 133–144 (2018).
[63] Maverakis, E. et al. Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review. J. Autoimmun. 57, 1–13 (2015).
[64] Zaia, J. Mass spectrometry of oligosaccharides. Mass Spectrom. Rev. 23, 161–227 (2004).
[65] Leymarie, N. & Zaia, J. Effective use of mass spectrometry for glycan and glycopeptide structural analysis. Anal. Chem. 84, 3040–3048 (2012).
[66] Kailemia, M.J., Ruhaak, L.R., Lebrilla, C.B. & Amster, I.J. Oligosaccharide analysis by mass spectrometry: a review of recent developments. Anal. Chem. 86, 196–212 (2014).

[67] Harvey, D.J. Negative ion mass spectrometry for the analysis of N-linked glycans. Mass Spectrom. Rev. 39, 586-679 (2020).
[68] Woodin, C.L., Maxon, M. & Desaire, H. Software for automated interpretation of mass spectrometry data from glycans and glycopeptides. Analyst 138, 2793–2803 (2013).
[69] Lauber, M.A. et al. Rapid preparation of released N-Glycans for HILIC analysis using a labeling reagent that facilitates sensitive fluorescence and ESI-MS detection. Anal. Chem. 87, 5401–5409 (2015).
[70] Jayo, R.G. et al. Simple capillary electrophoresis-mass spectrometry method for complex glycan analysis using a flow-through microvial interface. Anal. Chem. 86, 6479–6486 (2014).
[71] Taron, C.H., Bielik, A.M., Guthrie, E.P. & Shi, X. Redesigning Glycosidase Manufacturing Quality for Pharmaceutical and Clinical Applications. New England Biolabs Doi:10.13140/RG.2.2.27258.95684.
[72] Abzalimov, R.R. & Kaltashov, I.A. Electrospray ionization mass spectrometry of highly heterogeneous protein systems: protein ion charge state assignment via incomplete charge reduction. Anal. Chem. 82, 7523–7526 (2010).
[73] Kuster, B. & Mann, M. 18O-labeling of N-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. Anal. Chem. 71, 1431–1440 (1999).
[74] Gilar, M. et al. Characterization of glycoprotein digests with hydrophilic interaction chromatography and mass spectrometry. Anal. Biochem. 417, 80–88 (2011).
[75] Lam, M.P. et al. Online combination of reversed-phase/reversed-phase and porous graphitic carbon liquid chromatography for multicomponent separation of proteomics and glycoproteomics samples. Electrophoresis 32, 2930–2940 (2011).
[76] Kaltashov, I.A. et al. Advances and challenges in analytical characterization of biotechnology products: mass spectrometry-based approaches to study properties and behavior of protein therapeutics. Biotechnol. Adv. 30, 210–222 (2012).
[77] Dutta, D., Mandal, C. & Mandal, C. Unusual glycosylation of proteins: beyond the universal sequon and other amino acids. Biochim. Biophys. Acta. 1861, 3096–3108 (2017).
[78] Yan, Y., Xing, T., Wang, S. & Li, N. Versatile, sensitive, and robust native LC-MS platform for intact mass analysis of protein drugs. J. Am. Soc. Mass Spectrom. 31, 2171–2179 (2020).
[79] Wang, S. et al. Simple approach for improved LC-MS analysis of protein biopharmaceuticals via modification of desolvation gas. Anal. Chem. 91, 3156–3162 (2019). https://pubs.acs.org/doi/10.1021/acs.analchem.8b05846.
[80] Zhang, M. & Kaltashov, I.A. Mapping of protein disulfide bonds using negative ion fragmentation with a broadband precursor selection. Anal. Chem. 78, 4820–4829 (2006).
[81] Han, J.C. & Han, G.Y. A procedure for quantitative determination of Tris(2-carboxyethyl) phosphine, an odorless reducing agent more stable and effective than dithiothreitol. Anal. Biochem. 220, 5–10 (1994).
[82] Li, X. et al. Disulfide bond assignment of an IgG1 monoclonal antibody by LC-MS with post-column partial reduction. Anal. Biochem. 436, 93–100 (2013).
[83] Cramer, C.N., Haselmann, K.F., Olsen, J.V. & Nielsen, P.K. Disulfide linkage characterization of disulfide bond-containing proteins and peptides by reducing electrochemistry and mass spectrometry. Anal. Chem. 88, 1585–1592 (2016).
[84] Nicolardi, S., Deelder, A.M., Palmblad, M. & Van Der Burgt, Y.E.M. Structural analysis of an intact monoclonal antibody by online electrochemical reduction of disulfide bonds and fourier transform ion cyclotron resonance mass spectrometry. Anal. Chem. 86, 5376–5382 (2014).

[85] Bilusich, D., Brinkworth, C.S., McAnoy, A.M. & Bowie, J.H. The fragmentations of $[M-H]^-$ anions derived from underivatised peptides. The side-chain loss of H_2S from Cys. A joint experimental and theoretical study. Rapid Commun. Mass Spectrom. 17, 2488–2494 (2003).
[86] Wu, S.L. et al. Mass spectrometric determination of disulfide linkages in recombinant therapeutic proteins using online LC-MS with electron-transfer dissociation. Anal. Chem. 81, 112–122 (2009).
[87] Quick, M.M., Crittenden, C.M., Rosenberg, J.A. & Brodbelt, J.S. Characterization of disulfide linkages in proteins by 193 nm Ultraviolet Photodissociation (UVPD) mass spectrometry. Anal. Chem. 90, 8523–8530 (2018).
[88] Li, H. & O'Connor, P.B. Electron capture dissociation of disulfide, sulfur-selenium, and diselenide bound peptides. J. Am. Soc. Mass Spectrom. 23, 2001–2010 (2012).
[89] Chen, X., Zhou, Y., Peng, X. & Yoon, J. Fluorescent and colorimetric probes for detection of thiols. Chem. Soc. Rev. 39, 2120–2135 (2010).
[90] Wang, S. & Kaltashov, I.A. A new strategy of using O18-labeled iodoacetic acid for mass spectrometry-based protein quantitation. J. Am. Soc. Mass Spectrom. 23, 1293–1297 (2012).
[91] Kita, A., Ponniah, G., Nowak, C. & Liu, H. Characterization of cysteinylation and trisulfide bonds in a recombinant monoclonal antibody. Anal. Chem. 88, 5430–5437 (2016).
[92] Cohen, S.L., Price, C. & Vlasak, J. Beta-elimination and peptide bond hydrolysis: two distinct mechanisms of human IgG1 hinge fragmentation upon storage. J. Am. Chem. Soc. 129, 6976–6977 (2007).
[93] Gu, S. et al. Characterization of trisulfide modification in antibodies. Anal. Biochem. 400, 89–98 (2010).
[94] Houde, D.J., Bou-Assaf, G.M. & Berkowitz, S.A. Deciphering the biophysical effects of oxidizing sulfur-containing amino acids in interferon-beta-1a using MS and HDX-MS. J. Am. Soc. Mass Spectrom. 28, 840–849 (2017).
[95] Wang, S. & Kaltashov, I.A. Identification of reduction-susceptible disulfide bonds in transferrin by differential alkylation using O(16)/O(18) labeled iodoacetic acid. J. Am. Soc. Mass Spectrom. 26, 800–807 (2015).
[96] Liu, H., Chumsae, C., Gaza-Bulseco, G., Hurkmans, K. & Radziejewski, C.H. Ranking the susceptibility of disulfide bonds in human IgG1 antibodies by reduction, differential alkylation, and LC-MS analysis. Anal. Chem. 82, 5219–5226 (2010).
[97] Wang, S., Liu, A.P. & Li, N. An (18)O-Labeling Assisted LC-MS Method for Accurate Quantitation of Unprocessed C-Terminal Lysine in Therapeutic Monoclonal Antibodies. J. Am. Soc. Mass Spectrom. 31, 1587–1592 (2020).
[98] Hu, Z. et al. Carboxypeptidase D is the only enzyme responsible for antibody C-terminal lysine cleavage in Chinese hamster ovary (CHO) cells. Biotechnol. Bioeng. 113, 2100–2106 (2016).
[99] Khawli, L.A. et al. Charge variants in IgG1: Isolation, characterization, in vitro binding properties and pharmacokinetics in rats. mAbs 2, 613–624 (2010).
[100] Boswell, C.A. et al. Effects of charge on antibody tissue distribution and pharmacokinetics. Bioconjug. Chem. 21, 2153–2163 (2010).
[101] Hintersteiner, B. et al. Charge heterogeneity: basic antibody charge variants with increased binding to Fc receptors. mAbs 8, 1548–1560 (2016).
[102] Sricholpech, M. et al. Lysyl hydroxylase 3-mediated glucosylation in type I collagen: molecular loci and biological significance. J. Biol. Chem. 287, 22998–23009 (2012).
[103] Xie, Q., Moore, B. & Beardsley, R.L. Discovery and characterization of hydroxylysine in recombinant monoclonal antibodies. mAbs 8, 371–378 (2016).
[104] Vlasak, J. et al. Identification and characterization of asparagine deamidation in the light chain CDR1 of a humanized IgG1 antibody. Anal. Biochem. 392, 145–154 (2009).

[105] Yan, B. et al. Succinimide formation at Asn 55 in the complementarity determining region of a recombinant monoclonal antibody IgG1 heavy chain. J. Pharm. Sci. 98, 3509–3521 (2009).
[106] Alam, M.E. et al. Deamidation can compromise antibody colloidal stability and enhance aggregation in a pH-dependent manner. Mol. Pharm. 16, 1939–1949 (2019).
[107] Van Beers, M.M. & Bardor, M. Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins. Biotechnol. J. 7, 1473–1484 (2012).
[108] Ren, D. et al. An improved trypsin digestion method minimizes digestion-induced modifications on proteins. Anal. Biochem. 392, 12–21 (2009).
[109] Huang, H.Z., Nichols, A. & Liu, D. Direct identification and quantification of aspartyl succinimide in an IgG2 mAb by RapiGest assisted digestion. Anal. Chem. 81, 1686–1692 (2009).
[110] Liu, S., Moulton, K.R., Auclair, J.R. & Zhou, Z.S. Mildly acidic conditions eliminate deamidation artifact during proteolysis: digestion with endoprotease Glu-C at pH 4.5. Amino Acids 48, 1059–1067 (2016).
[111] Gaza-Bulseco, G., Li, B., Bulseco, A. & Liu, H.C. Method to differentiate asn deamidation that occurred prior to and during sample preparation of a monoclonal antibody. Anal. Chem. 80, 9491–9498 (2008).
[112] Du, Y., Wang, F., May, K., Xu, W. & Liu, H. Determination of deamidation artifacts introduced by sample preparation using 18O-labeling and tandem mass spectrometry analysis. Anal. Chem. 84, 6355–6360 (2012).
[113] Yu, X.C. et al. Accurate determination of succinimide degradation products using high fidelity trypsin digestion peptide map analysis. Anal. Chem. 83, 5912–5919 (2011).
[114] Wang, S. & Kaltashov, I.A. An 18O-labeling assisted LC-MS method for assignment of aspartyl/isoaspartyl products from Asn deamidation and Asp isomerization in proteins. Anal. Chem. 85, 6446–6452 (2013).
[115] Edman, P. A method for the determination of amino acid sequence in peptides. Arch. Biochem. 22, 475 (1949).
[116] Winter, D., Pipkorn, R. & Lehmann, W.D. Separation of peptide isomers and conformers by ultra performance liquid chromatography. J. Sep. Sci. 32, 1111–1119 (2009).
[117] Cournoyer, J.J. et al. Deamidation: differentiation of aspartyl from isoaspartyl products in peptides by electron capture dissociation. Protein Sci: Publ Protein Soc 14, 452–463 (2005).
[118] Yang, H., Fung, E.Y., Zubarev, A.R. & Zubarev, R.A. Toward proteome-scale identification and quantification of isoaspartyl residues in biological samples. J. Proteome Res. 8, 4615–4621 (2009).
[119] Cournoyer, J.J., Lin, C. & O'Connor, P.B. Detecting deamidation products in proteins by electron capture dissociation. Anal. Chem. 78, 1264–1271 (2006).
[120] Drazic, A. & Winter, J. The physiological role of reversible methionine oxidation. Biochim. Biophys. Acta. 1844, 1367–1382 (2014).
[121] Burkitt, W., Domann, P. & O'Connor, G. Conformational changes in oxidatively stressed monoclonal antibodies studied by hydrogen exchange mass spectrometry. Protein Sci. 19, 826–835 (2010).
[122] Kang, P. et al. Use of a stable-isotope-labeled reporter peptide and antioxidants for reliable quantification of methionine oxidation in a monoclonal antibody by liquid chromatography/mass spectrometry. Rapid Commun. Mass Spectrom. 30, 1734–1742 (2016).
[123] Liu, H., Ponniah, G., Neill, A., Patel, R. & Andrien, B. Accurate determination of protein methionine oxidation by stable isotope labeling and LC-MS analysis. Anal. Chem. 85, 11705–11709 (2013).

[124] Grassi, L. & Cabrele, C. Susceptibility of protein therapeutics to spontaneous chemical modifications by oxidation, cyclization, and elimination reactions. Amino Acids 51, 1409–1431 (2019).
[125] Wong, C., Strachan-Mills, C. & Burman, S. Facile method of quantification for oxidized tryptophan degradants of monoclonal antibody by mixed mode ultra performance liquid chromatography. J. Chromatogr. A 1270, 153–161 (2012).
[126] Yang, J., Wang, S., Liu, J. & Raghani, A. Determination of tryptophan oxidation of monoclonal antibody by reversed phase high performance liquid chromatography. J. Chromatogr. A 1156, 174–182 (2007).
[127] Giulivi, C., Traaseth, N.J. & Davies, K.J. Tyrosine oxidation products: analysis and biological relevance. Amino Acids 25, 227–232 (2003).
[128] Aeschbach, R., Amado, R. & Neukom, H. Formation of dityrosine cross-links in proteins by oxidation of tyrosine residues. Biochim. Biophys. Acta. 439, 292–301 (1976).
[129] Xu, C.F. et al. Discovery and characterization of histidine oxidation initiated cross-links in an IgG1 monoclonal antibody. Anal. Chem. 89, 7915–7923 (2017).
[130] Mukherjee, S. et al. Characterization and Identification of Dityrosine Cross-Linked Peptides Using Tandem Mass Spectrometry. Anal. Chem. 89, 6136–6145 (2017).
[131] Camperi, J., Guillarme, D., Lei, M. & Stella, C. Automated middle-up approach for the characterization of biotherapeutic products by combining on-line hinge-specific digestion with RPLC-HRMS analysis. J. Pharm. Biomed. Anal. 182, 113130 (2020).
[132] Viski, K., Gengeliczki, Z., Lenkey, K. & Baranyáné Ganzler, K. Parallel development of chromatographic and mass-spectrometric methods for quantitative analysis of glycation on an IgG1 monoclonal antibody. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 1032, 198–204 (2016).
[133] Zhang, Q. et al. Enrichment and analysis of nonenzymatically glycated peptides: boronate affinity chromatography coupled with electron-transfer dissociation mass spectrometry. J. Proteome Res. 6, 2323–2330 (2007).
[134] Brady, L.J., Martinez, T. & Balland, A. Characterization of nonenzymatic glycation on a monoclonal antibody. Anal. Chem. 79, 9403–9413 (2007).
[135] Liu, H., Ponniah, G., Neill, A., Patel, R. & Andrien, B. Identification and comparative quantitation of glycation by stable isotope labeling and LC-MS. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 958, 90–95 (2014).
[136] Yang, W. et al. Mechanism of reduction in IgG and IgE binding of β-lactoglobulin induced by ultrasound pretreatment combined with dry-state glycation: a study using conventional spectrometry and high-resolution mass spectrometry. J. Agric. Food Chem. 65, 8018–8027 (2017).
[137] Priego-Capote, F. et al. Glycation isotopic labeling with 13C-reducing sugars for quantitative analysis of glycated proteins in human plasma. Mol. Cell. Proteomics 9, 579–592 (2010).
[138] Rabbani, N. & Thornalley, P.J. Glycation research in amino acids: a place to call home. Amino Acids 42, 1087–1096 (2012).
[139] Ahmed, M.U., Thorpe, S.R. & Baynes, J.W. Identification of N epsilon-carboxymethyllysine as a degradation product of fructoselysine in glycated protein. J. Biol. Chem. 261, 4889–4894 (1986).
[140] Zhang, Q., Ames, J.M., Smith, R.D., Baynes, J.W. & Metz, T.O. A perspective on the Maillard reaction and the analysis of protein glycation by mass spectrometry: probing the pathogenesis of chronic disease. J. Proteome Res. 8, 754–769 (2009).
[141] Liu, H. et al. In vitro and in vivo modifications of recombinant and human IgG antibodies. mAbs 6, 1145–1154 (2014).

[142] Dick, L.W. Jr., Kim, C., Qiu, D. & Cheng, K.-C. Determination of the origin of the N-terminal pyro-glutamate variation in monoclonal antibodies using model peptides. Biotechnol. Bioeng. 97, 544–553 (2007).
[143] Liu, Y.D., Goetze, A.M., Bass, R.B. & Flynn, G.C. N-terminal glutamate to pyroglutamate conversion in vivo for human IgG2 antibodies. J. Biol. Chem. 286, 11211–11217 (2011).
[144] Yin, S., Pastuskovas, C.V., Khawli, L.A. & Stults, J.T. Characterization of therapeutic monoclonal antibodies reveals differences between in vitro and in vivo time-course studies. Pharm. Res. 30, 167–178 (2013).
[145] Chelius, D. et al. Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin gamma antibodies. Anal. Chem. 78, 2370–2376 (2006).
[146] Yu, L., Remmele, R.L. Jr. & He, B. Identification of N-terminal modification for recombinant monoclonal antibody light chain using partial reduction and quadrupole time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. 20, 3674–3680 (2006).
[147] Kumar, M., Chatterjee, A., Khedkar, A.P., Kusumanchi, M. & Adhikary, L. Mass spectrometric distinction of in-source and in-solution pyroglutamate and succinimide in proteins: a case study on rhG-CSF. J. Am. Soc. Mass Spectrom. 24, 202–212 (2013).
[148] Vlasak, J. & Ionescu, R. Fragmentation of monoclonal antibodies. mAbs 3, 253–263 (2011).
[149] Cordoba, A.J., Shyong, B.J., Breen, D. & Harris, R.J. Non-enzymatic hinge region fragmentation of antibodies in solution. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 818, 115–121 (2005).
[150] Yang, B. et al. Discovery and characterization of CHO host cell protease-induced fragmentation of a recombinant monoclonal antibody during production process development. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 1112, 1–10 (2019).
[151] Cohen, S.L., Price, C. & Vlasak, J. β-Elimination and Peptide Bond Hydrolysis: Two Distinct Mechanisms of Human IgG1 Hinge Fragmentation upon Storage. J. Am. Chem. Soc. 129, 6976–6977 (2007).
[152] Abzalimov, R.R., Frimpong, A. & Kaltashov, I.A. Structural characterization of protein-polymer conjugates. I. Assessing heterogeneity of a small PEGylated protein and mapping conjugation sites using ion exchange chromatography and top-down tandem mass spectrometry. Int. J. Mass Spectrom. 312, 135–143 (2012).
[153] González-Valdez, J., Rito-Palomares, M. & Benavides, J. Advances and trends in the design, analysis, and characterization of polymer-protein conjugates for "PEGylaided" bioprocesses. Anal. Bioanal. Chem. 403, 2225–2235 (2012).
[154] Muneeruddin, K. et al. Characterization of a PEGylated protein therapeutic by ion exchange chromatography with on-line detection by native ESI MS and MS/MS. Analyst 142, 336–344 (2017).
[155] Selby, T.L., Wesdemiotis, C. & Lattimer, R.P. Dissociation characteristics of $[M+X]^+$ ions (X=H, Li, K) from linear and cyclic polyglycols. J. Am. Soc. Mass Spectrom. 5, 1081–1092 (1994).
[156] Seyfried, B.K. et al. MALDI linear TOF mass spectrometry of PEGylated (glyco)proteins. J. Mass Spectrom. 45, 612–617 (2010).
[157] Forstenlehner, I.C. et al. A Direct-Infusion- and HPLC-ESI-Orbitrap-MS Approach for the Characterization of Intact PEGylated Proteins. Anal. Chem. 86, 826–834 (2014).
[158] Gerislioglu, S., Adams, S.R. & Wesdemiotis, C. Characterization of singly and multiply PEGylated insulin isomers by reversed-phase ultra-performance liquid chromatography interfaced with ion mobility mass spectrometry. Analyt. Chim. Acta. 1004, 58–66 (2018).
[159] Vestling, M.M. et al. A strategy for characterization of polyethylene glycol-derivatized proteins – a mass-spectrometric analysis of the attachment sites in polyethylene glycol-derivatized superoxide-dismutase. Drug Metab. Dispos. 21, 911–917 (1993).

[160] Veronese, F.M. et al. New PEGs for peptide and protein modification, suitable for identification of the PEGylation site. Bioconjugate Chem. 12, 62–70 (2001).
[161] Lee, D.L., Sharif, I., Kodihalli, S., Stewart, D.I.H. & Tsvetnitsky, V. Preparation and characterization of monopegylated human granulocyte-macrophage colony-stimulating factor. J. Interf. Cyt. Res. 28, 101–112 (2008).
[162] Na, D.H. & Lee, K.C. Capillary electrophoretic characterization of PEGylated human parathyroid hormone with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal. Biochem. 331, 322–328 (2004).
[163] Cindric, M. et al. Structural characterization of PEGylated rHuG-CSF and location of PEG attachment sites. J. Pharm. Biomed. Anal. 44, 388–395 (2007).
[164] Lee, H. & Park, T.G. A novel method for identifying PEGylation sites of protein using biotinylated PEG derivatives. J. Pharm. Sci. 92, 97–103 (2003).
[165] Lu, X.J., Gough, P.C., DeFelippis, M.R. & Huang, L.H. Elucidation of PEGylation site with a combined approach of in-source fragmentation and CID MS/MS. J. Am. Soc. Mass Spectrom. 21, 810–818 (2010).
[166] Tsuchikama, K. & An, Z. Antibody-drug conjugates: recent advances in conjugation and linker chemistries. Protein Cell 9, 33–46 (2018).
[167] Janin-Bussat, M.C., Dillenbourg, M., Corvaia, N., Beck, A. & Klinguer-Hamour, C. Characterization of antibody drug conjugate positional isomers at cysteine residues by peptide mapping LC-MS analysis. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 981–982, 9–13 (2015).
[168] Han, L., Zhao, Y. & Zhang, Q. Conjugation site analysis by MS/MS protein sequencing. Methods Mol. Biol. 2078, 221–233 (2020).
[169] Hernandez-Alba, O. et al. A case study to identify the drug conjugation site of a site-specific antibody-drug-conjugate using middle-down mass spectrometry. J. Am. Soc. Mass Spectrom. 30, 2419–2429 (2019).
[170] Rainov, N.G. & Soling, A. Technology evaluation: TransMID, KS Biomedix/Nycomed/Sosei/PharmaEngine. Curr. Opin. Mol. Ther. 7, 483–492 (2005).
[171] Murer, P. & Neri, D. Antibody-cytokine fusion proteins: a novel class of biopharmaceuticals for the therapy of cancer and of chronic inflammation. N. Biotechnol. 52, 42–53 (2019).
[172] Pardridge, W.M. Re-engineering biopharmaceuticals for delivery to brain with molecular Trojan horses. Bioconjugate Chem. 19, 1327–1338 (2008).
[173] Liu, A. et al. Achieving 100% Sequence Coverage of Monoclonal Antibodies by Tryptic Digestion Using a Dual-Column LC-MS System. Proceedings of the 66thASMS Conference on Mass Spectrometry and Allied Topics, San Diego, California, June 3-7, (2018).
[174] Marcoux, J., Champion, T., Colas, O., Wagner-Rousset, E., Corvaïa, N., Van Dorsselaer, A., Beck, A. & Cianférani, S. Native mass spectrometry and ion mobility characterization of trastuzumab emtansine, a lysine-linked antibody drug conjugate. Protein Sci. 24, 1210-1223 (2015).

Chapter 4
Characterization of higher order structure and protein interactions

Higher order structure is an important quality attribute of protein therapeutics, which dictates their potency and safety. Indeed, protein conformation is an important determinant of its function, and the loss of native conformation even within a small subset of the protein drug molecules has a negative impact on the potency. Even more important is the safety aspect associated with the integrity of the higher order structure, as even a relatively small-scale loss of structure may expose neo-epitopes that can potentially trigger an immune response to the protein drug. Mass spectrometry has been a relatively recent addition to the analytical/biophysical armamentarium used to study the higher order structure of protein therapeutics. Nevertheless, it is now routinely used to probe conformation of macromolecular medicines as complex as monoclonal antibodies at a level of detail that is not unattainable by the traditional techniques. At present, mass spectrometry is an integral part of the analytical toolbox across the industry, which is critical not only for the quality control work, but also in the discovery and development stages.

4.1 Mass spectrometry and its place in the analytical toolbox used for higher order structure characterization of protein therapeutics

The integrity of the higher order structure of a protein therapeutic may be compromised at any stage during its life cycle from production to formulation to storage to administration, highlighting the need to have robust, sensitive, and highly reliable methods to monitor conformational integrity of biopharmaceutical products. Modern biophysics offers an impressive arsenal of state-of-the-art tools for elucidating most intimate details of protein higher order structure [1]; however, many of them are poorly suited for the characterization of biopharmaceutical products. For example, high-resolution NMR still suffers from molecular weight limitations which place detailed structural analysis of most protein therapeutics with natural isotopic abundance outside of its reach. This limits the scope of NMR in biopharmaceutical analysis to fingerprinting [2, 3]. These rather unforgiving molecular weight limitations are frequently lifted when X-ray crystallography is employed as a means of structural analysis of proteins and their assemblies, making this technique an invaluable tool for characterizing higher order structure not only of protein therapeutics but also their complexes with physiological partners and therapeutic targets [4, 5]. While the role of X-ray crystallography in the development of biologics is indisputable [6], this technique also has significant limitations that make it less suited for routine analysis of the higher order structure of biopharmaceutical products. The two most important limitations within the context of our discussion are (i) the inability of X-ray

https://doi.org/10.1515/9783110546187-004

crystallography to observe dynamic events directly and (ii) the need to employ crystallization conditions which are frequently very different from those that are relevant to the production or storage conditions of most protein therapeutics. As a result, until very recently the field of biophysical characterization of protein therapeutics was dominated by classical techniques, such as optical spectroscopy, light scattering, analytical ultracentrifugation, calorimetry, and size-exclusion chromatography (SEC) [7]. While being robust and free of molecular weight limitations, these techniques do not reveal structural information at high resolution. Instead, they provide cumulative or average characteristics, such as the overall degree of aromatic residues solvent exposure (fluorescence), total percentages of helical, β-sheet or random coil structures within a protein molecule (circular dichroism), or molecular weight of a multi-unit protein assembly from which the association stoichiometry can be deduced (light scattering and analytical ultracentrifugation).

Above and beyond being an excellent tool for characterization of covalent structure of recombinant proteins and related biopharmaceutical products at a variety of levels (as outlined in Chapter 3), mass spectrometry has the capability to provide information on their higher order structure as well [8, 9]. In fact, it was not long after the introduction of ESI MS as a bioanalytical tool [10, 11] that its potential to probe various aspects of protein higher order structure became evident. This included the ability to monitor changes in the protein tertiary structure by analyzing ionic charge state distributions in ESI mass spectra [12, 13], as well as to observe non-covalent protein–protein and protein–small ligand complexes that were preserved during the ESI process [14–16] (Section 4.2). Availability of commercial ESI MS (and, to a lesser extent, MALDI MS) instruments also provided further impetus to the development of experimental strategies that rely on hydrogen/deuterium exchange as a means of studying protein dynamics in solution [17, 18] (Section 4.3). Above and beyond proteins, higher order structures of nucleic acids were also shown to be amenable to interrogation by ESI MS [19] (these methods will be reviewed in Chapter 8). Therefore, it is not surprising that both academic laboratories and the key stakeholders in industry were keen on evaluating the capabilities of mass spectrometry vis-à-vis characterization of the higher order structure of biopharmaceutical products as early as mid-1990s [20], with the first reports of successful use of mass spectrometry for these tasks published in the following decade [21, 22]. In the years passed since the publication of these initial reports, mass spectrometry was demonstrated to be a robust and reliable tool capable of characterizing various aspects of the higher order structure of protein therapeutics and related products at an unprecedented level of detail. It is an integral part of the analytical toolbox across the industry, playing critical roles not only in the quality control but also at the discovery and development stages. In fact, a brief survey of recent publications and conference presentations has shown that the majority of biopharma laboratories are now applying MS-based characterization of the higher order structure at certain stages during drug development, covering

applications ranging from epitope mapping, structural comparability assessment, and aggregation interface analysis to name just a few.

4.2 Native electrospray ionization MS

Following the advent of ESI MS [23, 24], the explosive growth of this technique's popularity in the field of bioanalysis was fueled primarily by its ability to measure masses of intact biopolymers (polypeptides in particular). However, it was soon realized that the gentle nature of this ionization technique also allows the higher order structure of proteins in many cases to be preserved, as long as the solution conditions used for sample preparation do not cause protein denaturation [13, 14]. This eventually led to the introduction of a technique which is now known as "native MS" and widely used in a variety of applications ranging from assessing the conformational integrity of biopharmaceutical products to elucidation of their quaternary structure to monitoring their interactions with physiological partners and therapeutic targets (Figure 4.1).

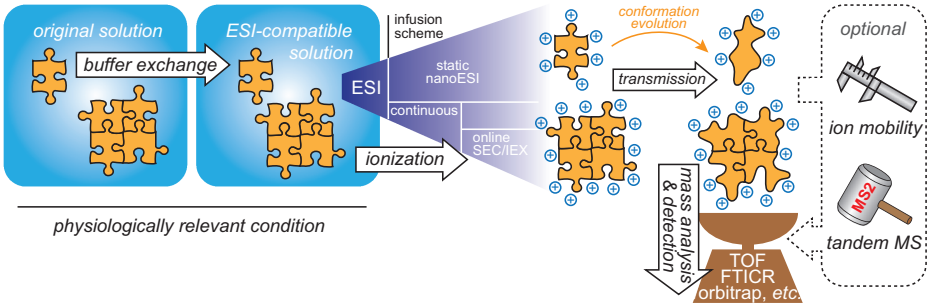


Figure 4.1: Schematic representation of major processes and measurements in native MS analysis. Adapted from Tong et al. [154] with permission from Elsevier.

4.2.1 Protein ion charge state distributions: conformational integrity of monomeric proteins

One of the unique features of ESI-generated macromolecular ions is that they carry multiple charges. Although once thought to be a reflection of solution-phase acid–base equilibria [25–27], it is now firmly established that the extent of multiple charging is dictated primarily by the physical dimensions of the macromolecule in solution [28] (with its solvent-accessible surface area showing a particularly strong correlation with the average ionic charge in ESI MS [29]), and attenuated by a range of gas-phase processes in the ESI interface [30–34]. As far as proteins are concerned, even partial

unfolding in solution results in a significant increase of their hydrodynamic radii, and the loss of conformational integrity is reflected in the changes of the protein charge state distributions [35]. This phenomenon is shown in Figure 4.2, where a mass spectrum of a monoclonal antibody (mAb) acquired under near-native conditions in solution (physiological pH and ionic strength, no organic co-solvent) exhibits a narrow distribution of charge states (+26 through +30), while the very same protein gives rise to a dramatically different ionic signal (with the number of charges exceeding +50) when placed in a strongly denaturing solution (low pH/low ionic strength and organic co-solvent). It is possible in many cases to obtain more nuanced information on protein conformations from the charge state distributions, as suggested by the appearance of the mAb mass spectrum acquired under mildly denaturing conditions (pH 3.0, physiological ionic strength, represented with a black trace in Figure 4.2). A subset of ions in this distribution (m/z > 5,000) are clearly aligned with the mass spectrum acquired under native conditions, and likely represent the sub-population of mAb molecules maintaining the near-native fold. The extent of the multiple charging for the rest of ionic population is notably higher, indicating partial loss of the native structure. The uneven shape of this part of the charge state distributions (the presence of two local maxima at z = +47 and +37) suggests that the population of mAb molecules giving rise to the ionic signal in this m/z region is also heterogeneous, and contains contributions from different non-native conformations. All these states retain a significant amount of residual structure, as the extent of multiple charging displayed by these ions is below that of mAb ions generated under strongly denaturing conditions in solution. An important feature of the charge state distribution analysis in ESI MS is that it can be readily carried out even if multiple species are present in solution, as long as a distinction can be made among them based on the mass difference(s). For example, the mass spectra shown in Figure 4.2 reveal the presence of multiple mAb glycoforms, including a glycan-free species. The charge state distribution of the a-glycosylated (glycan-free) mAb species closely mirrors that of the glycosylated molecules both under native and mildly denaturing conditions, providing a clear indication that in this particular case the absence of the glycans does not compromise the higher order structure.

While the conformational integrity of the protein therapeutic discussed in the preceding paragraph was compromised by placing it in a denaturing solvent, partial unfolding is frequently triggered by stress-related PTMs. An example is shown in Figure 4.3, where identical (near-native) solvent conditions were used to obtain mass spectra of intact and stress-oxidized forms of recombinant acid-β-glucocerebrosidase (VPRIV^{TM}), a protein therapeutic used for enzyme replacement therapy treatment of Gaucher's disease [36]. As a lysosomal protein, acid-β-glucocerebrosidase is natively folded at mildly acidic (lysosomal) pH, and its ESI mass spectrum under these conditions displays a narrow charge state distribution (ranging from +14 to +17), consistent with a compact conformation in solution. However, oxidative stress (which results in

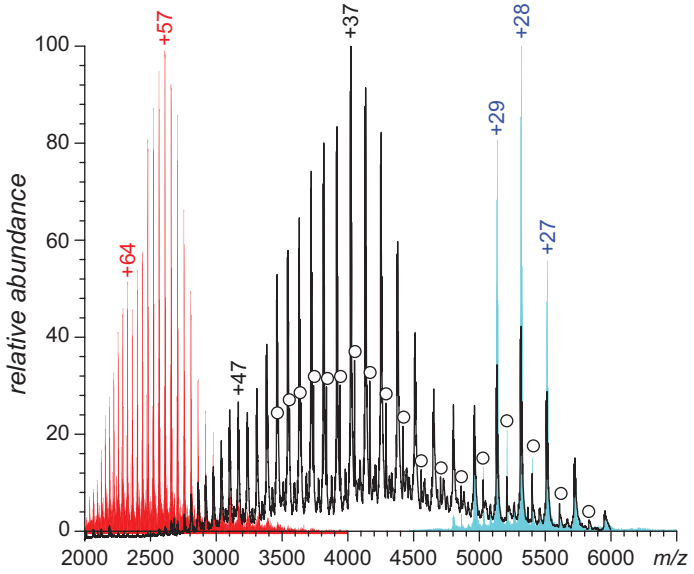


Figure 4.2: Mass spectra of a monoclonal antibody acquired under strongly denaturing (H_2O/CH_3OH/CH_3CO_2H, 47:50:3 v:v:v, red trace), mildly denaturing (150 mM $NH_4CH_3CO_2$, pH adjusted to 3.0, black trace) and near-native conditions (150 mM $NH_4CH_3CO_2$, pH 7.0, cyan trace). Open circles indicate ions representing the glycan-free form of the protein. Adapted from Pawlowski et al. [9] with permission from Elsevier.

a moderate mass increase, consistent with oxidation of 1-2 methionine residues per polypeptide chain) gives rise to a sub-population of ions with a significantly higher charge density (up to +30), consistent with the notion of a partial unfolding of the protein. The majority of the protein ions are still in the high m/z (low charge density) range, consistent with the authors' interpretation that the stress oxidation leads to a transient unfolding of the recombinant acid-β-glucocerebrosidase (affecting all protein molecules), rather than creating a distinct subset of the partially unfolded species [37]. Above and beyond the influence of stress-related PTMs on the higher order structure of protein therapeutics, charge state distribution analysis has also been used to evaluate the influence of specific production systems (cell lines) [38] and storage conditions [39] on the protein conformational integrity.

4.2.2 Native mass spectrometry and non-covalent assemblies: protein quaternary structure and interactions of protein therapeutics with their targets and physiological partners

The gentle nature of the ion production process in ESI allows a large variety of non-covalent assemblies of proteins (as well as other biopolymers) to be preserved in the

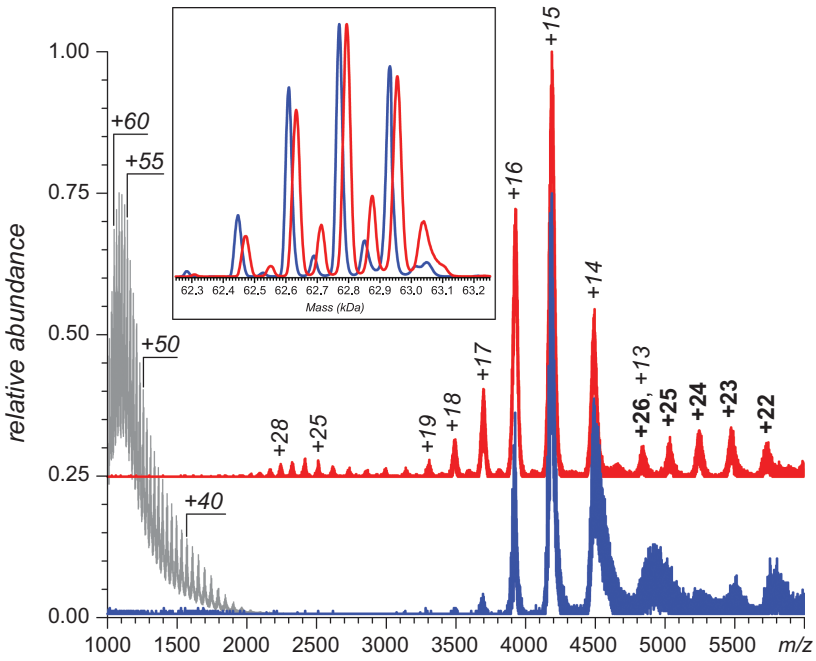


Figure 4.3: Nano-ESI mass spectra of 6 µM solutions (in 50 mM $NH_4CH_3CO_2$, pH adjusted to 4.5) of acid-β-glucocerebrosidase (velaglucerase, VPRIV™) in its intact (blue) and stress-oxidized (red) forms. The gray trace shows a reference mass spectrum of the protein under denaturing conditions ($H_2O/CH_3CN/CH_3CO_2H$, 47:50:3 v:v:v). The deconvoluted mass spectra of both forms of the protein are shown in the inset. Reproduced with permission from Bobst et al. [37].

gas phase, enabling their MS analysis. Mass measurement of a non-covalent assembly allows in many cases its composition and stoichiometry to be determined. An example of using native ESI MS to evaluate the quaternary structure of a protein therapeutic is shown in Figure 4.4, where the two mass spectra are used to identify the quaternary structure of recombinant human arylsulfatase A (*rh*ASA) at neutral and physiological pH. This protein is a therapeutic used in enzyme replacement therapy of a lysosomal storage disease metachromatic leukodystrophy (MLD). Upon reaching its destination, the protein transitions from a homodimeric form (at neutral pH) to an octameric one (at lysosomal pH). The octamer assembly failure is one of the factors contributing to the development of MLD, as it makes the protein susceptible to degradation by lysosomal proteases such as cathepsin L. However, the early stages of the protein's processing and trafficking occur at neutral pH, where it exists as a dimer, and it is the dimeric form of arylsulfatase A that is recognized by the mannose 6-phosphate receptor and internalized via endocytosis. Native MS allows both physiologically relevant quaternary structures of *rh*ASA to be identified based on the masses of the intact assemblies (Figure 4.4), and provides strong evidence of the dimer-to-octamer transition state that occurs within a narrow pH range (6.0–7.0) [40].

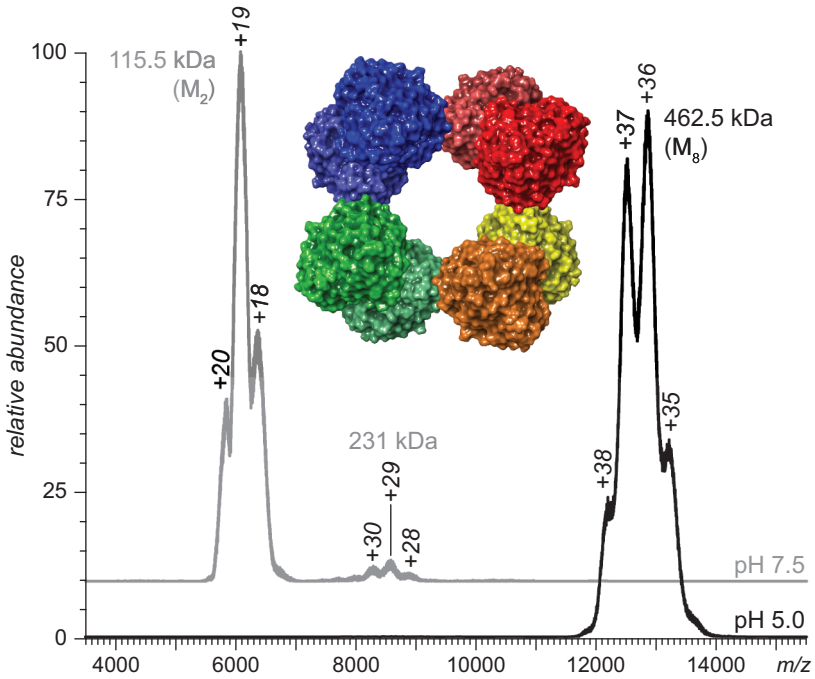


Figure 4.4: ESI mass spectra of *rh*ASA buffer-exchanged into 100 mM ammonium acetate at pH 7.5 (gray trace) and 5.0 (black trace). Protein concentration was ca. 0.9 mg/mL in each case. Adapted with permission from Abzalimov et al. [40]. The crystal structure of the octamer (pdb 1AUK) is shown at the top (the like colors are chosen to indicate pairs of monomers forming distinct homodimers) Copyright 2013 American Chemical Society.

Above and beyond characterization of the quaternary structure of protein therapeutics, native MS provides clear advantages in the studies of their interactions with therapeutic targets, such as mAb/antigen complexes [41–44]. Likewise, native MS has been used extensively to probe interactions of protein therapeutics with their key physiological partners, including the influence of the payload on recognition of a carrier protein by its cognate receptor enabling delivery via endo- and/or transcytosis [45], the influence of oligomerization and antigen binding on complement activity [43], and the influence of stress-related PTMs on the protein/protein interactions dictating the half-life of the protein drugs in circulation, such as the mAb/FcRn association [46]. Continuous improvements in MS hardware in the past decade have allowed meaningful measurements to be made for therapeutically relevant assemblies exceeding 1 MDa, such as oligomeric mAb/antigen complexes associated with complement component C1 (Figure 4.5) [43, 47].

Another type of non-covalent associations that are highly relevant to the analysis of protein therapeutics are those encountered during protein aggregation [48, 49]. The level of protein aggregation must be tightly controlled during all stages of the

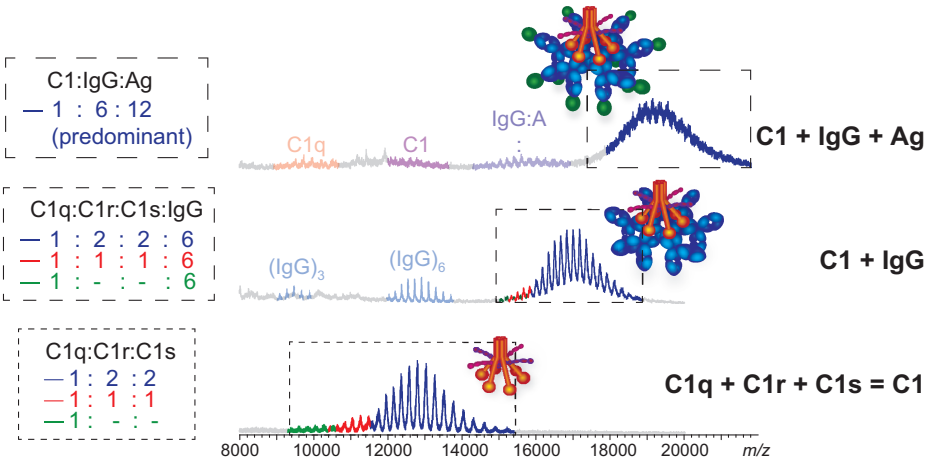


Figure 4.5: Native MS analysis of reconstituted C1, C1:IgG, and C1:IgG:Ag complexes. The signals shown in dashed boxes are color coded according to the stoichiometries of C1 assemblies, as specified to the left of the corresponding spectra. Adapted from G. Wang et al. [88] with permission from Elsevier.

biopharmaceutical product's life cycle from development to production to storage, and this task had been traditionally accomplished either by analytical ultracentrifugation [50] or (more commonly) by SEC or SEC-MALS [51, 52]. Native MS is rarely used as a means of monitoring aggregation of protein therapeutics, with its role largely limited to either off-line analysis of SEC fractions [53, 54], or online detection/identification in SEC [55] (the latter will be discussed in more detail in Section 4.2.3). One important point that must be considered when native MS is used to monitor aggregation of protein therapeutics is that the intensity of the ionic signal reflects molar concentration of the corresponding protein oligomer, rather than its weight concentration (as reported by SEC with UV detection) [54]. Indeed, the ionic current is proportional to the number of ions (meaning that in the absence of any biases and discrimination the equimolar mixture of monomers, dimers and trimers of the same protein will give rise to equiabundant ionic signals for these species). In contrast, the UV absorbance signal is proportional either to the number of peptide bonds or the number of aromatic residues (depending on the detection wavelength). Therefore, the equimolar mixture of monomers, dimers, and trimers will give rise to UV absorbance signals scaled as 1:2:3 for these three species, respectively. It is not therefore surprising that a direct comparison of the native MS and the SEC chromatogram for the same mixture of small soluble aggregates appears to suggest a significantly higher extent of aggregation reported by SEC with UV detection compared to MS alone (compare panels B and C in Figure 4.6). However, once the appropriate correction is made, the two distributions become comparable.

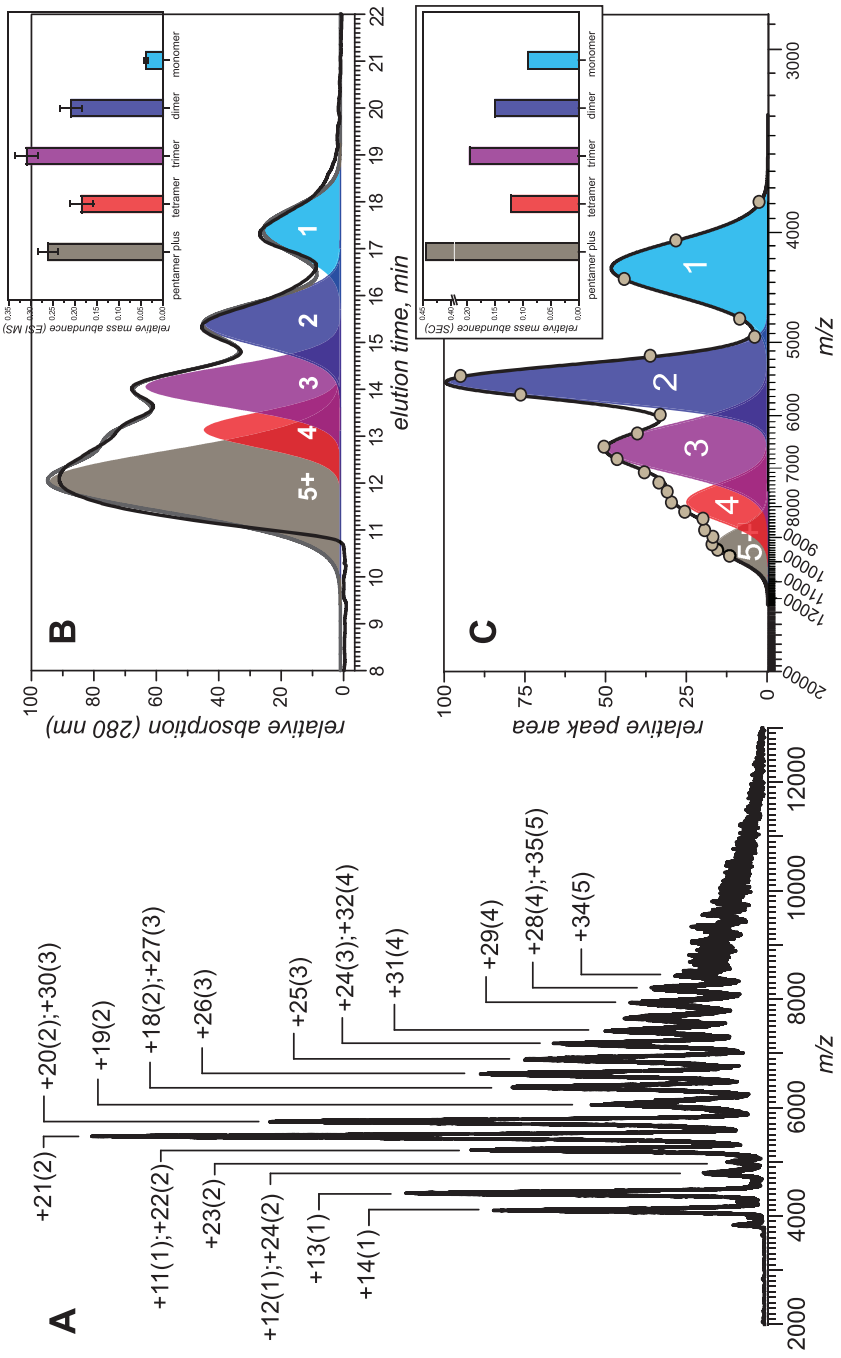


Figure 4.6: A: An ESI mass spectrum of heat-stressed recombinant human antithrombin (rhAT) (charge state assignment is shown for most peaks and numbers in parentheses indicate the size of

A unique advantage of native MS is the ability to measure simultaneously two parameters, ionic masses and their charge state distributions, making it possible to combine assessment of conformational integrity of protein therapeutics and monitoring their aggregation in a single experiment. Since the two processes are intimately linked (although the mechanistic details underlying this connection are rarely well understood), this feature of native MS can be exploited in applications seeking to understand the impact of stress conditions on the higher order structure of biopharmaceutical products. For example, temperature is a key environmental parameter that affects protein conformation, and heat stress is frequently used to assess the stability of protein therapeutics [48]. Aggregation propensity of mAbs has a strong correlation with their thermal stability [56], and differential scanning calorimetry (DSC) is frequently used to provide quantitative data on the latter [7]. Unfortunately, DSC does not provide much information on the mechanistic aspects of heat-induced conformational transitions, which could be useful not only for understanding the aggregation pathways but also for designing strategies to delay the onset of such processes. While the potential of ESI MS to probe behavior of biopolymers as a function of solution temperature had been recognized soon after the introduction of this ionization technique [57], various technical issues prevented wide adoption of temperature-controlled ESI MS measurements for at least a decade. One of the problems that is commonly encountered in such temperature-resolved studies is fast cooling of the protein solution just prior to spraying (when it flows through an unheated small-diameter capillary/needle). The fast solution cooling/temperature drop in this region (due to a large surface-to-volume ratio) frequently results in protein refolding, effectively erasing the memory of the reversible unfolding events occurring in the bulk of solution. While several approaches can be used to avoid this problem [58, 59], a great deal of care must be taken to ensure that the measurements are not affected by the heat loss/protein solution cooling in the ESI interface.

An example of using temperature-controlled ESI MS to study the behavior of a protein therapeutic under the heat-stress condition is presented in Figure 4.7. In this case the recombinant acid-β-glucocerebrosidase (already mentioned in Section 4.2.1) displays a narrow charge state distribution in the high m/z region when the protein solution is kept at room temperature. Only minimal changes are observed in the mass

Figure 4.6 (continued)
the protein oligomers). B: SEC of the heat-stressed *rh*AT (mobile phase 150 mM ammonium acetate, pH 8.0). The colored Gaussian curves represent contributions of the protein monomer (cyan) and oligomers to the chromatogram. C: Representation of the native MS data shown in panel A using a reversed m/z scale and deconvolution of this data set by finding the best fit with five Gaussian curves. The histograms shown in the insets in panels B and C show relative abundance of all detected *rh*AT species. The bar heights in the MS histogram were multiplied by the size of the corresponding oligomer (i.e., number of protein monomers within the corresponding *rh*AT species). Adapted with permission from G. Wang et al. [54]. Copyright 2012 American Chemical Society.

spectra as the temperature is increased up to 46 °C. However, ramping up the solution temperature just one degree over the DSC-determined melting point for this protein leads to a dramatic change in the appearance of the ionic signal (see the panel labeled "50 °C" in Figure 4.7). The protein ion charge state distribution has a clearly bimodal character (indicative of partial unfolding in solution), and the prominent ionic signal at m/z above 4,500 shows the presence of protein dimers and trimers. Further increase of the solution temperature (from 50 to 52 °C) results in a dramatic increase in the abundance of ionic species representing protein aggregates (dimers, trimers, and tetramers), and the size of the protein aggregates continues to grow as the solution temperature is increased (see the panels labeled "60 °C" and "70 °C" in Figure 4.7). The value of the melting temperature of recombinant acid-β-glucocerebrosidase reported by DSC is in agreement with the behavior revealed by the temperature-controlled ESI MS, but the latter offers much higher informational content, allowing both unfolding of the protein monomers and ensuing formation/evolution of the aggregates to be directly visualized as a function of the solution temperature. More recent reports show that temperature-controlled ESI MS can be used to study multi-stage transitions, such as sequential domain unfolding in mAbs [60]. A similar approach can also be used to monitor dissociation of multi-unit proteins connected by disulfide bridges as a result of the heat stress [61].

4.2.3 Native mass spectrometry of highly heterogeneous protein therapeutics

The examples of using native MS as a means of evaluating conformational integrity of biopharmaceuticals (Section 4.2.1) or determining their quaternary structure (Section 4.2.2) appear relatively straightforward. However, these tasks can become much more complicated for therapeutic proteins exhibiting a high degree of structural heterogeneity (e.g., extensively glycosylated proteins, PEGylated proteins and protein–drug conjugates). As has been already discussed in Chapter 3, ionic signals representing different charge states of such proteins can overlap, and in some extreme cases give rise to a near-continuum ionic signal in a mass spectrum. Charge state distribution analysis in such cases is frequently impossible, unless it is assisted by the tools of gas-phase ion chemistry, such as limited charge reduction [62]. This is illustrated in Figure 4.8, where the appearance of a poorly resolved mass spectrum of a stressed haptene-modified carrier protein could be interpreted in terms of a bimodal charge state distribution. This would imply that the ionic signal in the high m/z region (above m/z 3,500) represents compact conformers of the conjugate (nearly aligned with that of the intact unmodified protein), while the signal in the lower m/z region would correspond to partially unfolded states of the haptene-modified protein. Alternatively, one may argue that the signal in the lower m/z region represents lower molecular weight species, rather than partially unfolded forms of the vaccine component. Limited charge reduction provides unequivocal evidence that the ionic masses

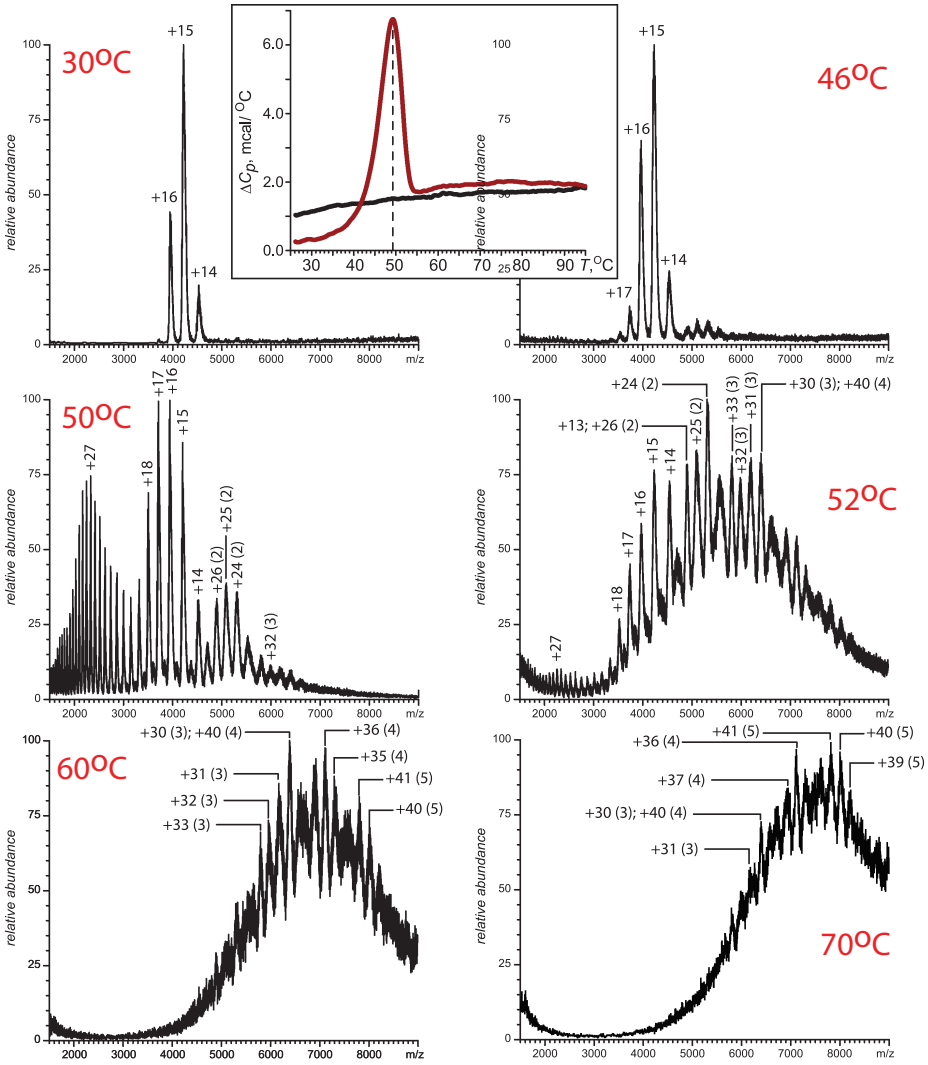


Figure 4.7: ESI mass spectra of recombinant acid-β-glucocerebrosidase (1 μM in 20 mM ammonium acetate, pH 4.7) recorded at various solution temperatures, as indicated on each panel. The numbers without parentheses indicate charge states of the protein monomers; the numbers in parentheses indicate the size of the protein oligomers (i.e., the number of monomeric units comprising the corresponding oligomer). Inset at the top: the DSC profile of recombinant acid-β-glucocerebrosidase (solution conditions are identical to those used in native MS measurements) showing a large-scale conformational transition at 49 °C (brown trace). The black trace in the inset represents a rerun of the DSC experiment for the acid-β-glucocerebrosidase sample that already went through a single cycle of DSC measurements. Adapted with permission from G. Wang et al. [58]. Copyright 2011 American Chemical Society.

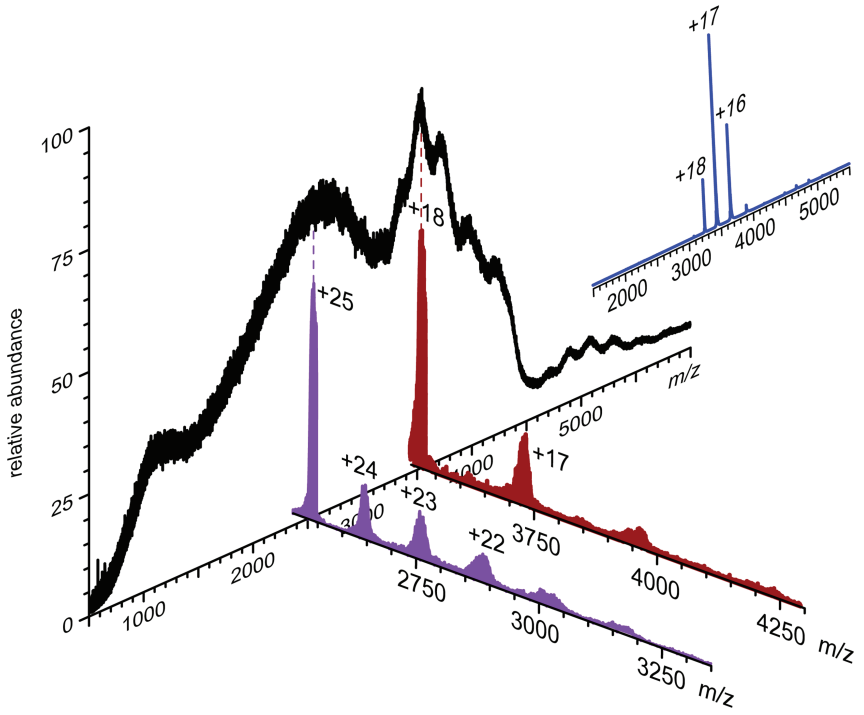


Figure 4.8: ESI mass spectrum of a stressed haptene-modified carrier protein (inactivated diphtheria toxin) with the limited charge reduction analysis of two ionic populations yielding identical masses (brown and purple, respectively). The reference mass spectrum of the intact (unmodified) carrier protein is shown in the inset. Unpublished data courtesy of Drs. C.E. Bobst (UMass-Amherst), O. Frieze and J. Perry (Pfizer).

in both subpopulations are nearly identical (62.5 kDa), which confirms that the stressed vaccine component is partially unfolded. Limited charge reduction can also be used to enable meaningful analysis of quaternary structure of highly heterogeneous recombinant proteins [40], their aggregates [54], and non-covalent complexes with physiological partners [63–65].

Crowded mass spectra can also be greatly simplified by combining front-end separations (either liquid chromatography or capillary electrophoresis) with online detection by native MS [66]. As already discussed in Chapters 2 and 3, reversed-phase LC-MS is one of the common tools in the field of the primary (covalent) structure analysis of biopharmaceutical products. Unfortunately, both mobile and stationary phases employed in the reversed-phase LC favor protein denaturation, making this type of chromatography unfit to serve as a front-end for native MS analyses. In contrast, several other types of LC, such as SEC and ion exchange (IEX), allow the biopolymer higher order structure to be preserved throughout the separation process, making them an ideal choice vis-à-vis front end separation for native MS

analyses. A wide range of commercial ESI MS instrumentation can be readily interfaced with either SEC or IEX using solvent conditions that preserve the integrity of protein higher order structures. Although the usual limitations apply (e.g., strong and non-volatile electrolytes must be excluded from the list of the mobile phase components – see Section 4.2.4 for more detail), a range of solvent systems are available that maintain near-native environment for most proteins and at the same time allow high-quality MS data to be acquired even for relatively large complexes.

Since SEC allows separation based on difference in hydrodynamic radii in ESI-compatible solvents under physiologically relevant pH and ionic strength, it is well suited for evaluating systems involving complex formation [43] and conformational changes [55]. Although in many cases the MS measurements can be performed off-line, species that may undergo rapid degradation or interconversion require online MS detection in real time. In fact, online SEC-MS lends itself as a powerful tool in the studies of protein conformational dynamics and aggregation [55], and online MS detection may provide a unique opportunity to interpret convoluted SEC chromatograms and resolve ambiguity when all other methods of detection fail to provide meaningful data. A rather trivial example would be the ability of MS to trace individual protein species that either co-elute or have overlapping chromatographic peaks, but a less trivial scenario is presented by metastable systems undergoing changes in their quaternary structure on the chromatographic timescale. The resulting SEC chromatograms are very different from those observed for non-covalent complexes that remain stable in solution on the chromatographic timescale and give rise to well-defined peaks.

An example of SEC-UV and SEC-MS analyses of a metastable system is shown in Figure 4.9, where SEC is used to determine quaternary structure of a familiar protein therapeutic arylsulfatase A (rhASA, see Section 4.2) under mildly acidic conditions (pH 6.4). As has been already mentioned earlier, this protein forms stable octamers (460 kDa) at lysosomal pH, although only dimers (115 kDa) exist at neutral pH [40] (see also Figure 4.4). Therefore, one would expect to observe a mixture of dimers and octamers at an intermediate pH level. However, SEC alone fails to provide meaningful data when used to analyze the oligomerization state of rhASA at such intermediate pH values, as the elution profiles become convoluted, and their appearance is strongly concentration dependent. Tailing of the SEC peak toward longer elution time might be indicative of limited $M_8 \rightleftharpoons M_2$ dissociation during the chromatographic run, but UV detection cannot discriminate between the two forms of the protein (black trace in Figure 4.9A). This task is readily accomplished when online detection is carried out by native ESI MS, which allows individual elution profiles to be obtained for both M_8 and M_2 species (red and blue curves in Figure 4.9A), and the appearance of these two chromatograms suggests that M_8 species dissociate as they move through the column. A schematic diagram in Figure 4.9B illustrates the behavior of this system using a 2-D plot (the species' position inside the column vs.

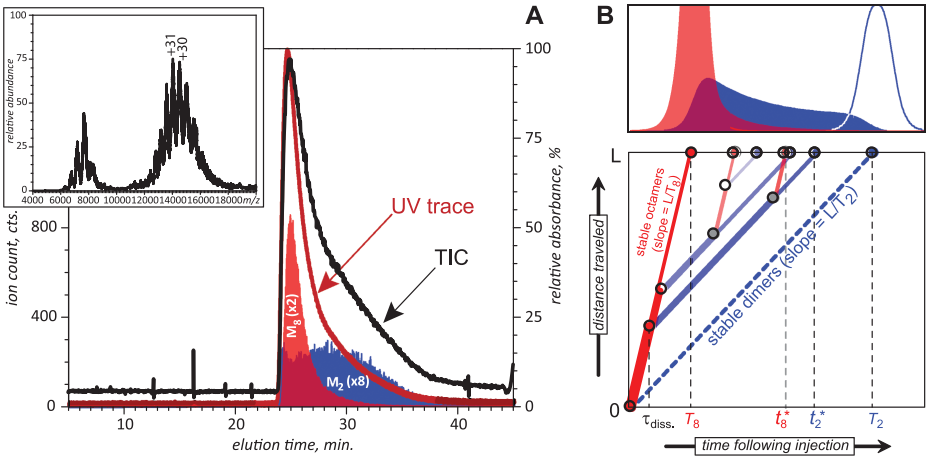


Figure 4.9: A: Online SEC-MS of 275 µg of rhASA (corresponding to 5 nmol of the octameric form) at pH 6.4: UV chromatogram (brown trace), TIC (black trace) and the cumulative extracted ion chromatograms of the dimeric (blue) and octameric (red) species of the protein. The inset shows a mass spectrum averaged across the 24–35 min elution window. B: A schematic 2-D diagram illustrating behavior of metastable M_8 species undergoing dissociation (to M_2) and re-association during the elution process. Stable oligomers would travel only along the trajectories on the far left (M_8) and far right (M_2), but the actual trajectories branch out as a result of M_8 dissociation (white circles) and M_2 re-association (gray circles). Red and blue circles show elution of M_8 and M_2, respectively. The top diagram shows a representative solution of ADR equations for the metastable M_8/M_2 system (color-filled curves); a reference solution for a non-reactive (stable) M_8/M_2 system is shown with overlaid color curves. Adapted from Kaltashov et al. [66] with permission from Elsevier.

the travel time), where the red and blue lines represent trajectories of protein octamers and dimers inside the column, and white and gray dots indicate random dissociation of the octamers and re-association of the dimers, respectively. Despite showing just a few examples of random dissociation and re-association events, this diagram is nonetheless very instructive as a means of understanding the behavior of "reactive" analytes during chromatographic analysis. Two other factors that influence the elution profiles of metastable oligomers are the diffusion inside the column and the finite duration of the sample injection, both of which result in the finite width of elution peaks even in the absence of $M_8 \rightleftharpoons M_2$ transformations. A mathematical model of the elution of the metastable complex can be constructed by taking into account both chemical (complex dissociation/re-association) and hydrodynamic processes (directed flow and diffusion) [66]. A detailed and rigorous mathematical treatment of this system is beyond the scope of this book (an interested reader is referred to the original paper by Kaltashov et al. [66]), and we will only present an end result (a system of partial differential equations that describe the evolution

of the concentrations of the two forms of the protein across the column length and as a function of time following the injection:

$$\begin{cases} \frac{\partial C_{M8}}{\partial t} = -u_{M8} \cdot \frac{\partial C_{M8}}{\partial x} + D_{M8} \cdot \frac{\partial^2 C_{M8}}{\partial x^2} - k_{diss.} C_{M8} + \frac{1}{2} \cdot k_{assoc.} C_{M2}^2 \\ \frac{\partial C_{M2}}{\partial t} = -u_{M2} \cdot \frac{\partial C_{M2}}{\partial x} + D_{M2} \cdot \frac{\partial^2 C_{M2}}{\partial x^2} + 4k_{diss.} C_{M8} - 2k_{assoc.} C_{M2}^2 \end{cases}$$

This is a greatly simplified one-dimensional version of the so-called advection–diffusion–reaction (ADR) equation [67], and a representative set of its numerical solutions is shown in the top panel of Figure 4.8B. While the numerical solutions of the ADR equations system are obviously sensitive to both the diffusion coefficients (D_{M8} and D_{M2}) and the dissociation/re-association rate constants ($k_{diss.}$ and $k_{assoc.}$), a visual inspection of the solutions presented in Figure 4.8B suggests that they adequately reflect the behavior of the real system, as observed experimentally with SEC-MS (Figure 4.8A). While finding the numerical solutions of the ADR equations system describing a specific metastable system requires the knowledge of all relevant parameters, a relatively straightforward qualitative analysis using a 2-D "distance vs. time" plot can be readily carried out in many cases, enabling meaningful interpretation of the most convoluted SEC chromatograms.

Ion exchange is another chromatographic format that can be interfaced with native MS [68]. While it also allows protein conformational integrity to be probed [68] (based on the ionic charge state distributions), most current applications of IEX-MS are still focused on rather mundane tasks of detection/characterization of charge variants of protein therapeutics (as already discussed in Chapter 3). An orthogonal approach to dealing with highly heterogeneous protein therapeutics is offered by ion mobility (IM) spectroscopy, a sister technique to MS, which is almost always used in combination with the latter in bioanalysis [69] (Chapter 2). Above and beyond the ability to separate isobaric ions based on differences in their shapes in the gas phase, IM measurements allow ionic collisional cross sections (CCSs) to be determined. The CCS values of biopolymer ions reflect their gas-phase conformations, naturally leading to a suggestion that the higher order structure analysis of proteins in general and protein therapeutics in particular should include IM in its arsenal [70]. However, application of IM-based methods to characterize protein conformations raises a range of questions (summarized in a recent review [71]), which undermine the value of the information deduced from such studies. The most serious issue is related to the phenomenon known as a conformational collapse or compaction in the gas phase. Non-globular proteins (such as antibodies and other proteins containing flexible hinge regions) are especially prone to the gas-phase compaction. This phenomenon leads to a significant underestimation of the physical size of solution-phase conformations when relying solely on the CCS values derived from IM measurements [71]. Another potential caveat in using IM to assess the integrity of native (solution-phase)

higher order structures relates to the fact that these measurements are not sensitive to the secondary structure, and compactness of the protein ions in the gas phase does not guarantee their assuming correct (native) conformations [71]. In contrast to this mixed record as a conformational analysis tool, IM has been highly successful as a means of ion separation in the analysis of complex and heterogeneous systems. Importantly, the separation dimension enabled by IM is both orthogonal and complementary to native MS and native LC-MS [71]. Availability of this unique separation dimension also proved very useful in facilitating other methods of analysis of highly heterogeneous macromolecular medicines, such as limited charge reduction [65], allowing the scope of the latter technique to be dramatically expanded to include macromolecular medicines as heterogeneous as heparin (to be reviewed in Chapter 8).

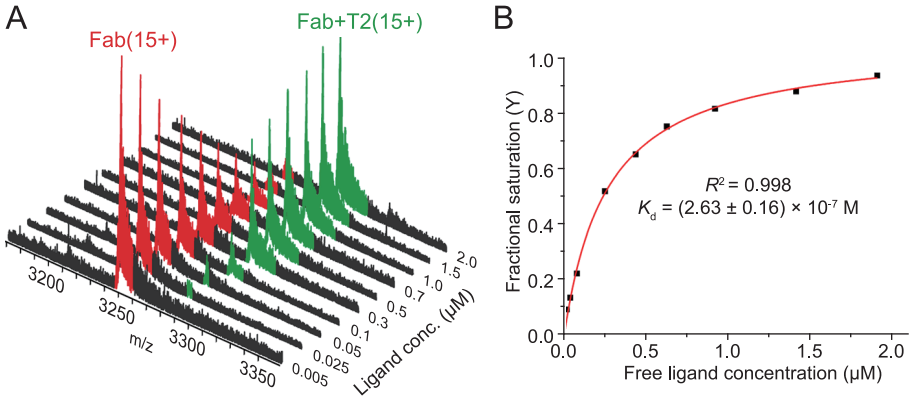


Figure 4.10: Direct ligand titration of a Fab fragment of anti-thyroxine (T4) antibody with a thyroxine analog T2. A: Native ESI mass spectra (+15 charge state) of 0.1 μM anti-T4 Fab with varying T2 concentrations measured in 20 mM ammonium acetate (pH 6.8). B: Titration plot (fractional saturation vs. free ligand concentration). The solid red line represents the best fit to the specific one-site binding model. Reproduced with permission from Thangaraj et al. [74]. Copyright 2019 American Chemical Society.

4.2.4 Can native MS be used to provide quantitative information on interactions between protein therapeutics and their targets?

One question that frequently arises when native MS measurements are discussed concerns the feasibility of extracting quantitative affinity data from such experiments. A relatively straightforward (but certainly not the most accurate) approach to this problem equates the relative ionic signal abundance of all involved species (e.g., the protein/ligand complex and its constituents in the unbound form) with

their fractional concentrations in solution, followed by K_d calculation for the protein/ligand complex. The validity of the basic underlying assumption, however, is far from certain, as the ionic signal in ESI is not determined solely by the protein concentration, and the response factors do influence the calculated K_d values [72]. These response factors can be determined in many cases [73], allowing the necessary corrections to be made in calculating the K_d values. Alternatively, one can use reference ligands, for which the K_d values are known, to enable measurements for other (structurally similar) ligands. A typical experiment involves incremental additions of the ligand to the protein solution (titration) and recording the changes in the relative abundance of relevant ionic signals, as illustrated in Figure 4.10 for the association between the Fab fragment of anti-thyroxin antibody and its antigen and structurally similar molecules [74]. Further improvements to the measurements protocol can be made by using SEC-MS as a platform [75], but the utility of this methodology is limited to probing interactions of protein therapeutics with small-molecule ligands. SEC-MS also enables quantitative assessment of transient protein/protein interactions, but evaluation of the binding strength in this cases is based on observing the kinetic processes within the column (as discussed in the preceding section, see also Figure 4.9), and the scope is limited to the relatively low affinity range (K_d values exceeding 10 nM) [66].

An orthogonal way to evaluating the binding affinity for a range of protein/protein interactions is offered by affinity chromatography (AC), which is now available in the high-pressure format and can be interfaced with ESI MS [76–78]. Using MS as an online detector for AC (AC-MS) allows the affinity values to be compared among multiple proteoforms of the protein therapeutic based on the differences among their distinct elution profiles (extracted ion chromatograms, as illustrated in Figure 4.11), although the absolute affinity values cannot be obtained. Although the use of an AC column is obviously limited to a particular ligand, it can be used for a variety of tasks ranging from evaluation of the impact of various PTMs on the mAb affinity toward a specific Fc receptor to resolving the glycosylation heterogeneity of mAbs at the intact protein level (Figure 4.11).

4.2.5 What needs to be considered at the planning stage and/or when analyzing the results of native MS measurements

One important consideration that should always be kept in mind when planning and/or analyzing the results of native MS measurements is the limited repertoire of ESI-compatible solvents that can be used in such experiments. As has been already discussed in Chapter 2, high-quality ESI MS data can be acquired only when volatile electrolytes are used to obtain the desired pH and ionic strength of the protein solution. Unfortunately, this means that most native MS measurements must be carried out in ammonium acetate and ammonium formate solutions (the use of ammonium

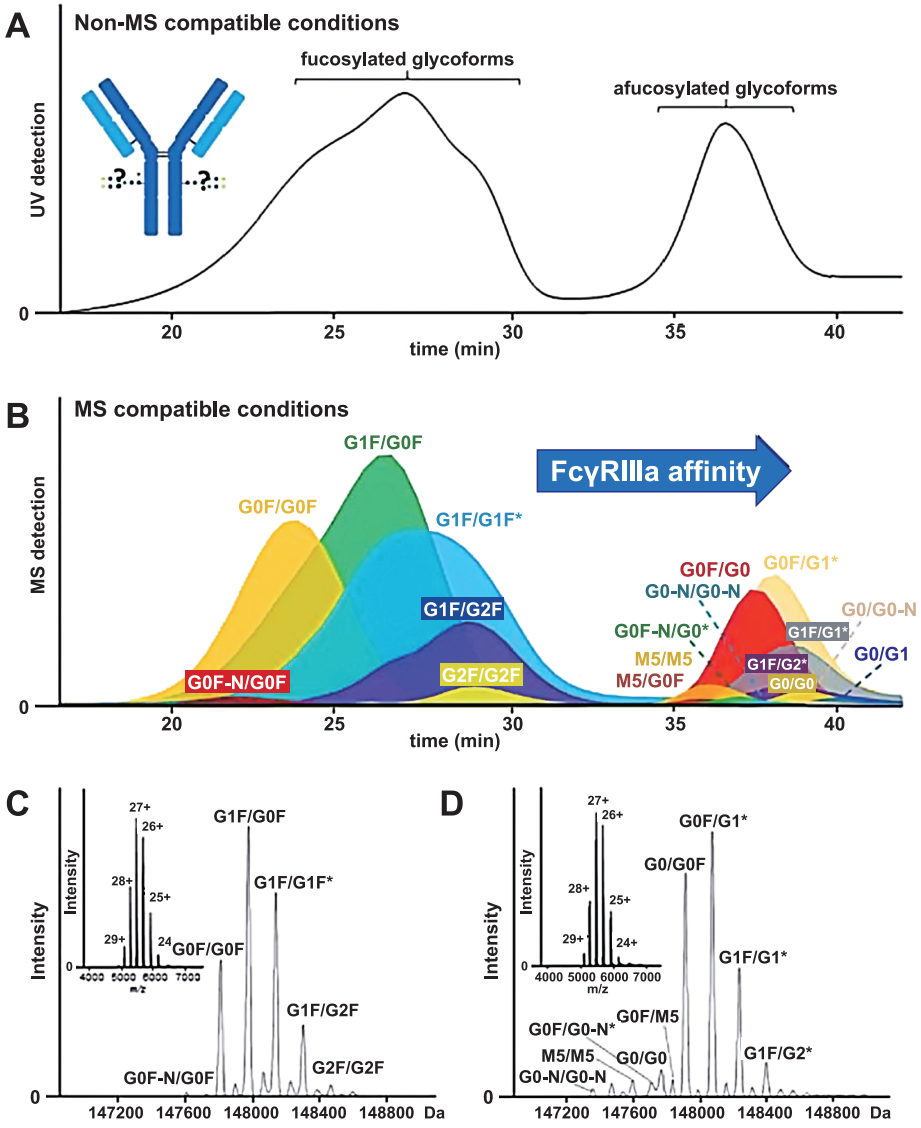


Figure 4.11: FcγRIIIa affinity chromatography for a therapeutic mAb. (a) UV chromatogram using reported non-MS compatible conditions. (b) AC-MS under MS-compatible conditions represented by extracted ion chromatograms of detected glycoforms. (c and d) Deconvoluted mass spectra and charge state distribution (inserts) of (c) 2× fucosylated (18–34 min) and (d) remaining glycoforms (34–42 min). In case of multiple possibilities (asterisk), the most probable glycoform is presented, based on reference data. Reproduced from Lippold et al. [77] with permission from Taylor & Francis Ltd (www.tandfonline.com).

bicarbonate – another ESI-compatible electrolyte – should be avoided, as it can trigger protein denaturation within ESI-generated aerosol particles [79]). The list of ESI-compatible salts can be expanded by using alkylated ammonia and substituted acetic acid (e.g., methylammonium dichloroacetate, $N(CH_3)H_3CHCl_2CO_2$) [80]. Although the use of the absolute majority of salts and other excipients that are common in biopharmaceutical formulations is certainly ruled out, presence of some non-volatile buffers in protein solutions can be tolerated by adding large amounts of ammonium acetate (to a final concentration ranging from 0.4 to 1.0 M) [81, 82]. High-intensity salt cluster background arising due to the presence of non-volatile electrolytes can also be reduced in native MS using specialized signal processing techniques [83], although this approach has not yet become routine in native MS. Despite the limitations on the solvent composition in native MS, the relative ease of the analysis of the ionic charge state distributions in many cases warrants its application as a means of the initial assessment of conformational integrity. Depending on the results, such measurements can be followed by more involved methods of analyses, such as the H/D exchange (Section 4.3). Lastly, it must be remembered that in most cases charge state distribution analysis allows only relatively large-scale conformational transitions to be detected; and the results of such analysis should not be over-interpreted. Indeed, even though a carefully designed experiment may allow small-scale transitions to be observed [80], the results may be affected by a range of gas-phase processes, including charge transfer reactions [30] and dissociation of non-covalent protein complexes via asymmetric charge partitioning mechanism [84].

Another important consideration related to native MS measurements discussed in Section 4.2.2 is that non-covalent assemblies remain stable in the gas phase only if they are maintained primarily by local electrostatic interactions (salt bridges) and hydrogen bonding. The protein/protein interactions driven primarily by hydrophobic forces become highly unstable in the absence of the solvent, which in most cases causes dissociation of these assemblies in the gas phase. Fortunately, in many cases the gas-phase dissociation proceeds via a very distinct mechanism (asymmetric charge partitioning [85, 86]) which generates an unambiguous signature in the mass spectrum and allows data misinterpretation to be avoided [84], although the existence of multiple dissociation pathways may complicate the data analysis. Importantly, preservation of the non-covalent complexes requires that the ion optics in the ESI interface be tuned in such a way as to minimize collisional activation of the ionized non-covalently bound associations in order to prevent their dissociation. The ion source temperature should also be set significantly below the levels that are used in ESI MS measurements described in Chapter 3.

Unfortunately, these necessary measures are not always sufficient to preserve the integrity of large metastable non-covalent complexes in the gas phase, unless the excess of internal energy is removed via collisions with the residual gas molecules (the phenomenon known as “collisional cooling” [87]). The latter is induced

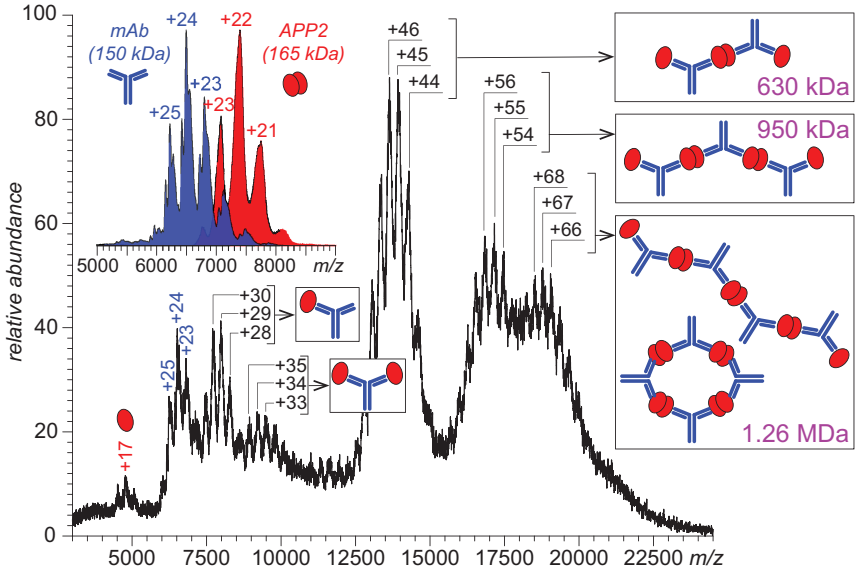


Figure 4.12: Native MS of large immune complexes produced by mAb and a bivalent (homodimeric) antigen, ectodomain of aminopeptidase P2 (APP2). The mass spectra of the antigen and the mAb are shown in the inset. Data courtesy of Y. Yang, a graduate research assistant at UMass-Amherst.

by introducing sufficient quantities of the "cooling" gas (typically N_2) post ion production, which cannot be implemented on many generic ESI MS instruments. Therefore, successful native MS measurements of large non-covalent complexes (of the type relevant for protein therapeutics) require either the use of dedicated instruments that have a built-in collisional cooling function or the hardware modification. The inability (or failure) to implement effective collisional cooling is arguably one of the most common reasons for the native MS failing to detect large non-covalent assemblies.

Lastly, in all examples shown in the preceding sections, the ionic signal in native MS extended into high m/z region (above 4,000 u). The reason for this is the relatively low number of charges that can be accommodated on the surface of compactly folded proteins. In fact, it is not uncommon to see native mass spectra of large multi-unit protein therapeutics or their complexes with physiological partners with the ionic signal populating the very high m/z range (>10,000 u) [43, 44, 88] See also an example shown in Figure 4.12. Unfortunately, this means that the most affordable mass spectrometers (based on the quadrupole mass analyzers) are poorly suited for native MS work; this also needs to be taken into consideration at the experiment planning stage.

4.3 Hydrogen deuterium exchange (HDX) MS

Native MS discussed in the preceding section can be a very effective tool for detecting an offset of unfolding/aggregation of protein therapeutics, as well as monitoring their interactions with a range of physiological partners. However, it rarely provides detailed structural information (e.g., it cannot identify specific segments within the protein that are affected by partial unfolding or participate in binding to its partners). This information can be provided in many cases by hydrogen/deuterium exchange (HDX) with MS detection. Another important advantage of HDX MS is its tolerance for a range of solvent systems that cannot be used in native MS measurements (as discussed in the preceding section). This allows the behavior of protein therapeutics to be studied under conditions that closely mimic or indeed duplicate either their physiological environment or the drug formulation. If executed properly, HDX MS measurements can provide a wealth of information on protein higher order structure and dynamics, and it is not surprising that in a mere decade that followed the first use of HDX MS to characterize a protein therapeutic [21] this technique had become widely adopted in the biopharmaceutical industry and is now routinely used not only in exploratory studies [89] but also in filings with regulatory agencies [90].

4.3.1 Global HDX MS measurements to monitor conformational integrity of protein therapeutics

In HDX MS experiments the protein higher order structure is probed by following the exchange of labile hydrogen atoms with the solvent. Protons that are involved in hydrogen bonding or sequestered from the solvent in the protein interior cannot be exchanged readily, unless they become exposed to the solvent molecules via dynamic events that affect the protein conformation either locally (e.g., via structural fluctuations) or globally (e.g., via transient unfolding events). The exchange reactions are usually initiated by diluting the protein solution in the appropriately buffered D_2O, triggering the exchange processes for all labile hydrogens that are exposed to the solvent (Figure 4.13). The chemical exchange rate (also called "intrinsic exchange rate") varies significantly among different types of labile hydrogens, and also strongly depends upon the solution pH (Figure 4.14). The exchange is quenched after a certain period of time by quickly acidifying the protein solution to pH 2.5–3.0 and lowering its temperature to 0–4 °C. This results in a dramatic deceleration of the intrinsic exchange reactions of the backbone amide hydrogen atoms with the solvent, while the exchange rates of other labile hydrogen atoms (those at the side chains and at the polypeptide termini) remain relatively high (Figure 4.14). As a result, the isotope labels acquired prior to the quench step are kept exclusively at the backbone amides (despite the loss of the higher order structure – and protection – that is inevitable for

nearly all proteins in the acidic environment). This provides a single reporter for each amino acid residue in the protein sequence with the exception of the N-terminal residue and all proline residues. Since each individual exchange event (substitution of a hydrogen atom with a deuterium atom) results in a mass increase of 1.01 Da, the total increase of the protein mass can be equated to the total number of amide groups that have acquired an isotopic label prior to the quench step. Measuring the protein mass increase as a function of the time interval length between the dilution step (initiation of the HDX reactions) and the quench step allows the kinetics of the exchange reactions to be followed at the intact protein level (see the "global HDX" box in Figure 4.13). Such measurements provide a relatively straightforward way to detect changes in the stability of the higher order structure as a result of certain modifications of the protein covalent structure and to evaluate their extent [21]. Importantly, the HDX step can be carried out in the formulation buffer if it is followed by rapid desalting of the protein under the slow exchange conditions (e.g., using either fast LC or solid-phase extraction approaches); the ability to evaluate conformational stability of protein therapeutics in relevant solvents/buffers provides a significant advantage over the native MS methods discussed in Section 4.2.

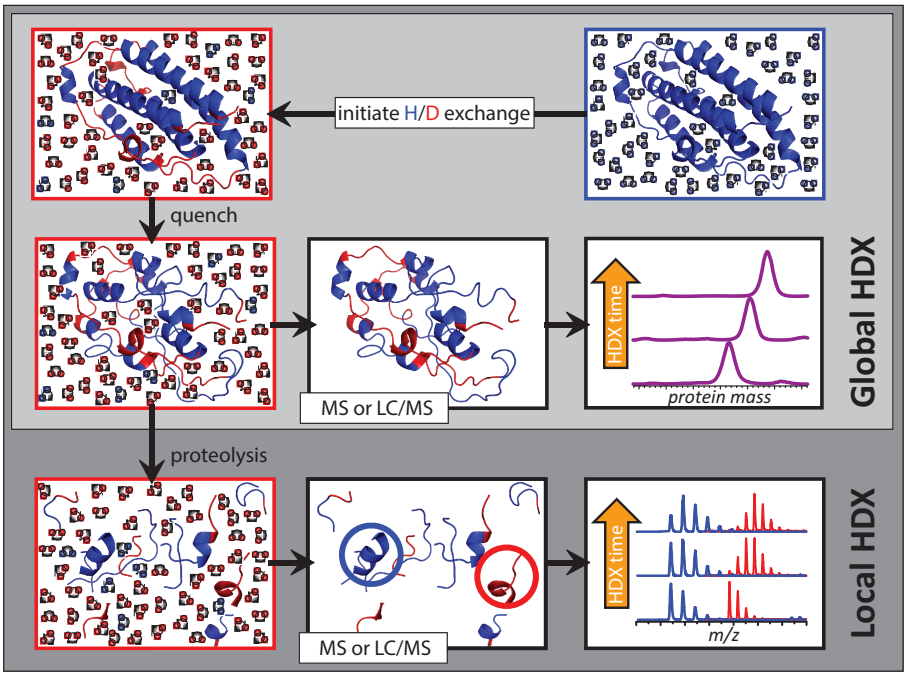


Figure 4.13: A schematic diagram of the workflows for global and local HDX MS measurements. Reproduced from Kaltashov et al. [9] with permission from Elsevier.

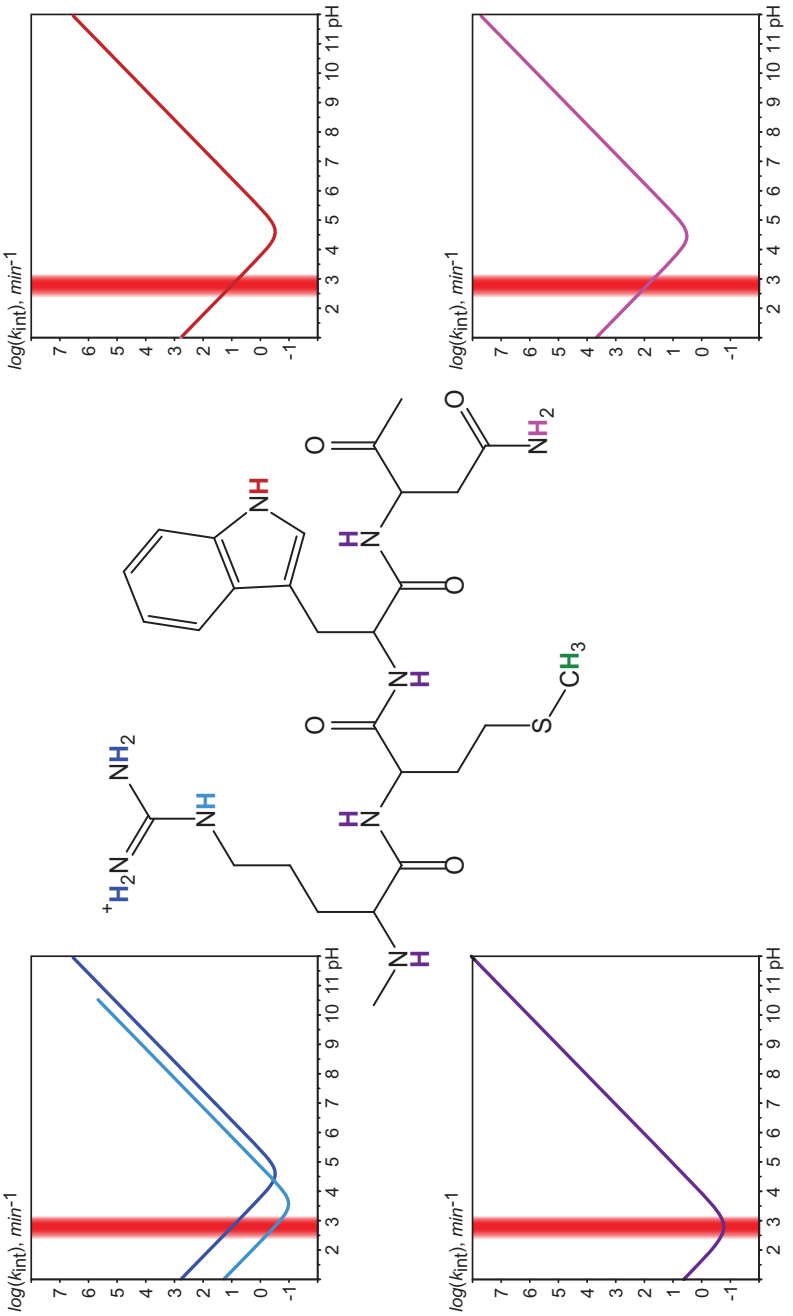


Figure 4.14: Intrinsic (chemical) exchange rates for different types of labile hydrogen atoms within polypeptides. The red stripe indicates the pH range used to quench the exchange of backbone amide hydrogen atoms.

4.3.2 Site-specific HDX MS measurements to identify instability hot spots

The ability to minimize the exchange of backbone amide hydrogen atoms under conditions when they are not afforded any protection by the protein higher order structure also allows the distribution of the labile isotopic labels along the polypeptide backbone to be determined by carrying out proteolysis prior to MS analysis (see the "local HDX" box in Figure 4.13). Pepsin is one of the very few proteases that remain active within the pH range 2.5–3.0, and is most commonly used to probe backbone amide protection following the quench of the exchange reactions [91, 92]. However, the slow exchange of the backbone amide hydrogen atoms continues even under the quench conditions, and the proteolytic step must be relatively fast in order to avoid the excessive loss of the deuterium labels (the so-called back-exchange [93]). The efficiency of the proteolytic step carried out under such sub-optimal conditions can be enhanced by reducing the disulfide bonds prior to the protein digestion. However, the need to maintain the slow-exchange conditions during the disulfide reduction places significant restrictions on the repertoire of reducing agents (with TCEP being the most common reagent used in such applications [94]). The proteolysis efficiency can also be enhanced by a judicial use of strong chaotropes, such as urea and guanidinium chloride [95]. Elimination of all ESI-incompatible components from the protein digest solution is usually achieved during a quick LC separation (followed by online MS analysis of the deuterium content of all produced peptic fragments). In addition to desalting, the LC step provides an advantage of reducing spectral crowding and, therefore, increases the likelihood of detecting difficult-to-ionize peptides (which frequently remain invisible in the ESI mass spectra of multi-component systems due to the signal suppression effects). The downside of using the LC step is the frequently encountered inability to detect small and highly hydrophilic peptides, as well as glycopeptides, which have poor retention characteristics on reversed-phase columns.

An example of using HDX MS to localize unstable elements within a protein [21] is illustrated in Figure 4.15. Here the evolution of the isotopic distribution of a peptic fragment (88–102) derived from intact (unmodified) interferon β1a reveals a rather anemic uptake of deuterium atoms within this segment of the protein, a hallmark of a stable conformation. However, the behavior of the same peptide derived from the protein bearing a single covalent modification (at a cysteine residue that is distal to this peptide in the protein sequence, but proximal in the three-dimensional structure) is markedly different. The bimodal shape of the isotopic distribution of the peptide ions in this case is indicative of the presence of two conformations in solution, one of which appears to lack significant backbone protection (red traces in Figure 4.15). In contrast, most other peptic fragments do not reveal a difference between the two forms of the protein, suggesting that the covalent modification does not compromise conformational integrity of the protein uniformly across the entire polypeptide chain. Instead, only several segments within the protein are

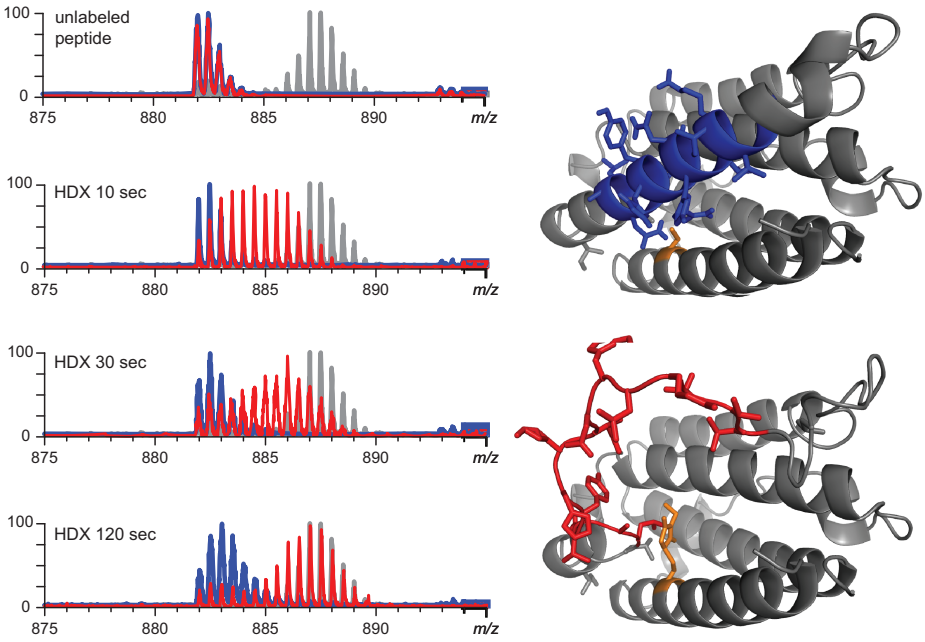


Figure 4.15: Localization of an instability hot spot within the covalently modified (NEM-alkylated) interferon β1a using HDX MS. The left-hand side panel shows the evolution of isotopic distributions of peptic fragments (88–102) derived from the intact (blue) and modified (red) proteins as a function of the exchange time in solution. The end point of the exchange reaction (taking into account the back-exchange) is indicated with a gray trace (isotopic distribution of a fully exchanged peptide). The panels on the right-hand side show location of the (88–102) segment (colored in blue) within the crystal structure of the intact protein (1AU1). Alkylation of Cys-17 (orange) inevitably leads to steric clashes within the native structure, which can be removed by unfolding the helical element containing the (88–102) segment (colored in red). Adapted with permission from Kaltashov et al. [96]. Copyright 2008 American Chemical Society.

affected, but the presence of even a small number of such instability hot spots results in a significant increase of the aggregation propensity [96].

The strategy illustrated in the example presented in the preceding paragraph can also be applied to detect/localize the instability regions that arise due to stress-related non-enzymatic PTMs targeting multiple sites within the protein, such as oxidation [37, 97–99], deamidation [100, 101], and glycation [102]. HDX MS can also be used to evaluate the conformational changes within protein therapeutics induced by non-covalent effectors of the higher order structure, such as small-molecule antimicrobials [103]. Influence of enzymatic PTMs (such as glycosylation) on the higher order structure integrity of a range of biopharmaceuticals can also be probed [104, 105], providing an important feedback for optimization of the design and production stages. HDX MS has also been used as a means of verifying the integrity of the higher

order structure following introduction of a "designer PTM," such as PEGylation [106] and small-molecule drug conjugation [107].

One important aspect of the HDX MS measurements as applied to detection of conformational changes within protein therapeutics is their sensitivity. The latter can be defined in two different ways, both of which are highly relevant in terms of the higher order structure characterization: (i) the smallest fraction of the conformationally perturbed protein molecules that can be detected and (ii) the smallest segment of the protein exhibiting physiologically and therapeutically relevant conformational changes. While the second aspect is highly specific for each product and requires an individualized approach in each case, the first one had been evaluated for several systems, yielding sensitivity levels in the 1–5% range [97, 108], surpassing those afforded by the traditional methods of biophysical characterization.

4.3.3 Site-specific HDX MS measurements to localize binding interfaces

Another important area of HDX MS applications is mapping binding interfaces in complexes formed by therapeutic proteins with their targets and physiological partners. The underlying assumption here is that association with a binding partner affords additional backbone protection to the protein segments within the binding interface, as illustrated in Figure 4.16 for the interaction of the recombinant form of human serum transferrin (Tf) and the ectodomain of its cognate receptor (TfR). The rates of deuterium incorporation into the protein backbone vary greatly across the entire complement of peptic fragments, but only few of them exhibit a noticeable difference in the peptide-level exchange kinetics between the datasets acquired in the absence and in the presence of the receptor. Specifically, Tf peptic fragment (71–81) shows robust isotope exchange in the absence of TfR within a minute of exposure to D_2O, but the deuterium incorporation becomes nearly halted on this timescale in the presence of the receptor [109].

A conclusion that naturally follows such observations is that the protein segments displaying the difference in exchange kinetics are parts of the binding interface. This is true in the specific example presented in Figure 4.16, for which the available X-ray crystallographic data confirm participation in the receptor binding for several residues from the Tf segment (71–81). However, one must remember that protein/protein associations frequently give rise to allosteric interactions that may alter conformational dynamics in protein segments that are distal to the binding interface. The occurrence of such effects may obviously complicate the interface localization efforts based on the HDX MS data; in fact, it is nearly impossible in many cases to distinguish the changes of the isotope exchange kinetics due to the solvent exclusion from the interface area upon protein/protein association from those induced by allosteric interactions. One possible approach to this problem makes use of HDX measurements on short (sub-second) timescales, with the binding partner

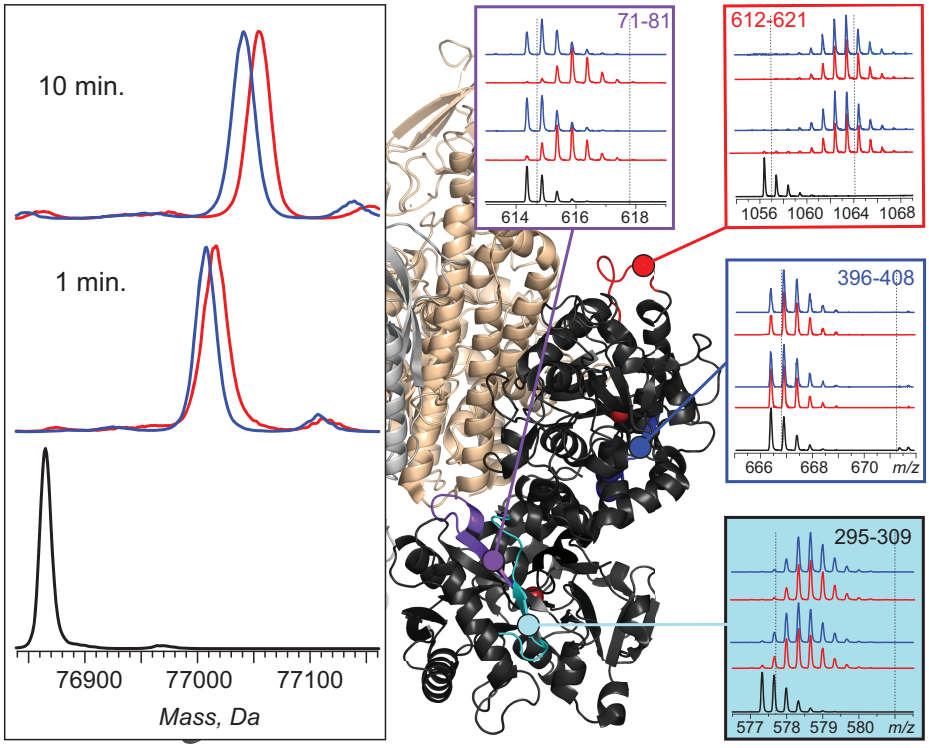


Figure 4.16: Influence of the receptor binding on the backbone protection of the recombinant form of human serum transferrin (Tf) and localization of the receptor interface with HDX MS. Left panel: HDX MS of Tf (global exchange) in the presence (blue) and the absence (red) of Tf receptor (TfR). The exchange was carried out by diluting the protein stock solution 1:10 in exchange solution (100 mM NH_4HCO_3 in D_2O, pH adjusted to 7.4). The black trace shows unlabeled (essentially deuterium-free) protein. Right panel: isotopic distributions of representative peptic fragments derived from Tf subjected to HDX in the presence (blue) and the absence (red) of the receptor. Dotted lines indicate deuterium content of unlabeled and fully exchanged peptides. Colored segments within the Tf/TfR complex show localization of these peptic fragments (based on the low-resolution structure of Tf/TfR). Adapted with permission from Kaltashov et al. [109] Copyright 2009 American Chemical Society.

being introduced simultaneously with the exchange buffer [110]. Synchronizing the onsets of the exchange reactions and the protein/protein association process allows the binding interface to be identified based on alteration of the exchange kinetics, while keeping the allosteric interactions below the detection threshold (presumably due to the longer time required for "propagating" the allosteric changes through the bulk of the protein) [110].

Availability of crystal structures of at least some components of a protein therapeutic is undoubtedly a boon for interpreting HDX MS data, particularly when the latter is used to identify binding interfaces/epitopes. If the crystal structures are not

available (which is the case for the majority of protein therapeutics), HDX MS data interpretation can be aided by molecular modeling, particularly in complicated cases, such as those involving discontinuous epitopes and size variants [111, 112].

4.3.4 HDX MS to probe aggregation of protein therapeutics

The centrality of the aggregation phenomenon to the issue of safety of protein therapeutics had fueled extensive efforts in the past decade to understand its molecular mechanisms. Specifically, the ability to identify the initiation regions of the aggregation processes within a given biopharmaceutical product is invaluable for the design of safe and effective constructs. HDX MS has been at the forefront of these efforts, mostly focusing on localization of the initiation regions [113]. One issue that greatly complicates the HDX MS work with protein aggregates is their dynamic character. Indeed, formation of the early-stage oligomers is usually a reversible process, and the equilibrium in such systems is inevitably upset by the very first step in the HDX MS work flow, protein dilution in D_2O (Figure 4.14). Several approaches are currently evaluated as a means of dilution-free HDX MS measurements, with the dialysis-based HDX [114] appearing to be particularly promising. Another intriguing approach to probing higher order structure of protein therapeutics which seems well-suited for the analysis of aggregates is the solid-state HDX MS [115]. This technique (which presently remains at the investigational stage) is designed to interrogate the structure of lyophilized proteins.

4.3.5 What needs to be considered at the planning stage and/or when analyzing the results of HDX MS measurements

From the early days of HDX MS work, multiple protocols were used by different groups to carry out the measurements and to interpret the results. The experimental variables were spread across the entire workflow (beginning with the selection of the dilution ratio to initiate the exchange reactions and ending with the duration of the time interval during each the protein and its proteolytic fragments are kept under the slow exchange conditions). These differences still persist with many laboratories using unique protocols, although in the past decade there was a steady movement towards harmonization of the ways the HDX MS work is executed and the results are reported. This has culminated in publication of a paper "Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments" endorsed by a large number of HDX MS practitioners in both industry and academia [116], and the readers are referred to this set of specific guidelines that cover every step of both the experiment and the data analysis. The NISTmAb (already mentioned in Chapter 3 as a standard reference

material for testing/validating methods of covalent structure characterization) is a useful resource for evaluation of HDX MS protocols, although the recent inter-laboratory study spearheaded by NIST revealed significant differences in the measurement outcomes. The continuous trend towards HDX MS automation [117] and acceptance of the unified guidelines [116] are likely to bring more consistency to the HDX MS work, but at least some differences are likely to persist for a long time.

4.3.6 Spatial resolution in site-specific HDX MS measurements and methods to improve it

The measurements discussed in the preceding section produce segment-specific (as opposed to "site-specific") information on deuterium distribution across the polypeptide backbone, and it is the length of the proteolytic fragments that limits the spatial resolution. The very few proteases that retain activity under the slow exchange conditions (pH 2.5–3 and low temperature) frequently result in generating relatively long (>10 amino acid residues) peptide fragments, and in many instances significantly longer segments are generated.

At least some improvement in the spatial resolution can be achieved by analyzing the deuterium content of overlapping proteolytic fragments, but the residue-level resolution remains out of reach for the majority of biopharmaceutical products. Supplementing the experimental scheme depicted in Figure 4.14 with the gas-phase fragmentation of the peptide ions is another strategy that may in some cases reveal the backbone protection patterns at a single residue level [118, 119]. Unfortunately, the choice of ion fragmentation techniques that can be used for such a task remains very limited due to the need to avoid hydrogen scrambling during the ion activation process [120, 121]. Consequently, the progress in this field has not moved much beyond feasibility studies carried out with small model proteins that are of little relevance vis-à-vis the needs of the biopharmaceutical sector.

An alternative approach to obtaining site-specific information on deuterium distribution along the polypeptide backbone utilizes the so-called top-down strategy, where proteolytic degradation is completely eliminated from the experimental work flow and is replaced with fragmentation of intact protein ions in the gas phase [109]. Once again, the need to suppress/eliminate hydrogen scrambling during the protein ion activation process places a very stringent limitation as far as the type of ion fragmentation techniques that can be employed in such measurements. Both electron capture [122] and electron transfer [123] dissociation can be used for this purpose, as long as any collisional activation of protein ions (which is frequently used to increase the dissociation yield) is completely avoided. More recently, ultraviolet photodissociation had also been shown to be a viable choice for the top-down HDX MS work [124]. While the top-down HDX MS is not by any means a routine tool in the

analysis of protein therapeutics, the extensive efforts invested in its development in the past decade appear to begin bearing fruit, with several reports on using top-down HDX MS to characterize proteins as large as mAbs now available [125, 126]. Another recently introduced approach that may become a helpful tool for probing conformational dynamics of mAbs and related biopharmaceuticals in a residue-specific fashion (dubbed "middle-down" HDX MS) takes advantage of the restricted use of pepsin (known to act specifically on immunoglobulin molecules to separate Fc from Fab regions) prior to the gas-phase dissociation of the resulting large proteolytic fragments [127].

Another issue that is frequently encountered in HDX MS measurements is inadequate sequence coverage (i.e., the presence of gaps in the peptic peptide maps). One particularly common reason for this is the desalting step that frequently precedes the fast LC separation of the peptic fragments prior to MS detection, resulting in the loss of highly hydrophilic peptides. In addition to very short hydrophilic peptides, it frequently leads to the loss of glycopeptides. Even if such glycopeptides are retained on the column, the intrinsic heterogeneity of glycans results in a dramatic reduction of the ionic signal that can be used to analyze the deuterium content within the corresponding segments of the protein. Removal of glycans from such peptides can solve both of these problems (poor retention and split-signal), but carrying out protein deglycosylation prior to the H/D exchange is likely to affect its conformation, defeating the purpose of the measurements. One elegant way to circumvent this problem involves the use of PNGase A after completion of the isotope exchange step (and concurrently with pepsin digestion under the slow exchange conditions) [128]. This enzyme is effective in removing all three types of N-glycans form glycopeptides (but is less effective when dealing with intact folded proteins), and remains active at pH as low as 2.5. More recently, another acidic glycosidase, PNGase H+, was demonstrated to be suitable for this purpose as well, and in fact showed greater tolerance vis-à-vis the acidic reducing agent TCEP, which is employed during the digestion step under the slow exchange conditions [129]. Unfortunately, O-glycosylation presents a noticeably less tractable problem due to the absence of enzymes that can remove this type of glycans from the proteins and/or peptides; however, these glycans are not encountered within biopharmaceutical products as frequently as the N-glycans.

4.4 Covalent labeling methods

HDX MS has become a commonly accepted tool for characterization of therapeutic proteins in recent years, but the search for alternative/complementary methods of analysis of protein higher order structure continues. One of the most annoying limitations of HDX is the labile character of the isotope label both in the solution phase (where the need to minimize the back-exchange dictates the solution conditions

that must be used for protein processing prior to MS detection – frequently leading to sub-optimal proteolysis, disulfide reduction, peptide separation, etc.) and in the gas phase (where the specter of hydrogen scrambling precludes the use of highly efficient methods of peptide and protein ion fragmentation). The realization that the ability to label the unprotected/solvent-exposed segments of the protein irreversibly would be a boon to the field incentivizes the relentless search for experimental strategies based on chemical labeling techniques.

4.4.1 Chemical labeling and cross-linking: what limits their use in characterization of biopharmaceutical products?

Chemical labeling [130] and cross-linking [131] are among the oldest biophysical techniques that remain in continuous use in studies aimed at characterization of protein (and, more generally, biopolymer) higher order structure. Unfortunately, these techniques suffer from a very significant limitation that makes the vast majority of them much less attractive compared to HDX MS vis-à-vis characterization of protein therapeutics. Most existing labeling protocols rely on extensive modification of the protein molecules in the sample in order to achieve the yields sufficient for MS characterization, and in most situations this results in placing multiple labels within a single protein molecule. However, even chemical modification of a single site within the protein molecule can result in a significant change of the local conformation as well as the global conformational stability. Therefore, placement of each additional label would not necessarily correlate with the solvent accessibility within the context of the native state, but rather reflect the compromised conformational integrity of the covalently modified protein. In fact, exactly this behavior was seen in the example presented in Figure 4.15, where alkylation of a single residue within interferon β1a had resulted in a dramatic decrease of conformational stability within several (but not all) segments of the protein. These changes were readily detected by HDX MS, and there is no doubt that the local unfolding triggered by a single covalent modification would result in a greater accessibility of the affected segment(s) to other chemical probes as well. This example is by no means unique, which is the reason why the classical chemical labeling techniques have failed so far to achieve any notable level of prominence in the analysis of the higher order structure of biopharmaceutical products (and while the exceptions to that do exist [132], they remain extremely rare). This problem would be further exacerbated should chemical cross-linking be used to probe the higher order structure of biopharmaceutical products, as placing even a single cross-link within a protein or between two proteins necessarily requires two chemical modifications. As a result, the uses of chemical cross-linking in the biopharmaceutical analysis also remain rare, although in some instances this technique had been shown to be useful vis-à-vis complementing information obtained via the use of orthogonal techniques, such as HDX MS [133]. In such situations an argument can

be made that once a specific method is validated using orthogonal (and commonly accepted) means, it can be used routinely even in the quality control settings.

4.4.2 Characterization of protein higher order structure with FPOP

While the classical chemical labeling and cross-linking techniques remain largely on the periphery of the efforts to characterize the higher order structure of protein therapeutics, a new labeling approach introduced by M.L. Gross and D. Hambly [134] gave rise to an experimental tool that now enjoys unprecedented popularity in both industry and academia. This technique (termed fast photo-oxidation of proteins, or FPOP [135]) provides an elegant solution to the problem that makes the use of chemical labeling impractical when applied to characterization of the higher order structure of protein therapeutics, namely distortions of the protein structure caused by multiple chemical modifications (*vide supra*). This problem is avoided in FPOP by making the duration of the chemical modification window narrow on the timescale of the protein conformational transitions. As a result, the higher order structure of a protein can be probed in its native state without being affected by structural perturbations (despite placing multiple labels on a single protein molecule). The labeling is irreversible (e.g., oxidation by $OH^{\cdot}$ radicals), which allows a great deal of flexibility as far as the downstream processing of the labeled proteins. This provides a significant advantage over the HDX MS-based methods of the higher order structure analysis, where lability of amide hydrogen atoms imposes rather unforgiving restrictions on all protein processing/analysis steps from proteolysis to MS/MS. One significant disadvantage of FPOP compared to HDX MS is the significantly lower level of labels/reporters that are typically placed on a single protein molecule (a few labels per protein molecule as opposed to labeling every amide backbone in HDX MS), although the recent efforts to expand the repertoire of FPOP probes are beginning to address this problem [136, 137].

The protein labeling in FPOP is triggered by a very short pulse of an excimer laser (Figure 4.17). The laser wavelength is selected such as to maximize the photolytic production of radicals (e.g., $OH^{\cdot}$ via photolysis of H_2O_2, which is present in the protein solution alongside the radical scavenger), and avoid interactions with all other components of the solution. The resultant radicals react with the protein by inducing oxidation, but the radical scavenger present in solution results in complete elimination of these reactive species almost immediately after the end of the laser pulse. The duration of the laser pulse (typically $<10^{-7}$ s) is short on the timescale of the protein conformational transitions. As a result, multiple oxidation events may occur within a single protein molecule without introducing artifacts (the laser pulse frequency and the protein solution flow rate are selected such as to avoid multiple irradiation of the same region of the flowing protein solution). It must be noted, however, that under some conditions the reactive species can persist in solution over

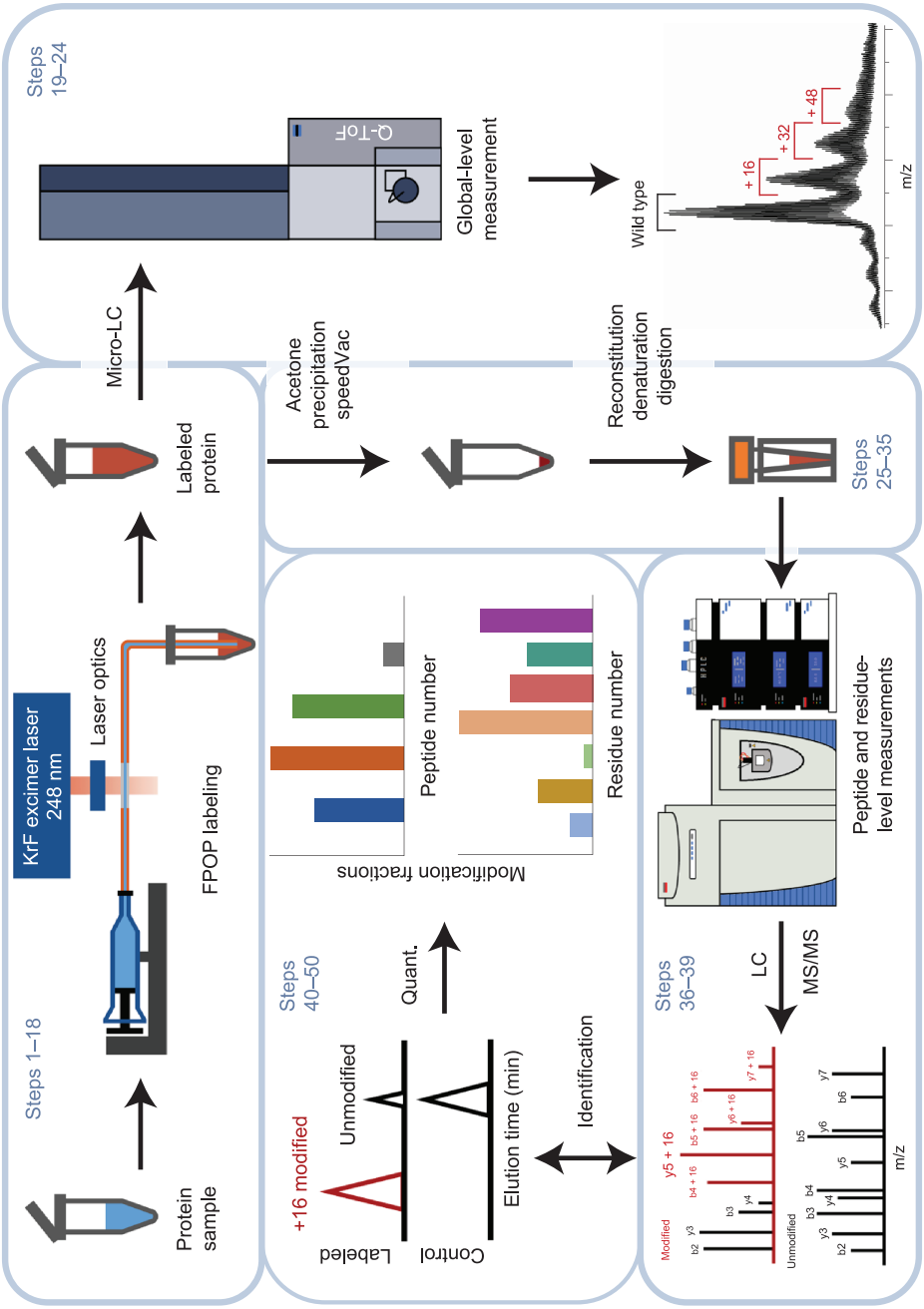


Figure 4.17: A schematic representation of a typical FPOP setup. Adapted by permission from Springer Nature Customer Service Centre GmbH: Liu et al. [147].

much longer time intervals (>10^{-2} s) [138], which would certainly be problematic vis-à-vis the reliability of FPOP measurements.

In recent years FPOP had been applied to solve multiple problems in the field of protein therapeutics ranging from assessing conformational integrity [139, 140] to mapping binding epitopes [141–143]. FPOP has a potential to tackle a range of other relevant problems, including assaying protein aggregation [144] and probing ligand-induced conformational changes [145]; consequently, the range of FPOP applications in the analysis of biopharmaceuticals is likely to continue to expand. Importantly, the information provided by FPOP is complementary to that derived from HDX MS measurements, which frequently prompts the deployment of integrated FPOP/HDX MS strategies [141, 143, 146].

4.5 An outlook for MS-based methods to probe higher order structure of protein therapeutics

The mass spectrometry-based methods discussed in this chapter have already become commonly accepted tools capable of dealing with highly complex systems. They are now routinely used in characterization of systems as large as mAbs, and have become common in regulatory filings [90]. It will be interesting to see if the great success enjoyed by MS as a tool to characterize the higher order structure of protein therapeutics in vitro could be translated to in vivo studies. Indeed, the conformational changes suffered by protein therapeutics post-administration undoubtedly have a significant impact on both efficacy and safety. At the same time, reliable experimental tools capable of detecting and characterizing such events in vivo remain wanting. Should MS prove capable of filling this gap, the impact on the studies of pharmacokinetics and pharmacodynamics of protein therapeutics would be enormous indeed (recent application of the FPOP technology to in vivo studies in human cells [148] look particularly promising in this regard). A related field where mass spectrometry can make significant contributions in the future is the process analytical technology (PAT) [149–151] in the productions of protein therapeutics (to be considered in Chapter 9). This field has been traditionally dominated by less technologically sophisticated, but rugged methods of analysis due to the need to carry out measurements online (and ideally in-line) in real time [152]. However, mass spectrometry is now being actively tested as an alternative, which is capable of providing information on the conformational integrity of protein therapeutics during their production [153]. Above and beyond protein therapeutics, MS-based methods for characterization of higher order structure are likely to accelerate the development of other types of biopharmaceutical products, such as gene therapy products (this will be discussed in Chapter 8).

References

[1] Dobson, C.M. Biophysical techniques in structural biology. Annu. Rev. Biochem. 88, 25–33 (2019).
[2] Poppe, L., Jordan, J.B., Rogers, G. & Schnier, P.D. On the analytical superiority of 1D NMR for fingerprinting the higher order structure of protein therapeutics compared to multidimensional NMR methods. Anal. Chem. 87, 5539–5545 (2015).
[3] Arbogast, L.W., Delaglio, F., Schiel, J.E. & Marino, J.P. Multivariate analysis of two-dimensional 1H, 13C Methyl NMR spectra of monoclonal antibody therapeutics to facilitate assessment of higher order structure. Anal. Chem. 89, 11839–11845 (2017).
[4] Jeffrey, P.D. et al. The X-ray structure of an anti-tumour antibody in complex with antigen. Nat. Struct. Biol. 2, 466–471 (1995).
[5] Maveyraud, L. & Mourey, L. Protein X-ray crystallography and drug discovery. Molecules 25, 1030 (2020).
[6] Brader, M.L., Baker, E.N., Dunn, M.F., Laue, T.M. & Carpenter, J.F. Using X-ray crystallography to simplify and accelerate biologics drug development. J. Pharm. Sci. 106, 477–494 (2017).
[7] Houde, D.J. & Berkowitz, S.A. Biophysical Characterization of Proteins in Developing Biopharmaceuticals. (Elsevier Science Bv, Amsterdam; 2015).
[8] Bobst, C.E. & Kaltashov, I.A. Advanced mass spectrometry-based methods for the analysis of conformational integrity of biopharmaceutical products. Curr. Pharm. Biotechnol. 12, 1517–1529 (2011).
[9] Kaltashov, I.A., Bobst, C.E., Pawlowski, J. & Wang, G. Mass spectrometry-based methods in characterization of the higher order structure of protein therapeutics. J. Pharm. Biomed. Anal. 184, 113169 (2020).
[10] Whitehouse, C.W., Dreyer, R.N., Yamashita, M. & Fenn, J.B. Electrospray interface for liquid chromatographs and mass spectrometers. Anal. Chem. 57, 675–679 (1985).
[11] Alexandrov, M.L. et al. Formation of beams of quasi-molecular ions of peptides from solutions. Bioorg. Khim 11, 700–704 (1985).
[12] Loo, J.A., Loo, R.R., Udseth, H.R., Edmonds, C.G. & Smith, R.D. Solvent-induced conformational changes of polypeptides probed by electrospray-ionization mass spectrometry. Rapid Commun. Mass Spectrom. 5, 101–105 (1991).
[13] Chowdhury, S.K., Katta, V. & Chait, B.T. Probing conformational changes in proteins by mass spectrometry. J. Am. Chem. Soc. 112, 9012–9013 (1990).
[14] Smith, R.D., Lightwahl, K.J., Winger, B.E. & Loo, J.A. Preservation of noncovalent associations in electrospray ionization mass spectrometry – Multiply charged polypeptide and protein dimers. Org. Mass Spectrom. 27, 811–821 (1992).
[15] Katta, V. & Chait, B.T. Observation of the heme globin complex in native myoglobin by electrospray-ionization mass-spectrometry. J. Am. Chem. Soc. 113, 8534–8535 (1991).
[16] Ganem, B., Li, Y.T. & Henion, J.D. Detection of noncovalent receptor-ligand complexes by mass spectrometry. J. Am. Chem. Soc. 113, 6294–6296 (1991).
[17] Katta, V. & Chait, B.T. Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. Rapid Commun. Mass Spectrom. 5, 214–217 (1991).
[18] Thevenon-Emeric, G., Kozlowski, J., Zhang, Z. & Smith, D.L. Determination of amide hydrogen exchange rates in peptides by mass spectrometry. Anal. Chem. 64, 2456–2458 (1992).
[19] Ganem, B., Li, Y.-T. & Henion, J.D. Detection of oligonucleotide duplex forms by ion-spray mass spectrometry. Tetrahedron Lett 34, 1445–1448 (1993).
[20] Nguyen, D.N., Becker, G.W. & Riggin, R.M. Protein mass spectrometry: applications to analytical biotechnology. J. Chromatogr. A 705, 21–45 (1995).

[21] Bobst, C.E. et al. Detection and characterization of altered conformations of protein pharmaceuticals using complementary mass spectrometry-based approaches. Anal. Chem. 80, 7473–7481 (2008).
[22] Houde, D., Arndt, J., Domeier, W., Berkowitz, S. & Engen, J.R. Characterization of IgG1 conformation and conformational dynamics by hydrogen/deuterium exchange mass spectrometry. Anal. Chem. 81, 2644–2651 (2009).
[23] Alexandrov, M.L. et al. Ion extraction from solutions at atmospheric pressure – a method of mass spectrometric analysis of bioorganic substances. Dokl. Acad. Nauk SSSR 277, 379–383 (1984).
[24] Yamashita, M. & Fenn, J.B. Electrospray ion source. Another variation on the free-jet theme. J. Phys. Chem. 88, 4451–4459 (1984).
[25] Guevremont, R., Siu, K.W.M., Le Blanc, J.C.Y. & Berman, S.S. Are the electrospray mass spectra of proteins related to their aqueous solution chemistry? J. Am. Soc. Mass Spectrom. 3, 216–224 (1992).
[26] Le Blanc, J.C.Y., Guevremont, R. & Siu, K.W.M. Electrospray mass spectrometry of some proteins and the aqueous solution acid/base equilibrium model in the negative ion detection mode. Int. J. Mass Spectrom. Ion Processes 125, 145–153 (1993).
[27] Peschke, M., Verkerk, U.H. & Kebarle, P. Prediction of the charge states of folded proteins in electrospray ionization. Eur. J. Mass Spectrom. 10, 993–1002 (2004).
[28] Konermann, L. & Douglas, D.J. Unfolding of proteins monitored by electrospray ionization mass spectrometry: a comparison of positive and negative ion modes. J. Am. Soc. Mass Spectrom. 9, 1248–1254 (1998).
[29] Kaltashov, I.A. & Mohimen, A. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. Anal. Chem. 77, 5370–5379 (2005).
[30] Gumerov, D.R., Dobo, A. & Kaltashov, I.A. Protein-ion charge-state distributions in electrospray ionization mass spectrometry: distinguishing conformational contributions from masking effects. Eur. J. Mass Spectrom 8, 123–129 (2002).
[31] Konermann, L., Ahadi, E., Rodriguez, A.D. & Vahidi, S. Unraveling the mechanism of electrospray ionization. Anal. Chem. 85, 2–9 (2013).
[32] Ogorzalek Loo, R.R., Lakshmanan, R. & Loo, J.A. What protein charging (and Supercharging) reveal about the mechanism of electrospray ionization. J. Am. Soc. Mass Spectrom. 25, 1675–1693 (2014).
[33] Mortensen, D.N. & Williams, E.R. Electrothermal supercharging of proteins in native MS: effects of protein isoelectric point, buffer, and nanoESI-emitter tip size. Analyst 141, 5598–5606 (2016).
[34] Konermann, L., Metwally, H., Duez, Q. & Peters, I. Charging and supercharging of proteins for mass spectrometry: recent insights into the mechanisms of electrospray ionization. Analyst 144, 6157–6171 (2019).
[35] Kaltashov, I.A. & Abzalimov, R.R. Do ionic charges in ESI MS provide useful information on macromolecular structure? J. Am. Soc. Mass Spectrom. 19, 1239–1246 (2008).
[36] Kacher, Y. et al. Acid beta-glucosidase: insights from structural analysis and relevance to Gaucher disease therapy. Biol. Chem. 389, 1361–1369 (2008).
[37] Bobst, C.E., Thomas, J.J., Salinas, P., Savickas, P. & Kaltashov, I.A. Impact of oxidation on protein therapeutics: conformational dynamics of intact and oxidized acid-β-glucocerebrosidase at near-physiological pH. Protein Sci. 19, 2366–2378 (2010).
[38] Zamani, L., Lindholm, J., Ilag, L.L. & Jacobsson, S.P. Discrimination among IgG1-º monoclonal antibodies produced by two cell lines using charge state distributions in nanoESI-TOF mass spectra. J. Am. Soc. Mass. Spectrom. 20, 1030–1036 (2009).

[39] Zhang, H.-M. et al. Structural and functional characterization of a hole–hole homodimer variant in a "Knob-Into-Hole" bispecific antibody. Anal. Chem. 89, 13494–13501 (2017).
[40] Abzalimov, R.R. et al. Studies of pH-dependent self-association of a recombinant form of arylsulfatase A with electrospray ionization mass spectrometry and size-exclusion chromatography. Anal. Chem. 85, 1591–1596 (2013).
[41] Tito, M.A. et al. Probing molecular interactions in intact antibody: antigen complexes, an electrospray time-of-flight mass spectrometry approach. Biophys. J. 81, 3503–3509 (2001).
[42] Atmanene, C. et al. Extending mass spectrometry contribution to therapeutic monoclonal antibody lead optimization: characterization of immune complexes using noncovalent ESI-MS. Anal. Chem. 81, 6364–6373 (2009).
[43] Wang, G. et al. Molecular basis of assembly and activation of complement component C1 in complex with immunoglobulin G1 and antigen. Mol. Cell. 63, 135–145 (2016).
[44] Wang, G., De Jong, R.N., Van Den Bremer, E.T.J., Parren, P.W.H.I. & Heck, A.J.R. Enhancing accuracy in molecular weight determination of highly heterogeneously glycosylated proteins by native tandem mass spectrometry. Anal. Chem. 89, 4793–4797 (2017).
[45] Nguyen, S.N., Bobst, C.E. & Kaltashov, I.A. Mass spectrometry-guided optimization and characterization of a biologically active transferrin-lysozyme model drug conjugate. Mol. Pharm. 10, 1988–2007 (2013).
[46] Haberger, M. et al. Functional assessment of antibody oxidation by native mass spectrometry. MAbs 7, 891–900 (2015).
[47] Strasser, J. et al. Unraveling the Macromolecular Pathways of IgG Oligomerization and Complement Activation on Antigenic Surfaces. Nano Lett. 19, 4787–4796 (2019).
[48] Joubert, M.K., Luo, Q., Nashed-Samuel, Y., Wypych, J. & Narhi, L.O. Classification and characterization of therapeutic antibody aggregates. J. Biol. Chem 286, 25118–25133 (2011).
[49] Wang, W., Nema, S. & Teagarden, D. Protein aggregation – Pathways and influencing factors. Int. J. Pharm 390, 89–99 (2010).
[50] Berkowitz, S.A. Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals. AAPS J. 8, E590–605 (2006).
[51] Sahin, E. & Roberts, C.J. Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanisms. Methods Mol. Biol. 899, 403–423 (2012).
[52] Some, D. & Razinkov, V. High-throughput analytical light scattering for protein quality control and characterization. Methods Mol. Biol. 2025, 335–359 (2019).
[53] Kükrer, B. et al. Mass spectrometric analysis of intact human monoclonal antibody aggregates fractionated by size-exclusion chromatography. Pharm. Res 27, 2197–2204 (2010).
[54] Wang, G., Johnson, A.J. & Kaltashov, I.A. Evaluation of electrospray ionization mass spectrometry as a tool for characterization of small soluble protein aggregates. Anal. Chem. 84, 1718–1724 (2012).
[55] Muneeruddin, K., Thomas, J.J., Salinas, P.A. & Kaltashov, I.A. Characterization of small protein aggregates and oligomers using size exclusion chromatography with on-line detection by native electrospray ionization mass spectrometry. Anal. Chem. 86, 10692–10699 (2014).
[56] Fukuda, J., Iwura, T., Yanagihara, S. & Kano, K. Factors to govern soluble and insoluble aggregate-formation in monoclonal antibodies. Anal. Sci. 31, 1233–1240 (2015).
[57] Mirza, U.A., Cohen, S.L. & Chait, B.T. Heat-induced conformational changes in proteins studied by electrospray ionization mass spectrometry. Anal. Chem. 65, 1–6 (1993).
[58] Wang, G., Abzalimov, R.R. & Kaltashov, I.A. Direct monitoring of heat-stressed biopolymers with temperature-controlled electrospray ionization mass spectrometry. Anal. Chem. 83, 2870–2876 (2011).

[59] Shoemaker, G.K., Soya, N., Palcic, M.M. & Klassen, J.S. Temperature-dependent cooperativity in donor-acceptor substrate binding to the human blood group glycosyltransferases. Glycobiology 18, 587–592 (2008).
[60] Wang, G., Bondarenko, P.V. & Kaltashov, I.A. Multi-step conformational transitions in heat-treated protein therapeutics can be monitored in real time with temperature-controlled electrospray ionization mass spectrometry. Analyst 143, 670–677 (2018).
[61] Brown, C.J., Woodall, D.W., El-Baba, T.J. & Clemmer, D.E. Characterizing thermal transitions of IgG with mass spectrometry. J. Am. Soc. Mass Spectrom. 30, 2438–2445 (2019).
[62] Abzalimov, R.R. & Kaltashov, I.A. Electrospray ionization mass spectrometry of highly heterogeneous protein systems: protein ion charge state assignment via incomplete charge reduction. Anal. Chem. 82, 7523–7526 (2010).
[63] Yang, Y., Du, Y. & Kaltashov, I.A. The utility of native MS for understanding the mechanism of action of repurposed therapeutics in COVID-19: heparin as a disruptor of the SARS-CoV-2 interaction with its host cell receptor. Anal. Chem. 92, 10930–10934 (2020).
[64] Niu, C., Yang, Y., Huynh, A., Nazy, I. & Kaltashov, I.A. Platelet factor 4 interactions with short heparin oligomers: implications for folding and assembly. Biophys. J. 118, 1371–1379 (2020).
[65] Minsky, B.B. et al. Mass spectrometry reveals a multifaceted role of glycosaminoglycan chains in factor Xa inactivation by antithrombin. Biochemistry 57, 4880–4890 (2018).
[66] Kaltashov, I.A. et al. LC/MS at the whole protein level: studies of biomolecular structure and interactionsusing native LC/MS and cross-path reactive chromatography (XP-RC) MS. Methods (San Diego, Calif.) 144, 14–26 (2018).
[67] Jost, J. Partial Differential Equations, Edn. 2nd. (Springer, New York; 2007).
[68] Muneeruddin, K., Nazzaro, M. & Kaltashov, I.A. Characterization of intact protein conjugates and biopharmaceuticals using ion-exchange chromatography with online detection by native electrospray ionization mass spectrometry and top-down tandem mass spectrometry. Anal. Chem. 87, 10138–10145 (2015).
[69] Bohrer, B.C., Mererbloom, S.I., Koeniger, S.L., Hilderbrand, A.E. & Clemmer, D.E. Biomolecule Analysis by ion mobility spectrometry. Ann. Rev. Anal. Chem. 1, 293–327 (2008).
[70] Campuzano, I.D. & Lippens, J.L. Ion mobility in the pharmaceutical industry: an established biophysical technique or still niche? Curr. Opin. Chem. Biol. 42, 147–159 (2018).
[71] D'Atri, V. et al. Adding a new separation dimension to MS and LC-MS: what is the utility of ion mobility spectrometry? J. Sep. Sci. 41, 20–67 (2018).
[72] Gabelica, V., Galic, N., Rosu, F., Houssier, C. & De Pauw, E. Influence of response factors on determining equilibrium association constants of non-covalent complexes by electrospray ionization mass spectrometry. J. Mass Spectrom. 38, 491–501 (2003).
[73] Gabelica, V., Rosu, F. & De Pauw, E. A simple method to determine electrospray response factors of noncovalent complexes. Anal. Chem. 81, 6708–6715 (2009).
[74] Thangaraj, S.K. et al. Quantitation of thyroid hormone binding to anti-thyroxine antibody fab fragment by native mass spectrometry. ACS Omega 4, 18718–18724 (2019). https://pubs.acs.org/doi/10.1021/acsomega.9b02659
[75] Ren, C. et al. Quantitative determination of protein–ligand affinity by size exclusion chromatography directly coupled to high-resolution native mass spectrometry. Anal. Chem. 91, 903–911 (2019).
[76] Gahoual, R. et al. Detailed characterization of monoclonal antibody receptor interaction using affinity liquid chromatography hyphenated to native mass spectrometry. Anal. Chem. 89, 5404–5412 (2017).
[77] Lippold, S. et al. Glycoform-resolved FcRIIIa affinity chromatography-mass spectrometry. MAbs 11, 1191–1196 (2019).

[78] Lippold, S., Nicolardi, S., Wuhrer, M. & Falck, D. Proteoform-resolved fcriiia binding assay for fab glycosylated monoclonal antibodies achieved by affinity chromatography mass spectrometry of fc moieties. Front. Chem. 7, 698 (2019).
[79] Hedges, J.B., Vahidi, S., Yue, X. & Konermann, L. Effects of ammonium bicarbonate on the electrospray mass spectra of proteins: evidence for bubble-induced unfolding. Anal. Chem. 85, 6469–6476 (2013).
[80] Frimpong, A.K., Abzalimov, R.R., Eyles, S.J. & Kaltashov, I.A. Gas-phase interference-free analysis of protein ion charge-state distributions: detection of small-scale conformational transitions accompanying pepsin inactivation. Anal. Chem. 79, 4154–4161 (2007).
[81] Hernandez, H. & Robinson, C.V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. Nat. Protoc. 2, 715–726 (2007).
[82] Saikusa, K., Kato, D., Nagadoi, A., Kurumizaka, H. & Akashi, S. Native mass spectrometry of protein and DNA complexes prepared in nonvolatile buffers. J. Am. Soc. Mass Spectrom. 31, 711–718 (2020).
[83] Cleary, S.P. & Prell, J.S. Liberating native mass spectrometry from dependence on volatile salt buffers by use of Gábor transform. ChemPhysChem 20, 519–523 (2019).
[84] Abzalimov, R.R., Frimpong, A.K. & Kaltashov, I.A. Gas-phase processes and measurements of macromolecular properties in solution: on the possibility of false positive and false negative signals of protein unfolding. Int. J. Mass Spectrom. 253, 207–216 (2006).
[85] Jurchen, J.C. & Williams, E.R. Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homodimers. J. Am. Chem. Soc. 125, 2817–2826 (2003).
[86] Felitsyn, N., Kitova, E.N. & Klassen, J.S. Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior. Anal. Chem. 73, 4647–4661 (2001).
[87] Chernushevich, I.V. & Thomson, B.A. Collisional cooling of large ions in electrospray mass spectrometry. Anal. Chem. 76, 1754–1760 (2004).
[88] Wang, G. et al. Molecular basis of assembly and activation of complement component C1 in complex with immunoglobulin G1 and antigen. Mol. Cell 63, 135–145 (2016).
[89] Wei, H. et al. Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications. Drug Discov. Today 19, 95–102 (2014).
[90] Rogstad, S. et al. A retrospective evaluation of the use of mass spectrometry in FDA biologics license applications. J. Am. Soc. Mass Spectrom. 28, 786–794 (2017).
[91] Englander, S.W. Hydrogen exchange and mass spectrometry: a historical perspective. J. Am. Soc. Mass Spectrom. 17, 1481–1489 (2006).
[92] Konermann, L., Stocks, B.B., Pan, Y. & Tong, X. Mass spectrometry combined with oxidative labeling for exploring protein structure and folding. Mass Spectrom. Rev. 29, 651–667 (2010).
[93] Walters, B.T., Ricciuti, A., Mayne, L. & Englander, S.W. Minimizing back exchange in the hydrogen exchange-mass spectrometry experiment. J. Am. Soc. Mass Spectrom. 23, 2132–2139 (2012).
[94] Han, J.C. & Han, G.Y. A procedure for quantitative determination of Tris(2-carboxyethyl) phosphine, an odorless reducing agent more stable and effective than dithiothreitol. Anal. Biochem. 220, 5–10 (1994).
[95] Giansanti, P., Tsiatsiani, L., Low, T.Y. & Heck, A.J. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. Nat. Protoc. 11, 993–1006 (2016).
[96] Kaltashov, I.A., Bobst, C.E., Abzalimov, R.R., Berkowitz, S.A. & Houde, D. Conformation and dynamics of biopharmaceuticals: transition of mass spectrometry-based tools from academe to industry. J. Am. Soc. Mass Spectrom. 21, 323–337 (2010).

[97] Bonnington, L. et al. Application of hydrogen/deuterium exchange-mass spectrometry to biopharmaceutical development requirements: improved sensitivity to detection of conformational changes. Anal. Chem. 89, 8233–8237 (2017).
[98] Bommana, R., Chai, Q., Schoneich, C., Weiss, W.F.T. & Majumdar, R. Understanding the increased aggregation propensity of a light-exposed IgG1 monoclonal antibody using hydrogen exchange mass spectrometry, biophysical characterization, and structural analysis. J. Pharm. Sci. 107, 1498–1511 (2018).
[99] Houde, D.J., Bou-Assaf, G.M. & Berkowitz, S.A. Deciphering the biophysical effects of oxidizing sulfur-containing amino acids in interferon-beta-1a using MS and HDX-MS. J. Am. Soc. Mass Spectrom. 28, 840–849 (2017).
[100] Gamage, C.L.D., Hageman, T.S. & Weis, D.D. Rapid prediction of deamidation rates of proteins to assess their long-term stability using hydrogen exchange-mass spectrometry. J. Pharm. Sci. 108, 1964–1972 (2019).
[101] Phillips, J.J. et al. Rate of asparagine deamidation in a monoclonal antibody correlating with hydrogen exchange rate at adjacent downstream residues. Anal. Chem. 89, 2361–2368 (2017).
[102] Mo, J. et al. Quantitative analysis of glycation and its impact on antigen binding. MAbs 10, 406–415 (2018).
[103] Arora, J., Joshi, S.B., Middaugh, C.R., Weis, D.D. & Volkin, D.B. Correlating the effects of antimicrobial preservatives on conformational stability, aggregation propensity, and backbone flexibility of an IgG1 mAb. J. Pharm. Sci. 106, 1508–1518 (2017).
[104] More, A.S. et al. Impact of glycosylation on the local backbone flexibility of Well-Defined IgG1-Fc glycoforms using hydrogen exchange-mass spectrometry. J. Pharm. Sci. 107, 2315–2324 (2018).
[105] Kuhne, F. et al. The impact of immunoglobulin G1 Fc sialylation on backbone amide H/D exchange. Antibodies (Basel, Switzerland) 8 (2019).
[106] Wei, H. et al. Using hydrogen/deuterium exchange mass spectrometry to study conformational changes in granulocyte colony stimulating factor upon PEGylation. J. Am. Soc. Mass Spectrom. 23, 498–504 (2012).
[107] Pan, L.Y., Salas-Solano, O. & Valliere-Douglass, J.F. Antibody structural integrity of site-specific antibody-drug conjugates investigated by hydrogen/deuterium exchange mass spectrometry. Anal. Chem. 87, 5669–5676 (2015).
[108] Hageman, T.S. & Weis, D.D. A Structural Variant Approach for Establishing a Detection Limit in Differential Hydrogen Exchange-Mass Spectrometry Measurements. Anal. Chem. 91, 8017–8024 (2019).
[109] Kaltashov, I.A., Bobst, C.E. & Abzalimov, R.R. H/D exchange and mass spectrometry in the studies of protein conformation and dynamics: Is there a need for a top-down approach? Anal. Chem. 81, 7892–7899 (2009).
[110] Deng, B. et al. Suppressing allostery in epitope mapping experiments using millisecond hydrogen / deuterium exchange mass spectrometry. MAbs 9, 1327–1336 (2017).
[111] Huang, R.Y. et al. Hydrogen/deuterium exchange mass spectrometry and computational modeling reveal a discontinuous epitope of an antibody/TL1A Interaction. MAbs 10, 95–103 (2018).
[112] Yan, Y. et al. Mapping the binding interface in a noncovalent size variant of a monoclonal antibody using native mass spectrometry, hydrogen-deuterium exchange mass spectrometry, and computational analysis. J. Pharm. Sci. 106, 3222–3229 (2017).
[113] Noda, M. et al. Identification of IgG1 aggregation initiation region by hydrogen deuterium mass spectrometry. J. Pharm. Sci. 108, 2323–2333 (2019).

[114] Houde, D., Nazari, Z.E., Bou-Assaf, G.M., Weiskopf, A.S. & Rand, K.D. Conformational analysis of proteins in highly concentrated solutions by dialysis-coupled hydrogen/deuterium exchange mass spectrometry. J. Am. Soc. Mass Spectrom. 27, 669–676 (2016).
[115] Kumar, L. et al. Optimizing the formulation and lyophilization process for a fragment antigen binding (Fab) protein using solid-state hydrogen-deuterium exchange mass spectrometry (ssHDX-MS). Mol. Pharm. 16, 4485–4495 (2019).
[116] Masson, G.R. et al. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. Nat. Methods 16, 595–602 (2019).
[117] Espada, A. et al. A decoupled automation platform for hydrogen/deuterium exchange mass spectrometry experiments. J. Am. Soc. Mass Spectrom. 30, 2580–2583 (2019).
[118] Rand, K.D., Zehl, M. & Jorgensen, T.J. Measuring the hydrogen/deuterium exchange of proteins at high spatial resolution by mass spectrometry: overcoming gas-phase hydrogen/deuterium scrambling. Acc. Chem. Res. 47, 3018–3027 (2014).
[119] Abzalimov, R.R., Bobst, C.E. & Kaltashov, I.A. A new approach to measuring protein backbone protection with high spatial resolution using H/D exchange and electron capture dissociation. Anal. Chem. 85, 9173–9180 (2013).
[120] Abzalimov, R.R. & Kaltashov, I.A. Controlling hydrogen scrambling in multiply charged protein ions during collisional activation: implications for top-down hydrogen/deuterium exchange MS utilizing collisional activation in the gas phase. Anal. Chem. 82, 942–950 (2010).
[121] Hoerner, J.K., Xiao, H., Dobo, A. & Kaltashov, I.A. Is there hydrogen scrambling in the gas phase? Energetic and structural determinants of proton mobility within protein ions. J. Am. Chem. Soc. 126, 7709–7717 (2004).
[122] Pan, J., Han, J., Borchers, C.H. & Konermann, L. Hydrogen/deuterium exchange mass spectrometry with top-down electron capture dissociation for characterizing structural transitions of a 17 kDa protein. J. Am. Chem. Soc. 131, 12801–12808 (2009).
[123] Abzalimov, R.R., Kaplan, D.A., Easterling, M.L. & Kaltashov, I.A. Protein conformations can be probed in top-down HDX MS experiments utilizing electron transfer dissociation of protein ions without hydrogen scrambling. J. Am. Soc. Mass Spectrom. 20, 1514–1517 (2009).
[124] Brodie, N.I. et al. Top-down hydrogen-deuterium exchange analysis of protein structures using ultraviolet photodissociation. Anal. Chem. 90, 3079–3082 (2018).
[125] Pan, J., Zhang, S., Parker, C.E. & Borchers, C.H. Subzero temperature chromatography and top-down mass spectrometry for protein higher-order structure characterization: method validation and application to therapeutic antibodies. J. Am. Chem. Soc. 136, 13065–13071 (2014).
[126] Pan, J., Zhang, S. & Borchers, C.H. Comparative higher-order structure analysis of antibody biosimilars using combined bottom-up and top-down hydrogen-deuterium exchange mass spectrometry. Biochim. Biophys. Acta, Proteins Proteomics 1864, 1801–1808 (2016).
[127] Pan, J., Zhang, S., Chou, A. & Borchers, C.H. Higher-order structural interrogation of antibodies using middle-down hydrogen/deuterium exchange mass spectrometry. Chem. Sci. 7, 1480–1486 (2016).
[128] Jensen, P.F. et al. Removal of N-Linked Glycosylations at Acidic pH by PNGase A Facilitates Hydrogen/Deuterium Exchange Mass Spectrometry Analysis of N-Linked Glycoproteins. Anal. Chem. 88, 12479–12488 (2016).
[129] Comamala, G. et al. Deglycosylation by the acidic glycosidase PNGase H(+) enables analysis of N-linked glycoproteins by hydrogen/deuterium exchange mass spectrometry. J. Am. Soc. Mass Spectrom. 31, 2305–2312 (2020).

[130] Limpikirati, P., Liu, T. & Vachet, R.W. Covalent labeling-mass spectrometry with non-specific reagents for studying protein structure and interactions. Methods (San Diego, Calif) 144, 79–93 (2018).
[131] Chu, F., Thornton, D.T. & Nguyen, H.T. Chemical cross-linking in the structural analysis of protein assemblies. Methods (San Diego, Calif) 144, 53–63 (2018).
[132] Madsen, J.A. et al. Covalent labeling denaturation mass spectrometry for sensitive localized higher order structure comparisons. Anal. Chem. 88, 2478–2488 (2016).
[133] Zhang, M.M. et al. An integrated approach for determining a protein-protein binding interface in solution and an evaluation of hydrogen-deuterium exchange kinetics for adjudicating candidate docking models. Anal. Chem. 91, 15709–15717 (2019).
[134] Hambly, D.M. & Gross, M.L. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. J. Am. Soc. Mass Spectrom. 16, 2057–2063 (2005).
[135] Li, K.S., Shi, L. & Gross, M.L. Mass spectrometry-based fast photochemical oxidation of proteins (FPOP) for higher order structure characterization. Acc. Chem. Res. 51, 736–744 (2018).
[136] Cheng, M., Zhang, B., Cui, W. & Gross, M.L. Laser-initiated radical trifluoromethylation of peptides and proteins: application to mass-spectrometry-based protein footprinting. Angew. Chem. Int. Ed. Engl. 56, 14007–14010 (2017).
[137] Zhang, M.M., Rempel, D.L. & Gross, M.L. A Fast Photochemical Oxidation of Proteins (FPOP) platform for free-radical reactions: the carbonate radical anion with peptides and proteins. Free Radic. Biol. Med. 131, 126–132 (2019).
[138] Vahidi, S. & Konermann, L. Probing the time scale of FPOP (Fast Photochemical Oxidation of Proteins): radical reactions extend over tens of milliseconds. J. Am. Soc. Mass Spectrom. 27, 1156–1164 (2016).
[139] Jones, L.M. et al. Complementary MS methods assist conformational characterization of antibodies with altered S-S bonding networks. J. Am. Soc. Mass Spectrom. 24, 835–845 (2013).
[140] Cornwell, O., Bond, N.J., Radford, S.E. & Ashcroft, A.E. Long-range conformational changes in monoclonal antibodies revealed using FPOP-LC-MS/MS. Anal. Chem. 91, 15163–15170 (2019).
[141] Li, J. et al. Mapping the energetic epitope of an antibody/interleukin-23 interaction with hydrogen/deuterium exchange, fast photochemical oxidation of proteins mass spectrometry, and alanine shave mutagenesis. Anal. Chem. 89, 2250–2258 (2017).
[142] Li, K.S. et al. Orthogonal mass spectrometry-based footprinting for epitope mapping and structural characterization: the IL-6 receptor upon binding of protein therapeutics. Anal. Chem. 89, 7742–7749 (2017).
[143] Zhang, Y., Wecksler, A.T., Molina, P., Deperalta, G. & Gross, M.L. Mapping the binding interface of VEGF and a monoclonal antibody Fab-1 fragment with Fast Photochemical Oxidation of Proteins (FPOP) and mass spectrometry. J. Am. Soc. Mass Spectrom. 28, 850–858 (2017).
[144] Li, K.S., Rempel, D.L. & Gross, M.L. Conformational-sensitive fast photochemical oxidation of proteins and mass spectrometry characterize amyloid beta 1–42 aggregation. J. Am. Chem. Soc. 138, 12090–12098 (2016).
[145] Liu, X.R., Rempel, D.L. & Gross, M.L. Composite conformational changes of signaling proteins upon ligand binding revealed by a single approach: calcium-calmodulin study. Anal. Chem. 91, 12560–12567 (2019).
[146] Shi, L., Liu, T., Gross, M.L. & Huang, Y. Recognition of human IgG1 by Fcgamma receptors: structural insights from hydrogen-deuterium exchange and fast photochemical oxidation of proteins coupled with mass spectrometry. Biochemistry 58, 1074–1080 (2019).
[147] Liu, X.R., Rempel, D.L. & Gross, M.L. Protein higher-order-structure determination by fast photochemical oxidation of proteins and mass spectrometry analysis. Nat. Protoc. 15, 3942–3970 (2020).

[148] Johnson, D.T., Punshon-Smith, B., Espino, J.A., Gershenson, A. & Jones, L.M. Implementing In-cell fast photochemical oxidation of proteins in a platform incubator with a movable XY Stage. Anal. Chem. 92, 1691–1696 (2020).
[149] Rathore, A.S., Bhambure, R. & Ghare, V. Process analytical technology (PAT) for biopharmaceutical products. Anal. Bioanal. Chem. 398, 137–154 (2010).
[150] Read, E.K. et al. Process analytical technology (PAT) for biopharmaceutical products: part II. Concepts and applications. Biotechnol. Bioeng. 105, 285–295 (2010).
[151] Read, E.K. et al. Process analytical technology (PAT) for biopharmaceutical products: part I. concepts and applications. Biotechnol. Bioeng. 105, 276–284 (2010).
[152] Guerra, A., Von Stosch, M. & Glassey, J. Toward biotherapeutic product real-time quality monitoring. Crit. Rev. Biotechnol. 39, 289–305 (2019).
[153] Furuki, K., Toyo'oka, T. & Yamaguchi, H. A novel rapid analysis using mass spectrometry to evaluate downstream refolding of recombinant human insulin-like growth factor-1 (mecasermin). Rapid Commun. Mass Spectrom. 31, 1267–1278 (2017).
[154] Tong, W. & Wang, G. How can native mass spectrometry contribute to characterization of biomacromolecular higher-order structure and interactions? Methods (San Diego, Calif.) 144, 3–13 (2018).

Chapter 5
Biosimilars and comparability studies

The term "biosimilars" refers to biopharmaceutical products that are copycats of approved biologics for which the patent protection/exclusivity rights have been already lost or are about to expire. Although in the past they were sometimes referred to as "generics," the latter term only applies to small-molecule medicines that are produced by chemical synthesis, and for which the exact replicas can be made. With very few exceptions, it is impossible to create an exact replica of a protein therapeutic, and the manufacturer of a biosimilar product must prove that its deviations from the characteristics of the originator's product do not affect the product's efficacy and safety. This is addressed by performing similarity studies to demonstrate that either the differences are statistically insignificant (which is relatively rare) or practically insignificant. A similar task arises when changes are introduced into the manufacturing process of biopharmaceutical products both during development and after approval, in which case it must be shown to be "comparable" to the appropriate pre-change products (e.g., reference materials, market batches). Analytical characterization is a critical component of both similarity and comparability studies, and mass spectrometry plays a particularly important role.

5.1 Biogenerics or biosimilars?

The early efforts in the field of biosimilars were motivated by the success of the generic small-molecule drugs whose introduction in the USA was enabled by the 1984 Hatch-Waxman Act [2]. However, adoption of the generic drugs' approval pathway to follow on biologics was impossible due to a significant difference between the generic versions of small-molecule drugs and the follow-on biologic drugs. The former can be created as an exact copy of the originator's product, while the complexity of protein therapeutics makes it almost impossible to generate their exact replicas (hence the gradual disappearance of the early term biogenerics [3, 4] from the common use, which has been replaced by biosimilars). The sensitivity of the modern tools of chemical and biophysical characterization allows even the minute differences between the follow-on biologic and the originator's product to be detected; however, not all such differences affect safety and efficacy of the protein therapeutic. Therefore, the acceptance criteria for follow-on biologics were not defined in the USA until 2010 with the addition of Section 351(k) of the Public Health Service Act upon adoption of the Affordable Care Act [5]. Even with the enactment of Section 351(k), implementation of the abbreviated licensure pathway for biosimilars was not immediate, with first formal biosimilars approvals made only in 2015 (Figure 5.1). By that time a number of patents for the first-generation biopharmaceutical products had already expired, and the follow-on versions already developed, leading their manufacturers to seek alternative routes for approval and in some cases to pursue legal actions to force regulatory agencies to consider their applications [6, 7].

https://doi.org/10.1515/9783110546187-005

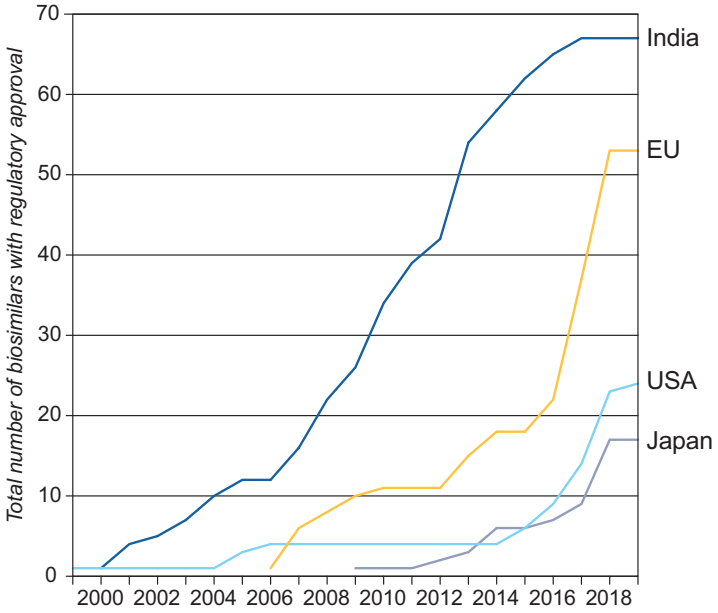


Figure 5.1: Global approvals of biosimilars in the past two decades. Note that the first formal biosimilar approval in the USA was made only in 2015; prior to that alternative approval pathways were used for this class of biopharmaceuticals. Likewise, in India the formal guidelines for biosimilars were not established until 2012, when dozens of follow-on biologics were already on the market. Adapted with permission from Eisenstein and Ashour [1].

Currently, there are several dozen approved biosimilars worldwide, and this number is set to grow as the patents protecting the original biopharmaceutical products continue to expire. The definition of a biosimilar product set forth by the FDA states that it is highly similar to the reference product notwithstanding minor differences in clinically inactive components, and that there are no clinically meaningful safety, purity, and potency differences between the biological product and the reference (originator's) product [8]. Analytical comparison is the cornerstone in establishing the biosimilarity [5, 9, 10]; in addition, both PK/PD and immunogenicity studies are required to demonstrate biosimilarity [11]. Similarity and comparability are two terms that are frequently used interchangeably in literature in connection with the biosimilar products. While comparability is frequently understood as comparing two products, this term is actually used in a narrower sense. It is based on the International Conference on Harmonization (ICH) Comparability Protocol Q5E, Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process that applies to the post-approval changes in the manufacturing process of a licensed product by the same manufacturer [8]. In contrast, similarity is a demonstration of the extent of the sameness of two products (e.g., the originator's products and its follow-on version). The FDA

adheres to this strict use of the vocabulary, with the testing of biosimilar products defined as a similarity exercise [8].

There are four tiers in the biosimilarity assessment: not similar, similar, highly similar, and fingerprint-like similar (with "highly similar" being the minimum qualifying level for a biosimilar product). Although the highly similar level is acceptable, developing the fingerprint-like similarity is likely to reduce the burden of proof on the sponsor to conduct any phase III trials [8]. Establishing the fingerprint-like similarity typically relies on assessment that includes inter alia characterization of the higher order structure that goes beyond standard biophysical techniques (e.g., optical spectroscopy methods and SEC) and uses stress tests to evaluate conformational changes that may occur during the product's life cycle (production/storage/use). Evaluation of the conformational integrity at higher level of spatial resolution compared to what can be achieved with standard biophysical tools (e.g., by using HDX MS discussed in Chapter 4) is another approach that may help achieve the highest tier in the biosimilarity assessment [8]. Importantly, the FDA does not specify explicitly what constitutes a fingerprint-like similarity; instead, the continuously evolving scientific view of this assessment (informed by the totality-of-evidence paradigm) serves as a guide in both design and evaluation of the biosimilarity assessment.

5.2 MS-based characterization of the covalent structure in the biosimilarity assessments

The foundation of the biosimilarity assessment is analytical comparison, and mass spectrometry plays a pivotal role in these studies. Any comparative differences identified (e.g., in the glycosylation profiles) must be considered in terms of their effect on biosimilarity (which frequently requires further investigations, e.g. via biological assays or indeed preclinical or clinical evaluations) [13]. The commonly applied MS-based analytical approaches include (a) determination of intact molecular mass by ESI MS; (b) amino acid sequence analysis using peptide mapping with LC–MS or MS/MS detection; (c) disulfide network assignment using peptide mapping under both reducing and non-reducing conditions; (d) glycosylation analysis; (e) analysis of other PTMs using peptide mapping with LC-MS detection; and (f) profiling the charge heterogeneity (which had been carried out until recently using ion exchange LC or isoelectric focusing, but currently benefits from inclusion of intact mass LC–MS analyses in its arsenal) [13]. Mass spectrometry also plays an important (and rapidly expanding) role in the higher order structural analysis and the purity analysis, but these aspects will be discussed in the subsequent sections.

The MS-based methods used for the assessment of the covalent structure (both the amino acid sequence and the PTMs) of biopharmaceutical products have been considered in great detail in Chapter 3. The majority of these methods are routinely

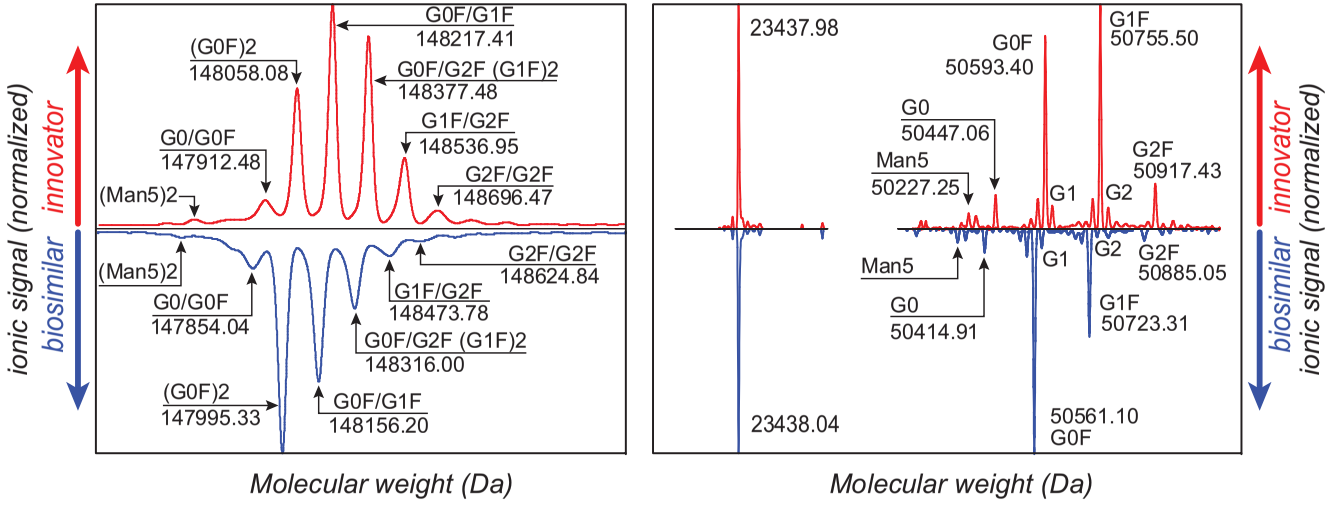


Figure 5.2: Comparative intact mass analysis of disulfide-intact (left) and reduced (right) mAb molecules in the reference sample (the originator's product, red traces) and its follow-on version (blue traces). These butterfly plots represent mass distributions obtained by deconvoluting the raw ESI MS data. The mass shift of the biosimilar drug candidate compared to the reference product is indicative of either unknown PTMs or differences in the amino acid sequences that are confined to the heavy chain of the antibody molecule. Adapted from Xie et al. [12] with permission from Taylor & Francis Ltd (www.tandfonline.com).

used in both biosimilarity and comparability assessments without any modifications. However, a unique aspect of these assessments is the comparative nature of the analyses involved: the attributes of the follow-on product that are identical to/indistinguishable from those of the originator product (both qualitatively and quantitatively) do not require an in-depth analysis. Instead, the biosimilarity assessments focus on the product's characteristics where differences can be detected. The latter are investigated in great detail with the goal of establishing their relevance vis-à-vis characteristics of the product that define/influence its safety and efficacy.

Intact mass measurement is typically the first step in both comparability and similarity studies, since it allows in many cases to determine with certainty if there is a difference in the covalent structure between the reference material and the product being examined. While the intact mass analysis under denaturing conditions provides superior sensitivity, larger and structurally heterogeneous proteins may be analyzed under native conditions to generate lower charge density ionic species and decrease the spectral crowding [14]. In the case of mAbs and other multi-unit protein therapeutics containing external disulfide bonds, intact mass analysis is also carried out following the reduction of disulfide bonds. This allows structural alterations (if any) to be localized to specific subunits of the protein [15, 16]. An example of such analysis is presented in Figure 5.2, where comparison of the mass distributions of an innovator's mAb and a biosimilar product not only reveals a notable difference in the glycan composition (relative abundance of various glycoforms), but also detects the presence of other differences. The latter are manifested by a consistent mass shift of 64 Da for all glycoforms when the originator and the follow-on products are compared. While such a mass shift might be caused by a range of non-enzymatic PTMs, they are expected to generate a distribution of species with varying extent of modifications, giving rise to a set of ionic peaks separated by a specific mass increment. When repeated under reducing conditions, intact mass analysis provides unequivocal evidence that the differences between the originator and the follow-on products occur within the heavy chains of the antibody molecules, while the light chains of the two mAbs are identical to each other (Figure 5.2B). Although this is a trivial conclusion vis-à-vis the differences in the glycan compositions (the light chains of IgG molecules are glycan-free), it allows the light chains to be excluded from consideration when searching for the modification (which results in a 32 Da mass decrease of each heavy chain).

Further localization of the difference(s) between the two species cannot be accomplished by means of intact mass measurements alone, and typically requires a combination of peptide mapping and peptide ion fragmentation in the gas phase. The two tasks can be combined, as most modern mass spectrometers allow LC–MS/MS to be used as a means of both detection and identification of proteolytic peptides. Typically, LC–MS data arrays are acquired for both samples, and alternating MS and MS/MS scans allow both mass and sequence information to be obtained for each proteolytic fragment. An example is shown in Figure 5.3, where

the total ion chromatograms of tryptic digests of the originator and the follow-on mAbs appear to be nearly identical, although a careful analysis allows a small difference to be detected. The only mismatched peptide pair in the digest accounts for the entire mass difference between the two heavy chain types (31.97 Da, see **panel B** in Figure 5.3). The amino acid sequences of both peptides in question are readily determined based on the MS/MS data, revealing the amino acid substitution at two positions (see **panel C** in Figure 5.3).

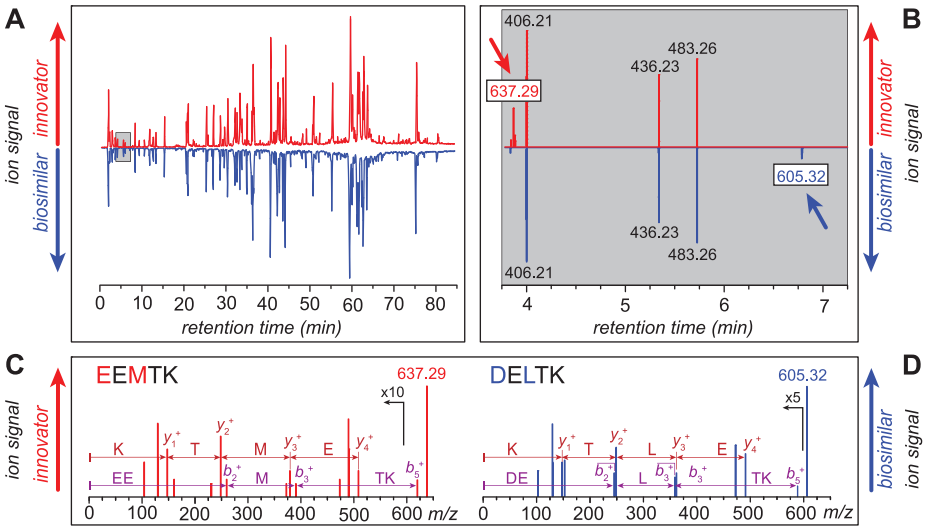


Figure 5.3: LC–MS tryptic peptide maps of innovator and biosimilar mAbs (A) and a zoomed view of the two chromatograms showing monoisotopic masses of tryptic peptide ions within the 3'45"–7'25"' elution window (B). Tandem mass spectra of the two unique peptide ions (m/z 637.29 for the innovator's mAb and m/z 605.32 for the biosimilar product) provide evidence of an amino acid sequence alteration. Adapted from Xie et al. [12] with permission from Taylor & Francis Ltd (www.tandfonline.com).

While intact mass analysis provides a relatively straightforward way to assess glycosylation differences between the reference material and the product being examined for either comparability or similarity, in many cases it does not provide sufficient sensitivity to afford reliable detection and quantitation of minor glycoforms. The latter tasks are typically accomplished by glycan release assays followed by quantitation of various glycans using HILIC chromatography and fluorescence detection (Figure 5.4); since most glycans lack chromophores, they must be labeled with fluorophores to enable their detection. The role of mass spectrometry in such exercises is largely confined to assigning LC peaks by measuring the masses of intact glycans (which can also be done using reference standards), although more involved MS-based methods of glycan analysis, including tandem MS, can also be used in biosimilarity assessments [17].

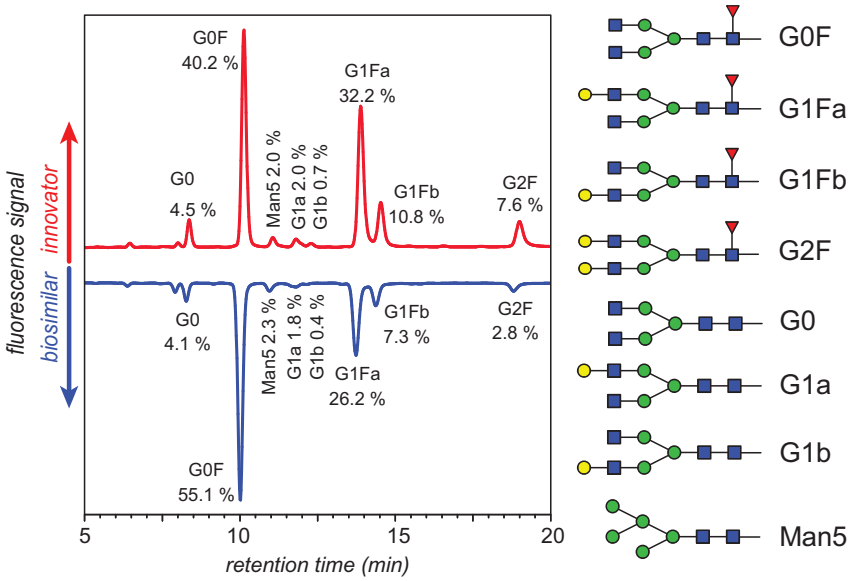


Figure 5.4: LC (HILIC) profiling of glycans released from the innovator and biosimilar mAbs. Detection and quantitation is carried out by fluorescence detection (all glycans are labeled with a fluorophore 2-aminobenzamide), and the LC peak assignments are verified using reference standards and/or MS analysis of LC fractions. Adapted from Xie et al. [12] with permission from Taylor & Francis Ltd (www.tandfonline.com).

Intact mass analysis of disulfide-rich proteins can be used as a means of evaluating the integrity of the disulfide networks if it is supplemented with ion mobility (IM) measurements [14]. Indeed, the presence of mAb isomers having different disulfide connectivity patterns gives rise to multiple gas-phase conformers that in some cases can be detected in IM-MS measurements [18]. An example of using native IM-MS to verify the absence of disulfide isomers within a re-engineered version of a mAb is shown in Figure 5.5, where its ionic drift time distributions are compared to those of the reference product. No evidence of multiple conformers can be seen for any of the charge states/glycoforms present in either sample, consistent with the notion that both products are homogeneous with respect to the disulfide linkage networks. Of course, a more rigorous characterization of both products would be required to establish comparability/similarity of the two products vis-à-vis their disulfide connectivity patterns; peptide mapping under both reducing and non-reducing conditions would likely be required for establishing a fingerprint-tier similarity.

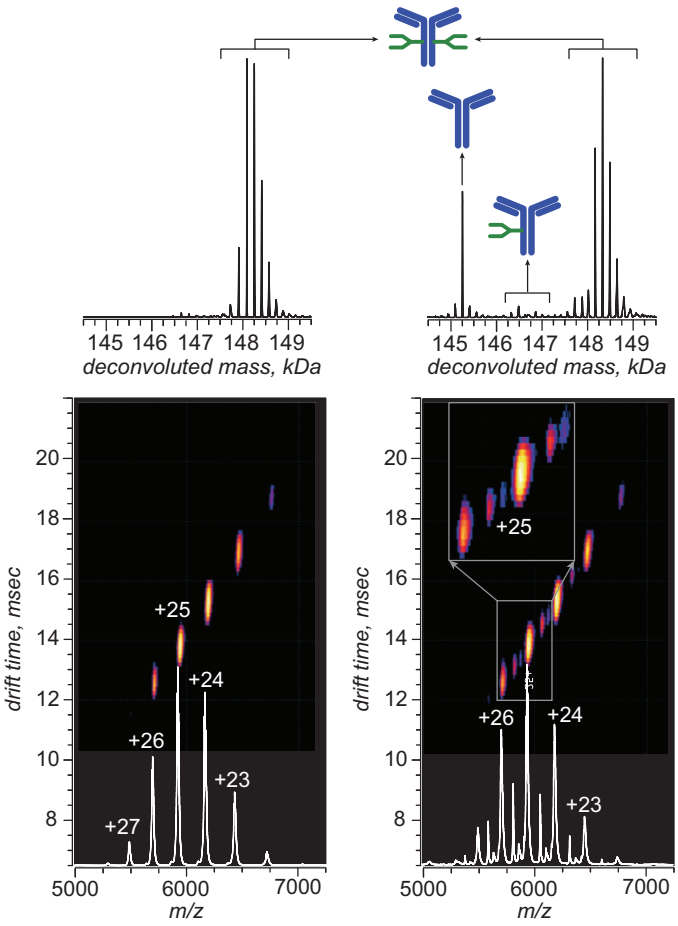


Figure 5.5: Native IM-MS of trastuzumab (left) and trastuzumab-B (right). The top diagrams show deconvoluted mass spectra for the two mAbs, and the bottom diagrams show raw mass spectra and 2-D plots of drift time versus *m*/*z*. Adapted with permission from Beck et al. [14].

5.3 MS-based characterization of the higher order structure in the biosimilarity assessments

Characterization of the higher order structure is an important component of both comparability and similarity assessments. The standard experimental arsenal includes far-UV circular dichroism spectroscopy and/or Fourier transform IR spectroscopy (for secondary structure analysis); and near-UV circular dichroism spectroscopy or intrinsic fluorescence spectroscopy, as well as differential scanning calorimetry (for tertiary structure analysis) [13]. However, reliance on these techniques alone would not be sufficient to establish a fingerprint-tier similarity. The latter requires

more sophisticated biophysical tools, with the MS-based methods of higher order structural analysis playing a particularly prominent role. HDX MS is undoubtedly the most commonly used MS-based tool in such studies [20]. Typically, a classical bottom-up approach is used for protein higher order structure characterization [19, 21–24] (see Chapter 4 for more detail), and an example of using this technique in comparability studies is shown in Figure 5.6.

A progress made in the field of HDX MS in recent years has dramatically expanded the capabilities of this technique [25], including further developments of the top-down HDX MS [26], middle-down HDX MS [27], and time-resolved ESI H/D exchange MS (TRESI-HDX-MS) [28]. One particularly exciting development is the incorporation of sub-zero temperature chromatography in the top-down HDX MS workflow, which allowed the scope of this technique to be expanded dramatically and enabled its use for characterization of the higher order structure of monoclonal antibodies [29, 30]. Although this technique has not been widely adopted in the biopharmaceutical community as yet, its application to the comparative higher order structure analyses of biosimilar mAbs has been demonstrated [31].

IM-MS is another technique that is actively explored by several groups vis-à-vis its utility for the higher order structure characterization in comparability and similarity studies. As already stated in Chapter 4, our position on using IM-MS as a means of structural analysis is that of caution. First, it relies on native MS to generate the proteins ions for subsequent mobility measurements, which requires the use of the ESI-friendly solvents and electrolytes (such as aqueous solutions of ammonium acetate), as opposed to using production- or formulation-specific solvents and buffers. Second, the measurements take place in the solvent-free environment, which is known to distort the native conformations. In fact, some published reports on using IM-MS measurements to support similarity studies appear to be meaningless, for example, protein ion mobility evaluation under denaturing conditions in solution yielding “similar” and “homogeneous” plots with no specific criteria given as to how both similarity and homogeneity should be evaluated [32]. Several other reports present a more nuanced use of IM-MS in biosimilarity and comparability assessments. Even though the “classical” IM-MS data (drift times of unexcited protein ions) in these reports do not provide information beyond what can be extracted from the ionic charge state distributions, application of collisional excitation allows the step-wise loss of structure of protein ions in the gas-phase to be monitored as a function of applied collisional energy [24, 33]. Arguably, the latter could be sensitive to certain structural defects within the protein that would decrease the energy thresholds for the gas-phase structure loss (e.g., the presence of disulfide-scrambled species), although specific examples relevant to comparability and biosimilarity assessments remain wanting.

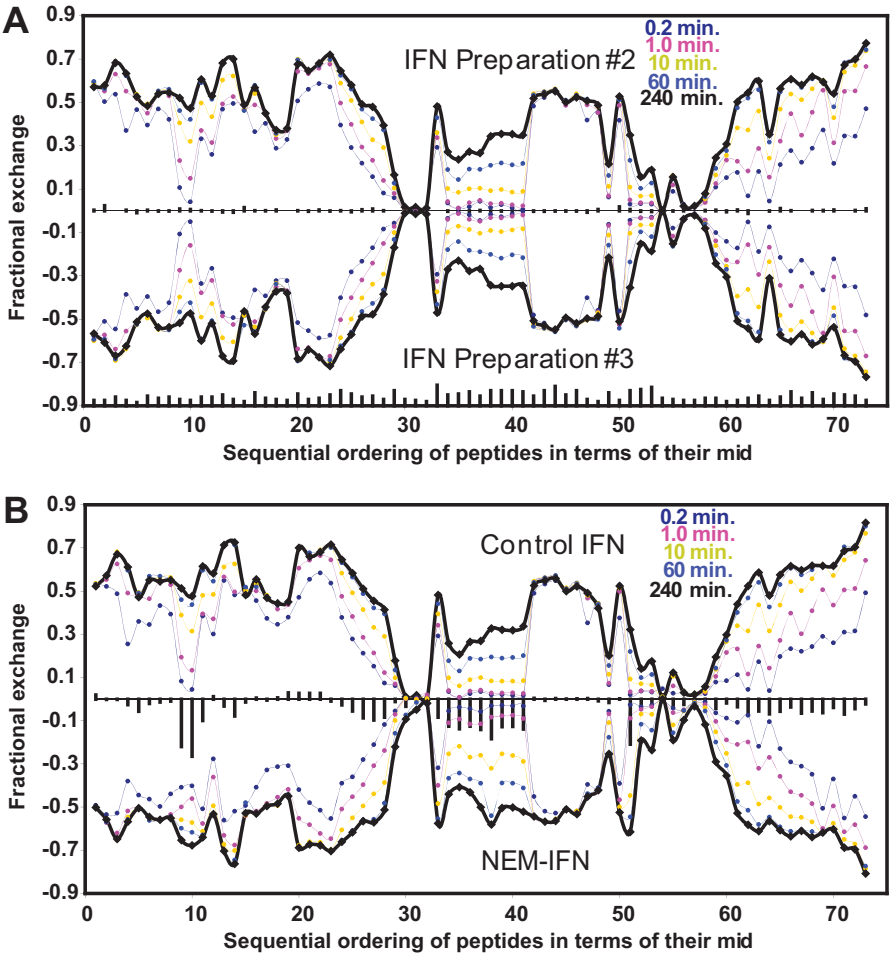


Figure 5.6: HDX MS comparability studies of two different preparations of interferon β1a (IFN) grown in different culture media (A) and chemically modified versus intact IFN (B). The fractional exchange for each peptide at each HDX time point (plotted in different colors) is represented as a single point in a graph where the *x*-axis shows the sequential order of each peptic fragment in the IFN sequence and the *y*-axis represents the fraction of exchanged amides for each peptide. To allow for better visualization, the data for the two samples are plotted in opposite directions (the butterfly plot). Differences in total HDX levels (summations of all time points) for each peptide are plotted as vertical bars across the *x*-axis (the average standard deviation of these values is below 2% of the total exchange level). The vertical lines at the bottom of the graph **A** are scaled to represent the relative size of each peptic fragment. Reproduced with permission from Kaltashov et al. [19].

5.4 MS-based characterization of the purity in biosimilarity assessments

Purity analysis is also a critical component of both comparability and biosimilarity assessments [13], and the role of mass spectrometry in this field continues to expand at an accelerating rate [9, 10]. Host cell proteins analysis is one particular area where MS-based methods are critical to the success of comparability and similarity evaluations [34–36]. These techniques will be considered in detail in Chapter 6 of this book.

References

[1] Eisenstein, M. & Ashour, M. Bring on the biosimilars. Nature 569, S2–S3 (2019).
[2] Mossinghoff, G.J. Overview of the Hatch-Waxman Act and its impact on the drug development process. Food Drug Law J. 54, 187–194 (1999).
[3] Dove, A. Betting on biogenerics. Nat. Biotechnol. 19, 117–120 (2001).
[4] Schellekens, H. & Ryff, J.C. 'Biogenerics': the off-patent biotech products. Trends Pharmacol. Sci. 23, 119–121 (2002).
[5] Nagel, K.M. Introduction to Biologic and Biosimilar Product Development and Analysis. (Springer Science+Business Media, New York, NY; 2018).
[6] Herrera, S. Biogenerics standoff. Nat. Biotechnol. 22,1343–1346 (2004).
[7] Fox, J.L. Sandoz sues FDA over delay in first biogeneric approval. Nat. Biotechnol. 23, 1327–1328 (2005).
[8] Niazi, S. Biosimilarity: The FDA Perspective. (CRC Press, Taylor & Francis Group, Boca Raton; 2019).
[9] Háda, V. et al. Recent advancements, challenges, and practical considerations in the mass spectrometry-based analytics of protein biotherapeutics: a viewpoint from the biosimilar industry. J. Pharm. Biomed. Anal. 161, 214–238 (2018).
[10] Ambrogelly, A. et al. Analytical comparability study of recombinant monoclonal antibody therapeutics. mAbs 10, 513–538 (2018).
[11] García, J.J., Raez, L.E. & Rosas, D. A narrative review of biosimilars: a continued journey from the scientific evidence to practice implementation. Transl. Lung Cancer Res. 9, 2113–2119 (2020).
[12] Xie H. et al. Rapid comparison of a candidate biosimilar to an innovator monoclonal antibody with advanced liquid chromatography and mass spectrometry technologies. mAbs 2, 379–394 (2010).
[13] Walsh, G. Biopharmaceutical benchmarks 2018. Nat. Biotechnol. 36, 1136 (2018).
[14] Beck, A. et al. Cutting-edge mass spectrometry characterization of originator, biosimilar and biobetter antibodies. J. Mass Spectrom. 50, 285–297 (2015).
[15] Xie, H.W. et al. Rapid comparison of a candidate biosimilar to an innovator monoclonal antibody with advanced liquid chromatography and mass spectrometry technologies. mAbs 2, 379–394 (2010).
[16] Li, W. et al. Discovery and characterization of antibody variants using mass spectrometry-based comparative analysis for biosimilar candidates of monoclonal antibody drugs. J. Chromatogr. B 1025, 57–67 (2016).

[17] Duivelshof, B.L. et al. Glycosylation of biosimilars: recent advances in analytical characterization and clinical implications. Anal. Chim. Acta. 1089, 1–18 (2019).
[18] Bagal, D., Valliere-Douglass, J.F., Balland, A. & Schnier, P.D. Resolving disulfide structural isoforms of IgG2 monoclonal antibodies by ion mobility mass spectrometry. Anal. Chem. 82, 6751–6755 (2010).
[19] Kaltashov, I.A., Bobst, C.E., Abzalimov, R.R., Berkowitz, S.A. & Houde, D. Conformation and dynamics of biopharmaceuticals: transition of mass spectrometry-based tools from academe to industry. J. Am. Soc. Mass Spectrom. 21, 323–337 (2010). https://pubs.acs.org/doi/10.1016/j.jasms.2009.10.013.
[20] Auclair, J., Rathore, A.S. & Krull, I.S. Utility of hydrogen-deuterium exchange together with mass spectrometry for determining higher order structures of proteins: comparisons between biosimilar and innovator biopharmaceuticals. LCGC N. Am. 37, 810–815 (2019).
[21] Deng, B., Lento, C. & Wilson, D.J. Hydrogen deuterium exchange mass spectrometry in biopharmaceutical discovery and development – A review. Anal. Chim. Acta. 940, 8–20 (2016).
[22] Upton, R. et al. Hybrid mass spectrometry methods reveal lot-to-lot differences and delineate the effects of glycosylation on the tertiary structure of Herceptin®. Chem. Sci. 10, 2811–2820 (2019).
[23] Fang, J. et al. Advanced assessment of the physicochemical characteristics of Remicade® and Inflectra® by sensitive LC/MS techniques. mAbs 8, 1021–1034 (2016).
[24] Kang, J. et al. Multifaceted assessment of rituximab biosimilarity: the impact of glycan microheterogeneity on Fc function. Eur. J. Pharm. Biopharm. 146, 111–124 (2020).
[25] Oganesyan, I., Lento, C. & Wilson, D.J. Contemporary hydrogen deuterium exchange mass spectrometry. Methods (San Diego, Calif.) 144, 27–42 (2018).
[26] Kaltashov, I.A., Bobst, C.E. & Abzalimov, R.R. H/D exchange and mass spectrometry in the studies of protein conformation and dynamics: is there a need for a top-down approach? Anal. Chem. 81, 7892–7899 (2009).
[27] Pan, J., Zhang, S., Chou, A. & Borchers, C.H. Higher-order structural interrogation of antibodies using middle-down hydrogen/deuterium exchange mass spectrometry. Chem. Sci. 7, 1480–1486 (2016).
[28] Brown, K.A., Rajendran, S., Dowd, J. & Wilson, D.J. Rapid characterization of structural and functional similarity for a candidate bevacizumab (Avastin) biosimilar using a multipronged mass-spectrometry-based approach. Drug Test. Anal. 11, 1207–1217 (2019).
[29] Pan, J., Zhang, S., Parker, C.E. & Borchers, C.H. Subzero temperature chromatography and top-down mass spectrometry for protein higher-order structure characterization: method validation and application to therapeutic antibodies. J. Am. Chem. Soc. 136, 13065–13071 (2014).
[30] Pan, J. & Borchers, C.H. Top-down structural analysis of posttranslationally modified proteins by Fourier transform ion cyclotron resonance-MS with hydrogen/deuterium exchange and electron capture dissociation. Proteomics 13, 974–981 (2013).
[31] Pan, J., Zhang, S. & Borchers, C.H. Comparative higher-order structure analysis of antibody biosimilars using combined bottom-up and top-down hydrogen-deuterium exchange mass spectrometry. Biochim. Biophys. Acta.1864, 1801–1808 (2016).
[32] Montacir, O. et al. Bioengineering of rFVIIa Biopharmaceutical and Structure Characterization for Biosimilarity Assessment. Bioengineering (Basel, Switzerland) 5 (2018).
[33] Pisupati, K. et al. A multidimensional analytical comparison of remicade and the biosimilar remsima. Anal. Chem. 89, 4838–4846 (2017)

[34] Reisinger, V., Toll, H., Mayer, R.E., Visser, J. & Wolschin, F. A mass spectrometry-based approach to host cell protein identification and its application in a comparability exercise. Anal. Biochem. 463, 1–6 (2014).
[35] Mihara, K. et al. Host cell proteins: the hidden side of biosimilarity assessment. J. Pharm. Sci. 104, 3991–3996 (2015).
[36] Liu, J., Eris, T., Li, C., Cao, S. & Kuhns, S. Assessing analytical similarity of proposed amgen biosimilar ABP 501 to adalimumab. BioDrugs 30, 321–338 (2016).

Chapter 6
Characterization of impurities in biopharmaceutical products

Protein biopharmaceuticals are intrinsically heterogeneous due to their complex structures as well as the propensity of acquiring both enzymatic and non-enzymatic modifications during the production and under the storage conditions. In addition, manufacturing of protein biopharmaceuticals is a highly complicated and involved process, which could also be a potential source of introducing undesired impurities ranging from small molecules to whole microorganisms. Therefore, even after a number of purification steps, low levels of impurities are always expected in the final drug substances and products. In general, the impurities from protein biopharmaceuticals can be divided into two categories: product-related impurities and process-related impurities. Those impurities, on a case-by-case basis, could exhibit undesired safety and efficacy profiles, which need to be thoroughly characterized and understood, preferably early in the drug development process. Subsequently, impurities deemed critical for the drug safety and efficacy should be rigorously monitored in routine release testing. In this chapter, instead of reviewing techniques that are routinely applied for the quality control purpose, we will focus on some of the most advanced MS-based technologies to characterize impurities in protein biopharmaceuticals.

6.1 Product-related substances and impurities

Compared to small-molecule drugs, protein biopharmaceuticals exhibit much greater heterogeneity due to their complex structures as well as the occurrence of a wide variety of post-translational modifications. Because of the heterogeneity, the protein biopharmaceutical product is often composed of the desired product and varying levels of product-related variants. The latter can be further divided into product-related substances and product-related impurities, depending on whether the variant species exhibit comparable properties to the desired product with respect to safety and efficacy [3]. In the following section, we will focus on the characterization of product-related variants without further differentiating between product-related substances and impurities, as the latter would require a comprehensive assessment of the potential impact of each variant species on drug safety and efficacy on a case-by-case basis.

6.2 Characterization of size variants of protein biopharmaceuticals

The size variants of protein biopharmaceuticals are normally considered as product-related impurities, as they often exhibit undesired properties with respect to safety and efficacy. Low-molecular-weight (LMW) variants, such as truncated forms

https://doi.org/10.1515/9783110546187-006

of a protein molecule, often exhibit reduced (or loss of) activity or altered pharmacokinetics properties, while high-molecular-weight (HMW) variants, such as protein aggregates, may even induce undesired immunogenicity. As a result, size variants of protein biopharmaceuticals need to be closely monitored and thoroughly characterized during the development stage.

The size variants of protein biopharmaceuticals are commonly monitored by a size-exclusion chromatography (SEC) method and a capillary electrophoresis sodium dodecyl sulfate (CE-SDS) method. It is important to point out that the size variants observed by those two methods are often very different, as the SEC method is normally performed under native conditions, while the CE-SDS method is performed under denaturing conditions. Taking therapeutic mAb molecule as an example, the commonly observed LMW species from CE-SDS analysis might comprise a mAb monomer losing a light chain (H2L), a half mAb (HL), and a free light chain species (L). However, none of those species would be detected by an SEC method, as they still bind to each other, forming an intact mAb molecule, due to the strong non-covalent interactions between the H-chain and L-chain N-terminal regions and between the two H-chain C-terminal regions. In addition, for HMW species (e.g., mAb aggregates), CE-SDS method only reveals the covalent aggregates due to the denaturing nature of the analysis, while SEC method often does not differentiate between covalent and non-covalent aggregates. As a result, when discussing MS-based characterization of size variants in protein biopharmaceuticals, it is important to specify the experimental conditions, as different MS methods often lead to the identification of different size variants.

6.2.1 MS-based characterization of size variants under denaturing conditions

We first focus on MS-based characterization of protein size variants under denaturing conditions. Historically, reversed-phase chromatography (RPLC) was the first separation modality coupled to ESI-MS to characterize the protein biopharmaceuticals at the intact-protein levels (see Chapter 3 for more detail). Although being highly compatible with MS detection, RPLC methods often lack the selectivity and resolution in separating size variants from the main protein product. In particular, the retention times of various size variants on a RP column often correlate poorly with their relative size, making it a suboptimal approach to systematically characterize the complete size heterogeneity. To facilitate a comprehensive size variant characterization, a separation method that enables size-based separation prior to MS detection will be ideal, as the presence or the absence of a particular size variant can be easily assessed based on the elution order. However, under denaturing conditions, there are very few MS compatible chromatographic or electrophoretic methods that have size-based separation capability. Notably, using a RP-like mobile-phase condition (e.g., 30% acetonitrile, 0.1% TFA, and 0.1% FA), Liu et al. have demonstrated that SEC

separation of mAb heavy and light chains could be well achieved under denaturing conditions, which was also compatible with online MS detection [4]. Unfortunately, the SEC resolution obtained under this solvent condition is significantly compromised compared to that from a conventional SEC method under native conditions. Therefore, this method has not been widely adopted to study the size variants of protein biopharmaceuticals. Nevertheless, this method might be of great value to study covalent aggregates that can be readily distinguished from non-covalent aggregates under those denaturing conditions.

More recently, using a hydrophilic interaction chromatography (HILIC) method, Wang et al. have reported the use of a wide-pore amide-bonded column to facilitate the MS-based characterization of the LMW impurities from a therapeutic mAb [5]. Although HILIC method is conventionally applied to separate glycans, glycopeptides, or glycoproteins, based on the hydrophilic interaction between the sugar moiety and the HILIC stationary phase, the authors demonstrated that it could also be used to achieve size-based separation of mAb-related fragments once the N-glycosylation was removed. The resolution achieved by this HILIC method in separating mAb-related size variants was superior to other chromatographic methods under denaturing conditions. As a result, a wide variety of LMW species from a purified mAb drug product, including the free light chain, the Fab fragment, the half antibody (HL), the Fab-clipped monomer, and the H2L species, could all be successfully identified in a single LC-MS analysis (Figure 6.1). In particular, this study provided the first example that H2L species could be chromatographically separated from the main mAb product, and therefore be unambiguously identified based on the measured mass. As the HILIC separation is performed under denaturing conditions (0.1% TFA, acetonitrile and 60–80 °C), the size variants identified by this method should be highly comparable to those from CE-SDS method, which is widely and routinely applied as a purity assay in release testing. However, unlike the CE-SDS method that is unable to provide identification of size variants, the HILIC-MS method offers a valuable approach to elucidate the nature of each LMW impurity in great detail.

6.2.2 MS-based characterization of size variants under non-denaturing conditions

Alternatively, the size variants of protein biopharmaceuticals can also be characterized under native conditions, using SEC-based methods. Thanks to the recent advances in both methodology and instrumentation of native MS techniques, online coupling of SEC separation to native MS analysis has become commonplace. Replacing conventional, involatile salts with ammonium-based volatile salts, native SEC-MS analysis has enabled direct identification of protein size variants in a high throughput manner. In a pioneering study, Muneeruddin et al. have demonstrated successful characterization of small protein aggregates and oligomers by a native

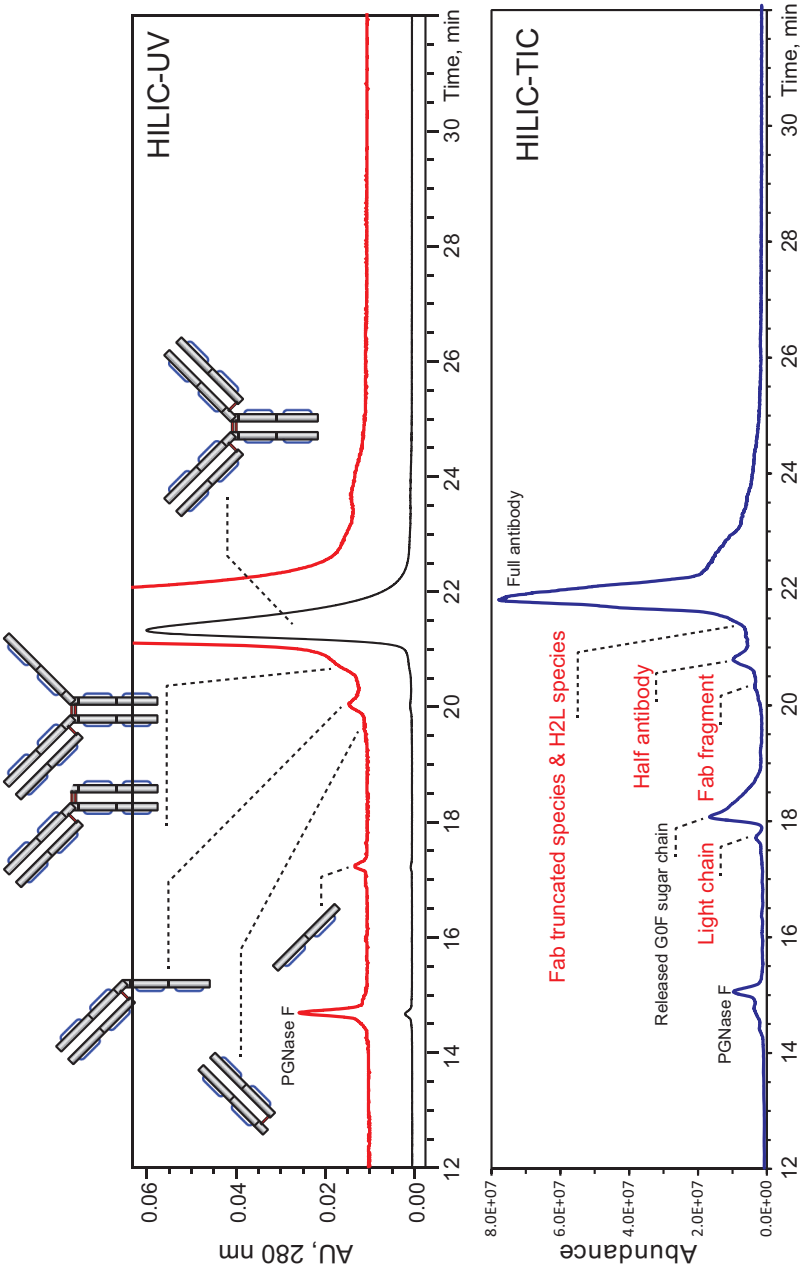


Figure 6.1: The UV (top) and TIC (bottom) profiles from a representative HILIC-MS analysis of mAb-1 drug product sample. The UV signal was amplified 10 times in the red trace to better visualize the LMW impurities. Adapted from Wang et al. [5] with permission from Elsevier.

SEC-MS method [6]. Compared to a conventional workflow requiring off-line fractionation and subsequent MS analysis [7], the online method was not only fast but also eliminated artifact formation during the fractionation processes. More recently, Haberger et al. reported the application of native SEC-MS method to characterize the formation of size variants from a bispecific mAb, under both storage (5 °C) and elevated temperature conditions (40 °C) [8]. Because of the asymmetric feature of a bispecific molecule, a significantly more complicated size variant profile was observed. The combination of SEC separation and the resolving power of native MS analysis was essential for structure elucidation of each size variant species with high confidence. As native SEC-MS methods have become increasingly routine in biopharma laboratories for product characterization, it is worth mentioning two important factors that might affect the method performance. First, although various ammonium-based volatile salts have been used in native SEC-MS applications, a recent study from Ventouri et al. has demonstrated a clear preference of using ammonium acetate over ammonium bicarbonate or ammonium formate, as the latter two are more likely to induce protein structural changes due to their chaotropic nature [9]. Furthermore, as ammonium-based volatile salts often exhibit considerably lower ionic strength than the conventional SEC buffers at equal concentrations, more pronounced secondary interactions between the protein and the column matrix are often observed in native SEC-MS applications. For example, the negatively charged silanol groups from a silica-based SEC column can interact with basic proteins via electrostatic forces. Although those interactions might be reduced by using higher salt concentrations, the sensitivity of subsequent native MS analysis could be compromised due to less efficient ionization/desolvation. In addition, higher salt concentrations might also induce hydrophobic interactions between the protein hydrophobic patches and the chemical groups (e.g., short alkyl chains or linkage of functional groups) from a surface-derivatized SEC column. In a recent report, Yan et al. studied the retention behaviors of different mAb molecules on a SEC column using varying ammonium-based salt concentrations (Figure 6.2) [2]. As demonstrated by the change in retention time of each mAb molecule at different salt concentrations, it is evident that either electrostatic interaction (enhanced at relatively low salt concentrations) or hydrophobic interaction (enhanced at relatively high salt concentrations) between the mAb molecule and the SEC column matrix could be readily induced under these solvent conditions that are commonly used for native SEC-MS applications. In addition, mAb molecules with relatively high basicity or hydrophobicity were expected to exhibit more pronounced secondary interactions with SEC column matrix. As these secondary interactions not only affect the retention behavior but could also lead to compromised recovery of protein analytes, the solvent conditions used in native SEC-MS methods should be carefully evaluated and tailored to specific molecules.

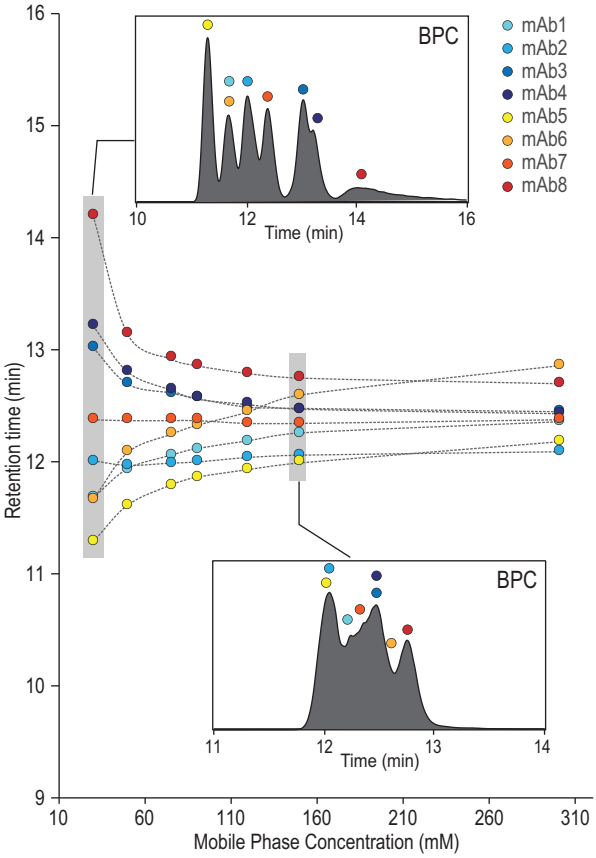


Figure 6.2: Plots of the retention time of eight mAbs on the BEH200 SEC column performed under ammonium salt concentrations ranging from 30 to 300 mM. (insets) Base peak chromatograms (BPCs) from the mmSEC-MS analysis of the eight mAbs at the corresponding concentrations. Adapted with permission from Yan et al. [2]. Copyright 2019 American Chemical Society.

6.2.3 MS-based characterization of protein aggregates

Characterization of HMW variants (e.g., protein aggregates) in protein biopharmaceuticals is of great importance, as these impurities are known to have the potential to trigger unwanted immune responses upon administration [10]. A wide variety of MS-based tools can facilitate the characterization of protein aggregates from different aspects, including investigation of aggregation states, dimerization mechanism and dimerization interfaces, and therefore provide improved understanding of their formation and impact on drug properties. For example, the aggregation states of HMW variants can be readily assessed by native ESI-MS analysis in either an offline or online fashion. In a pioneering study, Wang et al. studied the heat-induced

aggregation processes of human antithrombin III (AT) by direct infusion native MS, where the aggregation states of several oligomeric species could be unambiguously determined [11]. More recently, online native SEC-MS analysis has been more frequently applied to determine the aggregation states of significantly larger protein biopharmaceuticals. Compared to the direct infusion method, this online approach allows the artifacts caused by either gas-phase dissociation of the non-covalent assemblies or non-specific protein association in the ESI interface region to be eliminated, as SEC-based separation provides another degree of confidence to achieve the high-fidelity identification. Moreover, by performing the experiments under either native or denaturing conditions, MS-based tools could also be used to assess the covalent or non-covalent nature of protein aggregates. For example, unlike native MS analysis, both RPLC-MS and denaturing SEC-MS should readily dissociate non-covalent protein aggregates into monomeric species due to the high temperature, low pH, and high-organic solvent conditions applied in these methods. To better understand the covalent protein aggregates, it is also of great interest to elucidate the root cause for covalent cross-linking. Indeed, a wide variety of cross-linking chemistry, including inter-molecular disulfide bond scrambling, dityrosine cross-linking, histidine–histidine cross-linking, and advanced glycation end products (AGEs)-related cross-linking, are all potential causes for the formation of covalent protein aggregates in protein biopharmaceuticals. For example, using a bottom-up based approach, photo-stress related histidine–histidine cross-linking has been recently identified to be responsible for the formation of covalent dimer species of IgG products [12]. This information is obviously valuable for protein drug development, as it provides insights into how to reduce the formation of such covalent aggregates. Finally, for non-covalent aggregates, MS-based tools in protein higher-order structure characterization (see Chapter 4) could also be applied to probe the dimerization interfaces. For example, both HDX MS and hydroxyl radical footprinting (e.g., FPOP) techniques have been used to elucidate the dimerization/aggregation interfaces of therapeutic mAbs [13, 14]. The knowledge on which regions are responsible for dimerization is invaluable, as it not only provides the basis to understand the potential impact of dimerization on drug function, but also offers insights into strategies to reduce the dimerization propensity via protein engineering.

6.3 MS-based characterization of charge variants of protein biopharmaceuticals

Protein biopharmaceuticals can also exhibit a high degree of charge heterogeneity due to their complex structures as well as the propensity to acquire a wide variety of enzymatic or non-enzymatic PTMs. As charge variants are often associated with critical quality attributes of protein biopharmaceuticals, they also need to be thoroughly characterized and rigorously monitored during development and manufacturing.

Compared to many other biophysical methods, MS-based tools are highly effective as a means of elucidation of the underlying biochemical root cause of charge variants formation. Traditionally, MS-based workflow for charge variant characterization consists of off-line fractionation of charge variants via ion exchange chromatography (IEX) or off-gel isoelectric focusing (IEF) and subsequent analysis by different MS-based techniques, such as intact protein mass analysis and peptide mapping analysis. Although this workflow is a highly involved process that is both resource- and time-demanding, it is still a highly effective approach to simplify the charge variant complexity, and therefore facilitate the identification of each corresponding proteoform. Knowledge gained from such studies not only improves the understanding of the product charge heterogeneity, and thus provides a framework for risk assessment, but also guides the process development to better control or even remove undesired charge variants.

More recently, various charge-based separation techniques have been successfully coupled with the online MS detection and have seen increasing applications in the charge heterogeneity characterization of protein biopharmaceuticals. Using ammonium-based volatile salts as mobile phases, IEX with online native MS detection method (IEX-MS) has been pioneered by Muneeruddin et al. and used to study heterogeneous therapeutic proteins and protein conjugates at a variety of levels [15]. By far, the reported IEX-MS methods all sought to use ammonium-based volatile salt to achieve either a salt gradient, pH gradient, or salt-mediated pH gradient for protein elution and separation. For example, the pH-gradient method developed by Füssl et al. has been particularly successful in IEX-MS analysis of mAbs using analytical LC flow rates, as the salt concentrations used in both mobile phase A and B were kept very low (≤25 mM), and thus were highly compatible with direct MS detection [16]. Alternatively, salt-mediated pH-gradient methods have also gained popularity and offer improved separation in some cases due to a different elution mechanism. However, this approach often requires the use of more concentrated mobile phase B for successful elution, which might pose challenges for sensitive MS detection. To address this issue, Yan et al. used a post-column splitting strategy that dramatically reduced the flow rates from analytical IEX separation to be compatible with nanospray ESI [1]. Because of the application of nanospray, this developed method could tolerate relatively high salt concentrations (up to 150 mM) and also proved to be extremely sensitive. In a study, they showcased the successful detection of a previously unreported Fab glycosylation variant in NISTmAb reference standard, which was present at a level less than 0.1%. The authors also demonstrated that, using a generic LC gradient, the developed method could be readily applied to characterize different mAb molecules with a wide range of pI values (6.3–9.2) with both good LC resolution and excellent MS data quality (Figure 6.3).

Besides IEX methods, capillary electrophoresis (CE)-based methods have also been successfully coupled with MS detection to study the charge variants of protein biopharmaceuticals. For example, applying capillary zone electrophoresis (CZE) on a microfluidic platform, the charge heterogeneity of an IgG2 mAb and its drug load

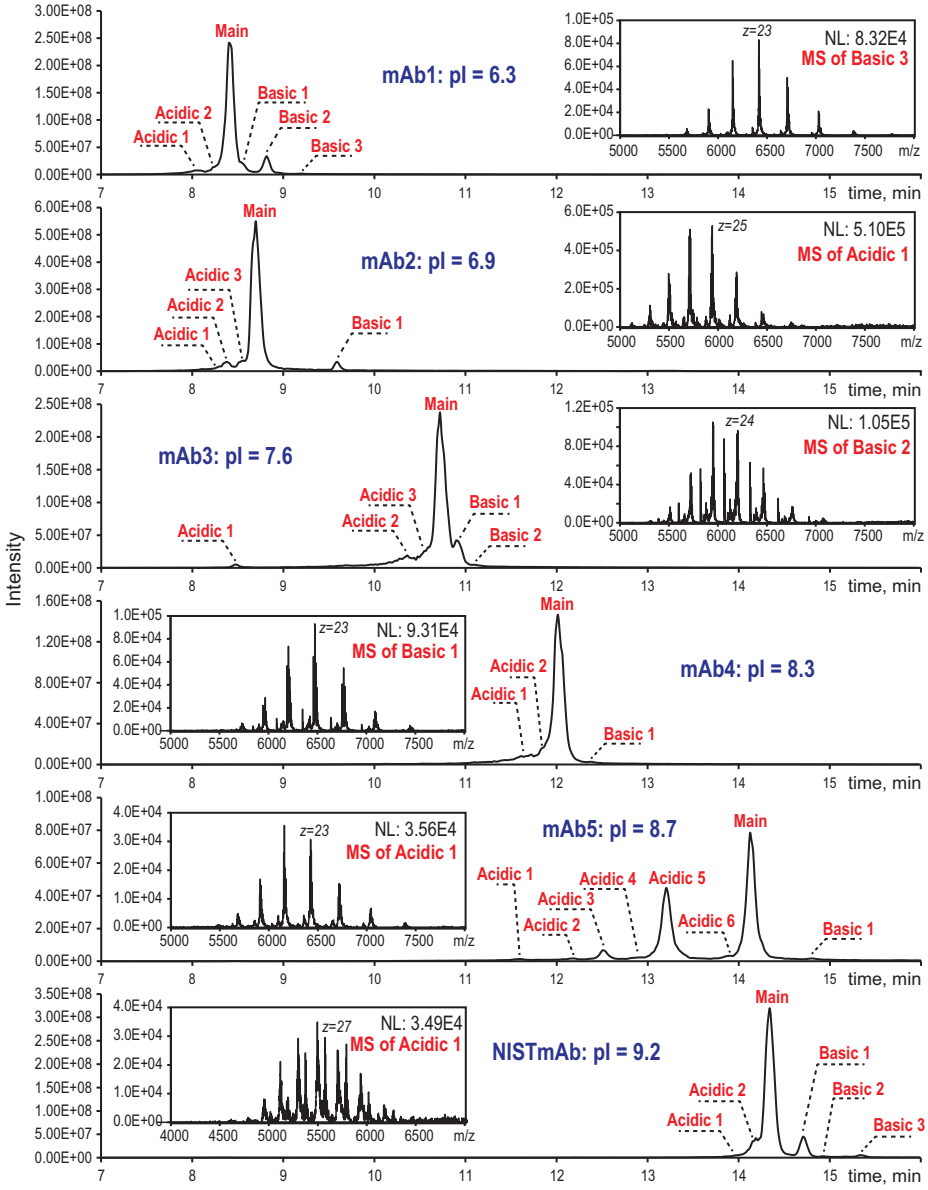


Figure 6.3: Total ion chromatograms (TICs) from the strong cation exchange chromatography – native MS analysis of different mAb samples with pI values ranging from 6.3 to 9.2. The raw mass spectrum of an exemplary minor charge variant peak from each sample is displayed in the corresponding inset. Adapted with permission from Yan et al. [1]. Copyright 2018 American Chemical Society.

variants (e.g., ADC) were extensively characterized using online MS detection [17]. When neutral-pH background electrolytes were applied, CZE could also be performed under near-native conditions and coupled to native MS detection. In a recent study, this "native" CZE-MS method was scrupulously compared with an IEX-MS method to study the charge variants of cetuximab. Both methods were able to achieve high resolution separation of cetuximab charge variants for sensitive MS detection, and provide deep insights into its charge heterogeneity [18]. More recently, thanks to the development of novel interfaces, capillary isoelectric focusing (cIEF) methods have also been successfully coupled to MS detection and used to characterize the charge variants of protein biopharmaceuticals [19, 20]. For example, using a flow-through microvial interface, Wang et al. have shown that cIEF-MS analysis can be readily achieved for peptides, proteins, and mAbs [21]. As cIEF methods have become increasingly popular in quality control (QC) testing for batch release, the ability to obtain cIEF peaks assignment based on online MS measurements is extremely valuable.

It is important to note that, although cIEF, CZE, and IEX methods all separate protein variants based on charge, their separation mechanisms are different. The separation of cIEF is mainly determined by the intrinsic charge (pI), while the separation of CZE and IEX is primarily affected by the surface charge. As a result, certain charge variants identified by one method might not be detected as charge variants in another method. For example, methionine oxidation (+16 Da) is often not considered to contribute to the charge heterogeneity by cIEF method, as it does not change the pI of the protein molecule. However, such oxidized species, on a case-by-case basis, could be identified as either acidic or basic charge variants by IEX method, as methionine oxidation may change protein surface charge in either direction by altering the protein higher order structures [22]. Therefore, application of orthogonal methods will likely lead to a more comprehensive characterization of product charge heterogeneity and is particularly beneficial for supporting late-stage programs, where a detailed understanding of product heterogeneity is often expected by regulatory agencies. Finally, although these charge-based intact protein mass methods (e.g., IEX-MS, CZE-MS, and cIEF-MS) are highly effective and can be applied in a relatively high-throughput manner, they are all somewhat limited in revealing site-specific modifications that are the underlying causes for protein charge variants. Incorporation of top-down techniques, as demonstrated in the previously discussed study [15], might be beneficial for relatively small proteins, but can be challenging for large proteins with abundant disulfide bonds (e.g., mAbs), as the protein ion fragmentation efficiency and sequence coverage can be significantly reduced. Therefore, the traditional workflow, consisting of off-line fractionation and peptide mapping analysis, is still the most effective approach to obtain site-specific information in order to fully elucidate the charge heterogeneity.

6.4 MS-based characterization of homodimer impurities in bispecific mAbs

By engaging two different targets and/or offering novel mechanisms of action, a rapidly growing number of bispecific mAbs (bsAbs) are currently being developed and evaluated in clinical settings to target various human diseases. Containing two different antigen-binding (Fab) arms, this new mAb modality is considerably more complicated than traditional mAbs with respect to both production and characterization. Assembly of the bsAb construct requires both correct pairing of cognate light and heavy chains and heterodimerization of two different half-molecules, a process that can also lead to the formation of homodimer species (i.e., monospecific mAbs). Regarded as product-related impurities, these homodimer species need to be actively monitored, as they are not only unable to achieve the desired therapeutic function, but also can pose significant safety risks (e.g., anti-CD3 antibody). Comparing to size- and charge-variant characterization, where a wide variety of analytical assays are readily available, characterization of homodimer impurities is a relatively new field. In the following paragraph, we will cover some emerging MS-based techniques specifically developed to study the homodimer impurities present in bsAb products.

As homodimer impurities often have different molecular weights compared to the corresponding bsAb, LC-MS based intact mass methods have been the method of choice for their characterization. Early attempts of using RPLC coupled to a high-resolution accurate-mass MS system have been successful and can readily detect homodimer impurities down to 1% [23, 24]. However, further pushing the limit of detection to even lower levels has proved challenging, largely due to the lack of chromatographic separation between the homodimer impurities and the main bsAb species. Alternatively, Phung et al. has shown that native IEX-MS method could also be applied to facilitate the detection of homodimer impurities because of improved chromatographic separation [25]. However, as suggested by Yan et al., to achieve reliable MS-based quantitation of homodimer impurities, an LC method with isocratic elution is highly preferred, as it eliminates the impact of varying solvent conditions on ionization efficiency, a common issue present in gradient-elution methods (e.g., RPLC-MS and IEX-MS) [2]. On this front, the authors developed a mixed-mode size-exclusion chromatography (mmSEC) coupled with native MS method (mmSEC-MS) to achieve highly sensitive detection and quantitation of the homodimer impurities in bsAb samples. The undesired secondary interactions frequently present in a regular SEC method, such as electrostatic interaction and hydrophobic interaction (*vide supra*), were carefully modulated by applying different SEC columns and salt concentrations, leading to improved chromatographic separation between bsAb and homodimer impurities. As shown in Figure 6.4, the achieved chromatographic separation is critical for successful detection of extremely low-abundance variants, as it greatly improves the dynamic range of native MS analysis. This is particularly true

for Orbitrap instruments, which allow the injection time to be effectively utilized to accumulate signals from low-abundance impurities without automatic gain control (AGC) being quickly filled by the main species. As a result, this method was shown to be able to detect homodimer impurities down to 0.1% to 0.01%. Moreover, because of the application of isocratic elution, this method also demonstrated excellent quantitation performance of homodimer impurities down to 0.1% (Figure 6.5).

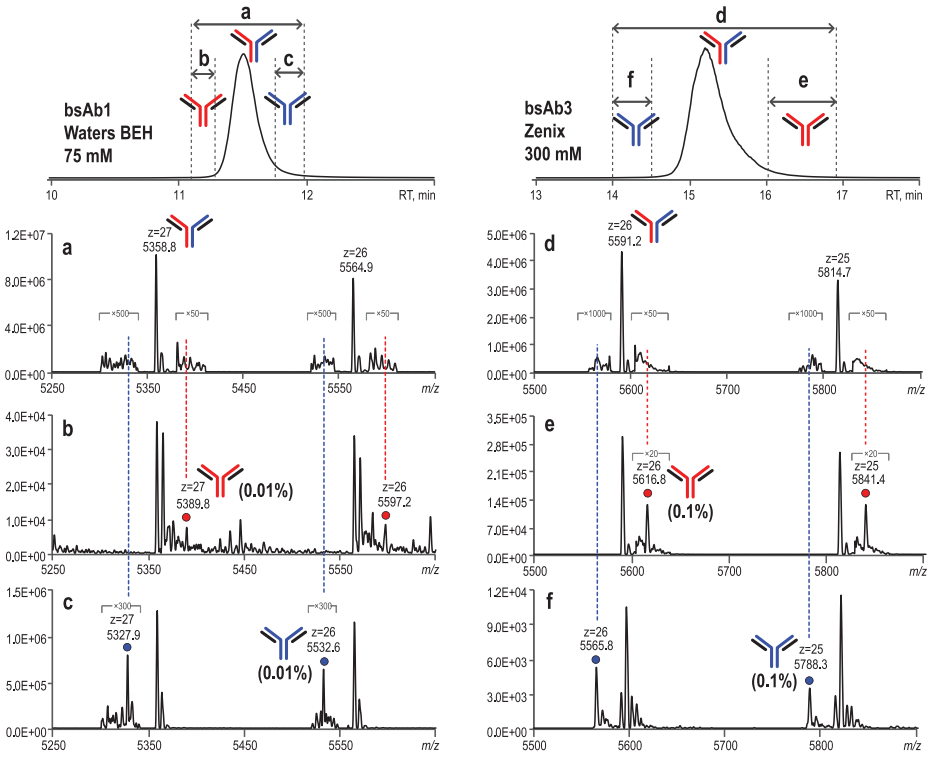


Figure 6.4: Sensitive detection of the homodimer impurities in bsAb1 (LOD: 0.01%, left panel) and bsAb3 (LOD: 0.1%, right panel) by mmSEC-MS. The native MS spectra (a–f) were averaged across the corresponding TIC regions, and the two most abundant charge states are shown. Adapted with permission from Yan et al. [2]. Copyright 2019 American Chemical Society.

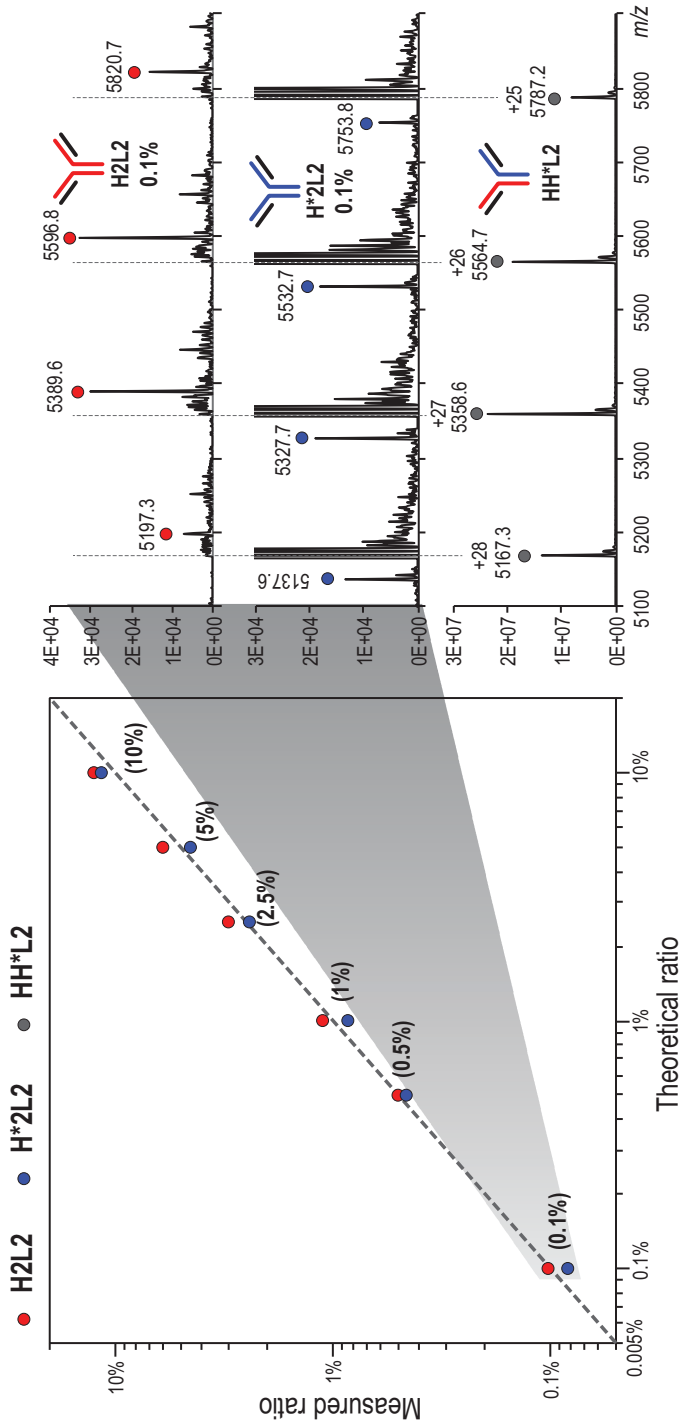


Figure 6.5: Quantitation performance of mmSEC-MS evaluated using the bsAb1 spiked-in standards containing the homodimer impurities ranging from 0.1–10%. The measured ratios of H2L2 homodimer (red) and H*2 L2 homodimer (blue) to bsAb1 are shown in the left panel. The raw mass spectrum of each homodimer at 0.1% relative abundance is shown in the right panel. Adapted with permission from Yan et al. [2]. Copyright 2019 American Chemical Society.

6.5 Process-related impurities: MS-based characterization of host cell proteins

In addition to the product-related impurities, the complex manufacturing workflow for protein biopharmaceuticals can also introduce process-related impurities into the final drug products, such as host cell DNA, host cell proteins (HCPs), as well as microbial and viral contaminants. These impurities are required by the agencies to be tested at product release and should meet pre-defined specifications to be used in clinical studies and for commercialization. In particular, the HCPs, derived from a mammalian expression system (e.g., Chinese hamster ovary (CHO) cells), can significantly impact the product safety, efficacy and stability, and should be rigorously monitored during process development. Although generally present at very low levels (<100 pm), certain HCPs are known to elicit immune responses upon injection or exhibit undesired biological activities. For example, many HCPs can exhibit protease activities that can degrade protein drugs under storage conditions. Moreover, certain HCPs were also found to exhibit lipase activities that can readily break down surfactants (e.g., polysorbate 20) used in the drug formulation, leading to reduced drug stability and shelf life. Although enzyme-linked immunosorbent assays (ELISA) have been routinely performed and generally accepted by regulatory agencies to monitor HCPs, there is growing evidence that ELISA methods cannot provide a complete coverage of HCPs and might leave many critical HCPs undetected. Indeed, using polyclonal antibodies derived against HCPs, ELISA methods may exhibit limited detection capacity toward weakly immunogenic and non-immunogenic HCPs. In addition, ELISA methods often provide quantitation of overall HCP levels, and therefore are unable to provide necessary information to perform risk assessment of individual HCPs.

To address these issues, LC-MS/MS based methods have been increasingly applied to complement the ELISA analysis in order to get a more comprehensive profiling of HCP impurities. In particular, because of the high specificity, LC-MS/MS based methods provide both identification and quantitation of individual HCPs. This information not only allows the risk assessment to be performed on individual HCPs, but also offers important insights on how each HCP gets cleared or even enriched through different process steps, and thus guides the process development. The main challenge of analyzing HCPs by LC-MS/MS in protein drug products is the large dynamic range presented by low levels of HCPs (e.g., low ppm) and the overwhelmingly more abundant protein drugs. To tackle this challenge, one strategy is to apply more efficient LC separation steps to improve the detection of HCP-derived tryptic peptides. For example, both online and off-line 2D LC-MS/MS methods have been reported for successful detection of low abundance HCPs [26, 27]. In particular, high pH RPLC has commonly been applied in the first dimension due to the excellent orthogonality and compatibility with the second dimension low pH RPLC-MS/MS analysis. Alternatively, strategies focused on sample

preparation have also been developed to either enrich HCPs [28] or deplete the protein drugs [29, 30] prior to LC-MS/MS analysis, with a goal to reduce the dynamic range demands. Notably, Huang et al. have recently reported an interesting approach to detect low abundance HCPs in therapeutic mAb products, which employed an unusual tryptic digestion procedure under near-native conditions [31]. Under this so-called native digestion conditions, the mAb molecule, which is structurally compact and rigid, largely remained intact, while the majority of HCPs were effectively digested into tryptic fragments. After a simple heat-induced protein precipitation step, the tryptic digests of HCPs could be readily analyzed by shotgun proteomics approach. This elegant solution effectively overcame the dynamic range issue present in MS-based HCP analysis and has been increasingly adopted in biopharma laboratories.

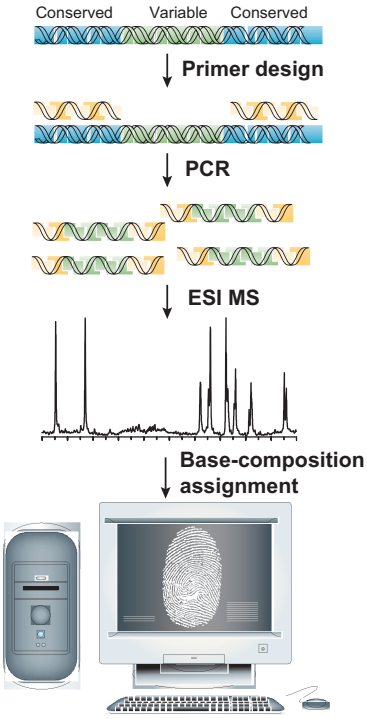


Figure 6.6: A flow scheme for Ibis T5000 analysis. The nucleic acids are purified from the sample, and the target sites on microbial genomes that are common across broad groups of microorganisms and that flank regions of high information content are selected for primer design. Broad-range PCR is conducted and, following an automated desalting step, the masses of the amplified nucleic acids are measured by ESI MS. The mass spectra are processed to identify the masses of the amplicons that are present with sufficient mass accuracy to unambiguously calculate the nucleotide base compositions present in each amplicon. A mathematical process known as triangulation is used to convert the base composition data to a list of the organisms that are present and their relative and absolute quantities. Adapted by permission from Springer Nature Customer Service Centre GmbH: Ecker et al. [32].

It is also important to note that the presence of many residual HCPs in the final products is believed to result from co-purification with protein drugs via a hitchhiking mechanism. Therefore, disrupting such interactions between the protein drugs and HCPs might be critical for a successful enrichment or depletion strategy. For example, Wang et al. recently reported the use of HILIC to isolate and enrich HCPs from protein therapeutics prior to LC-MS analysis [33]. The denaturing HILIC conditions applied in this method facilitated the dissociation of HCPs from protein drugs for subsequent enrichment. Alternatively, Chen et al. reported the use of an anionic detergent to dissociate interactions between mAbs and HCPs followed by a molecular weight cutoff (MWCO) filtration step to enrich HCPs prior to LC-MS analysis [34]. This simple approach was demonstrated to be highly effective in enriching low-abundance HCPs by up to 1,000-fold and showed successful detection of HCPs down to 0.01–8 ppm.

6.6 Viral and microbial contaminants: adventitious agents and bioburden

Ensuring the absence of pathogens in biopharmaceutical products is a very important aspect of the drug safety testing. Viral contamination of the biologic products is a particularly serious concern due to its repeated occurrence in the past. Although in the majority of known cases it was the vaccine products that were unintentionally contaminated with unwanted viruses, protein therapeutics can be affected as well [35]. Perhaps the most dramatic example is the global shortage of recombinant glucocerebrosidase triggered by viral contamination of one of the Genzyme's drug manufacturing facilities and the subsequent halt of production [36]. While pathogen testing typically relies on biological methods of analysis, there are several options that involve the use of MS-based techniques. The suitability of MS as a means of rapid DNA sequencing was realized over two decades ago [37, 38]. Furthermore, the dramatic expansion of the number of completed genomes for a variety of organisms since the turn of the century [39] presented an opportunity to utilize these vast data arrays in searches for genetic markers of pathogens in a variety of complex matrices, including both raw materials for production of protein therapeutics and their finished formulations. In particular, combination of the PCR technology with highly sensitive detection by ESI MS emerged as a powerful platform capable of identifying a wide range of pathogens not only in microbial isolates, but also in clinical samples [40, 41]. Initially developed as a microbial detection/identification tool [32], this approach was later adapted for detection of RNA viruses by including a reverse transcription step in the workflow [42]. This system, introduced in late 2000s as Ibis T5000 (Figure 6.6), went through several stages of upgrades and was commercialized as PLEX-ID, which was subsequently redesigned into a CE-marked IRIDICA system [43]. The latter offered bacterial, fungal, and viral assays, with a typical assay

turnaround time of 8 h (faster than 16S rRNA gene PCR/sequencing), and a maximum throughput of 15 samples per 24 h. Regrettably, production of new IRIDICA instruments and test kits for the existing ones was terminated by the manufacturer in 2017 despite a consensus opinion that the system had clinical utility and no obvious replacement [43]. Although this technology is no longer commercially available, the increased demands placed on microbial and viral detection in a variety of fields, including safety testing of biopharmaceutical products, make it almost certain that a similar state-of-the-art diagnostic technology will emerge in the near future [44]. Another possible approach to detection of microorganisms in a variety of complex matrices is based on application of analytical and bioinformatics tools from the armamentarium of proteomics [45]. Application of these tools to pathogen detection in either raw materials of finished biopharmaceutical products would be conceptually similar to the HCP detection (vide supra).

Finally, as is the case for any pharmaceutical preparations, protein drug formulations must be tested for endotoxins, the components of Gram-negative bacteria (lipopolysaccharides derived from their outer membranes) that are powerful pyrogens and have a potential to precipitate septic shock when present in circulation [47]. It is now recognized that even in the absence of the pyrogenic response, endotoxin can activate the immune system and stimulate the production of anti-protein drug antibodies [48]. MS-based characterization of endotoxins is a mature field [49, 50]; however, until recently MS analysis of endotoxins was mostly limited to purified lipopolysaccharide (LPS) samples, while the ability to detect these molecules in complex matrices remained wanting. Extensive structural heterogeneity exhibited by most bacterial lipopolysaccharides further complicates the use of MS-based tools for the purpose of endotoxin detection. Therefore, current protocols for endotoxin detection continue to rely on traditional tests, such as a rabbit pyrogen test, a limulus amebocyte lysate assay, a recombinant factor C assay, a bovine whole blood assay and a monocyte activation test [51], despite their well-documented shortcomings [52, 53]. Nevertheless, the continuous improvements in the sensitivity of MS measurements and the emergence of novel highly selective LPS extraction/enrichment techniques may lead to incorporation of MS into the endotoxin detection workflows in the future. Specifically, MS-based approaches relying on detection of small-molecule endotoxin markers derived from lipid A (the most conserved part of the LPS molecule that serves as its membrane anchor, see Figure 6.7) were shown to be particularly promising in this regard [46, 54, 55]. The majority of these studies are carried out using LC-MS (with ionization by ESI) to detect the biomarker signal focusing on fatty acids produced by either partial or complete O-deacylation of lipid A; however, sensitive detection of such molecules (following their derivatization) by means of GC-MS (with ionization by electron impact) is also feasible [56]. Lastly, utilization of the so-called ambient ionization techniques, such as direct analysis in real time (DART MS) [57], for endotoxin detection in complex pharmaceutically relevant matrices has also been reported [58].

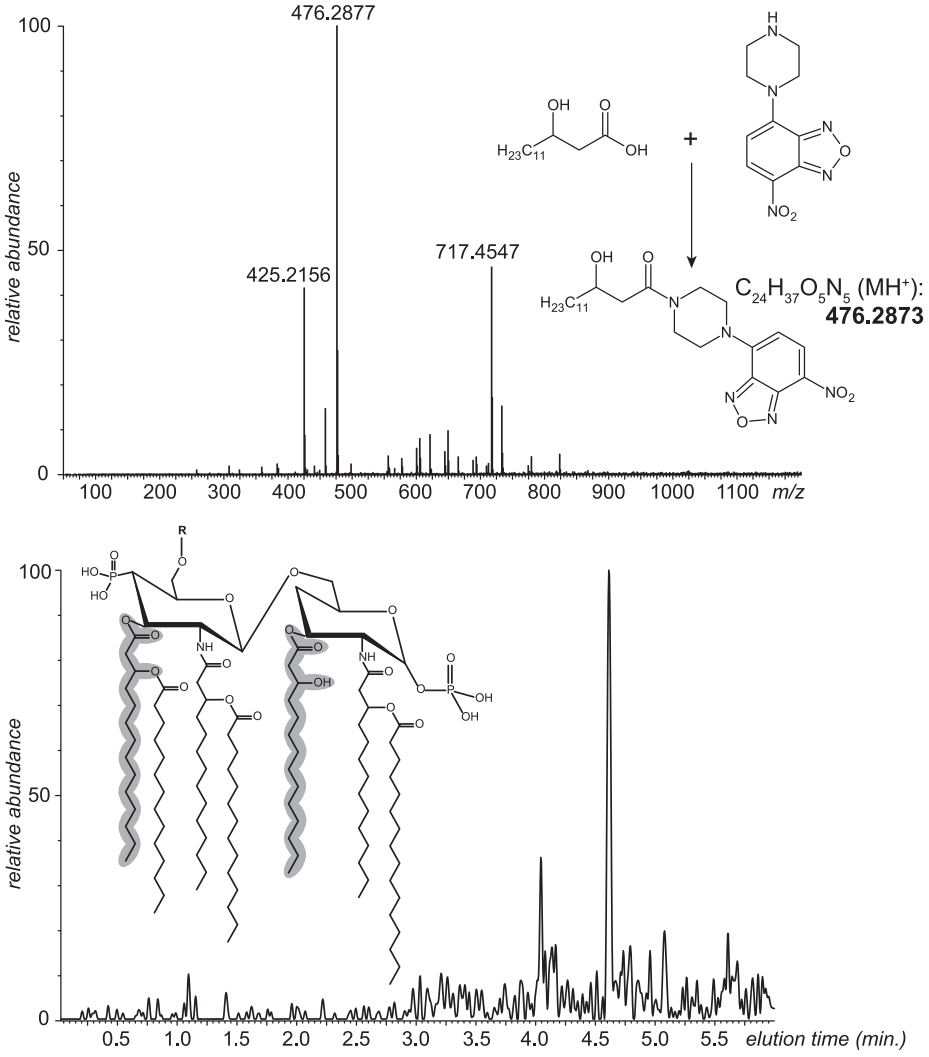


Figure 6.7: Evaluation of chemically modified β-hydroxymyristic acid as an endotoxin marker in pharmaceutically relevant matrices. Top: high-resolution mass spectrum of modified β-hydroxymyristic acid in LC-MS of a reference material. The inset shows the details of derivatization of β-hydroxymyristic acid with a fluorescent label. Bottom: LC-MS chromatogram of β-hydroxymyristic acid released from hydrolyzed LPS of *E. coli*. The inset shows a structure of lipid A (the most conserved part of LPS) from *S. enterica*; β-hydroxymyristoate chains are highlighted in gray. Adapted with permission from Mishra et al. [46] with permission from Elsevier.

6.7 Formulation-related impurities

This chapter would probably be incomplete without mentioning another class of impurities, which are introduced into a protein therapeutic formulation as a result of its interaction with the container, container closure or any other device used for either its storage or administration. Even though the materials used to manufacture the packaging/delivery systems for biopharmaceutical products are selected to be both chemically and physically inert, certain chemical components (collectively termed extractables) can be released from such materials as a result of solvent extraction and/or heat exposure. Although most extractables require fairly aggressive treatment of the packaging/delivery systems materials in order to be released, a subset of them (termed collectively leachables) can leach into a drug product formulation under normal conditions of storage/use. For the majority of pharmaceutical products (not limited to biopharmaceuticals), the interactions that result in leaching are mediated by solvent; in fact, leaching is frequently presumed to affect only liquid formulations. Nevertheless, there is evidence that leaching may also affect lyophilized formulations [59], which are common among the biopharmaceutical products.

Most leachables are either metal ions or small organic molecules, although certain polymers (such as silicone oil or polyethylene glycol) should also be included in this category. While very few leachables are toxic by themselves, many can exert a significant negative influence over safety profiles of biopharmaceutical products. For example, tungsten (the metal used in manufacturing prefilled syringes and frequently found as a contaminant in protein drug formulations) had been shown to trigger formation of the soluble protein aggregates in the pH- and metal concentration-dependent fashion [60]. Other metal leachables frequently encountered in borosilicate vials are aluminum and barium. These two metals form insoluble salts upon interaction with phosphate and sulfate, respectively, and the presence of these anions in formulations containing leached aluminum and barium results in formation of particles that may act as seeds or nucleation sites for protein aggregation [61]. Organic leachables may also have a profound effect on the stability of protein therapeutics. For example, disulfide-containing leachables (such as thiuram disulfides, which are frequently used as accelerators in rubber stoppers) may directly react with proteins containing free thiol groups, causing their degradation [61]. Protein stability can also be affected indirectly, for example, by changing the pH of liquid formulations.

Perhaps the most notorious case of the protein drug safety being compromised by small organic-molecule leachables is the spike in the occurrence of pure red cell aplasia among anemia patients receiving recombinant human erythropoietin treatment, which occurred 10 years after the introduction of this protein therapeutic [63]. The condition (worsening of the anemia, which becomes unresponsive to the increasing doses of the drug and ultimately makes the affected patients transfusion dependent [64]) was determined to be caused by erythropoietin-neutralizing antibodies

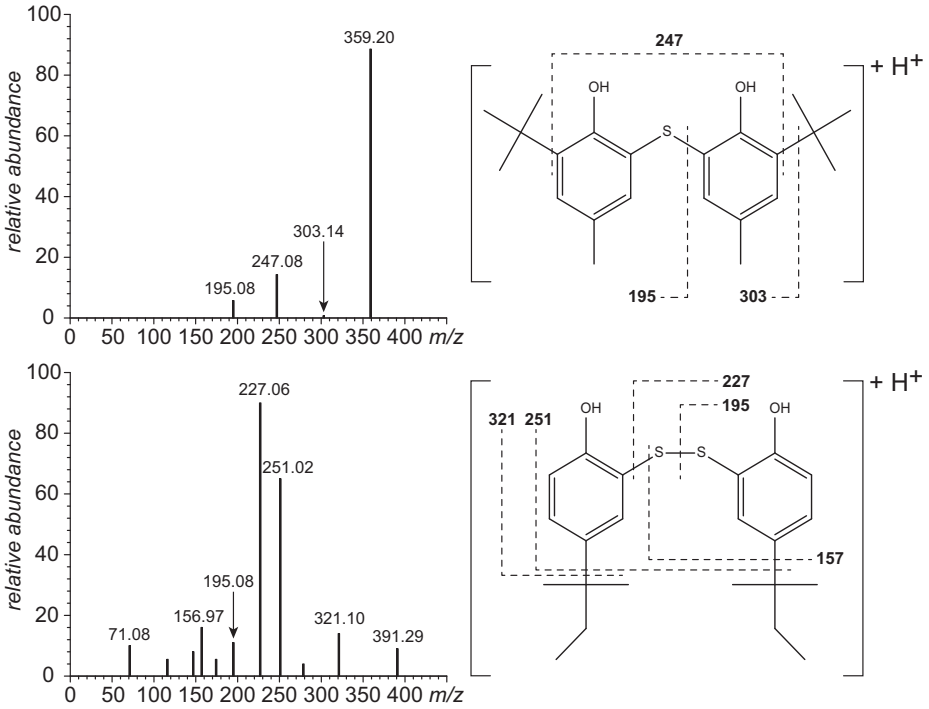


Figure 6.8: Identification of a leachable in EPREX® pre-filled syringes with LC-ESI/MS/MS. Top: a fragment ion mass spectrum of a reference material 2,2′-thiobis[6-(1,2dimethylethyl)-4-methyl] phenol and the proposed fragmentation pathways of the molecular ion at *m*/*z* 359. Bottom: a fragment ion mass spectrum of one of the non-volatile leachables separated by reversed-phase LC and the proposed structure based on the measured masses of the molecular ion (*m*/*z* 391) and the fragment ions as shown on the diagram. Adapted from Pang et al. [62] with permission: PDA, Inc.

and ultimately linked to one particular formulation of the protein therapeutic (Eprex) [63, 64]. Ironically, the new formulation sought to minimize the risk of viral contamination and other impurities by replacing albumin (used as a protein stabilizer) with polysorbate 80 (polyoxyethylene-sorbitan-20 mono-oleate, or Tween 80) and glycine. Although the initial studies of the root cause of the dramatic elevation of immunogenicity pointed at erythropoietin monomers and oligomers trapped in the polysorbate 80 micelles as the culprits [65], further studies failed to confirm this conclusion [66]. Further extensive investigations led to identification of leachables, a set of small-molecule aromatic compounds originating from the uncoated rubber syringe stoppers [62], which were shown to possess adjuvant-like properties in animal models [67]. While a variety of analytical techniques were used to support identification of these leachates, the critical contribution was made by LC-MS/MS (Figure 6.8).

In general, MS-based methods of detection and identification of leachables in biopharmaceutical products are similar to those employed for the analysis of impurities

and degradants present in small-molecule drug formulations. While detailed discussion of these methods goes beyond the scope of this book, an interested reader is referred to a comprehensive collection on this subject edited by Pramanik et al. [68]. Detection and quantitation of metal contaminants in biopharmaceutical products is best accomplished by inductively coupled plasma (ICP) MS, a well-established technique that allows a range of relevant metals to be detected in pharmaceutical preparations with unmatched selectivity and sensitivity (down to sub-ppb levels). An interested reader is referred to several excellent reviews on this subject [69–72].

References

[1] Yan, Y., Liu, A.P., Wang, S., Daly, T.J. & Li, N. Ultrasensitive characterization of charge heterogeneity of therapeutic monoclonal antibodies using strong cation exchange chromatography coupled to native mass spectrometry. Anal. Chem. 90, 13013–13020 (2018). https://pubs.acs.org/doi/10.1021/acs.analchem.8b03773.

[2] Yan, Y., Xing, T., Wang, S., Daly, T.J. & Li, N. Coupling mixed-mode size exclusion chromatography with native mass spectrometry for sensitive detection and quantitation of homodimer impurities in bispecific IgG. Anal. Chem. 91, 11417–11424 (2019). https://pubs.acs.org/doi/full/10.1021/acs.analchem.9b02793.

[3] Q6B Specifications: test Procedures and Acceptance Criteria for Biotechnological/Biological Products. Center for Drug Evaluation and Research; Center for Biologics Evaluation and Research; Food and Drug Administration, Rockville, MD, USA. (1999).

[4] Liu, H., Gaza-Bulseco, G. & Chumsae, C. Analysis of reduced monoclonal antibodies using size exclusion chromatography coupled with mass spectrometry. J. Am. Soc. Mass Spectrom. 20, 2258–2264 (2009).

[5] Wang, S., Liu, A.P., Yan, Y., Daly, T.J. & Li, N. Characterization of product-related low molecular weight impurities in therapeutic monoclonal antibodies using hydrophilic interaction chromatography coupled with mass spectrometry. J. Pharm. Biomed. Anal. 154, 468–475 (2018).

[6] Muneeruddin, K., Thomas, J.J., Salinas, P.A. & Kaltashov, I.A. Characterization of small protein aggregates and oligomers using size exclusion chromatography with online detection by native electrospray ionization mass spectrometry. Anal. Chem. 86, 10692–10699 (2014).

[7] Kükrer, B. et al. Mass spectrometric analysis of intact human monoclonal antibody aggregates fractionated by size-exclusion chromatography. Pharm. Res. 27, 2197–2204 (2010).

[8] Haberger, M. et al. Rapid characterization of biotherapeutic proteins by size-exclusion chromatography coupled to native mass spectrometry. MAbs 8, 331–339 (2016).

[9] Ventouri, I.K. et al. Probing protein denaturation during size-exclusion chromatography using native mass spectrometry. Anal. Chem. 92, 4292–4300 (2020).

[10] Moussa, E.M. et al. Immunogenicity of Therapeutic Protein Aggregates. J. Pharm. Sci. 105, 417–430 (2016).

[11] Wang, G., Johnson, A.J. & Kaltashov, I.A. Evaluation of electrospray ionization mass spectrometry as a tool for characterization of small soluble protein aggregates. Anal. Chem. 84, 1718–1724 (2012).

[12] Liu, M., Zhang, Z., Cheetham, J., Ren, D. & Zhou, Z.S. Discovery and characterization of a photo-oxidative histidine-histidine cross-link in IgG1 antibody utilizing (1)(8)O-labeling and mass spectrometry. Anal. Chem. 86, 4940–4948 (2014).

[13] Deperalta, G. et al. Structural analysis of a therapeutic monoclonal antibody dimer by hydroxyl radical footprinting. mAbs 5, 86–101 (2013).
[14] Iacob, R.E. et al. Investigating monoclonal antibody aggregation using a combination of H/DX-MS and other biophysical measurements. J. Pharm. Sci. 102, 4315–4329 (2013).
[15] Muneeruddin, K., Nazzaro, M. & Kaltashov, I.A. Characterization of intact protein conjugates and biopharmaceuticals using ion-exchange chromatography with online detection by native electrospray ionization mass spectrometry and top-down tandem mass spectrometry. Anal. Chem. 87, 10138–10145 (2015).
[16] Fussl, F. et al. Charge variant analysis of monoclonal antibodies using direct coupled pH gradient cation exchange chromatography to high-resolution native mass spectrometry. Anal. Chem. 90, 4669–4676 (2018).
[17] Redman, E.A., Mellors, J.S., Starkey, J.A. & Ramsey, J.M. Characterization of intact antibody drug conjugate variants using microfluidic capillary electrophoresis-mass spectrometry. Anal. Chem. 88, 2220–2226 (2016).
[18] Fussl, F., Trappe, A., Carillo, S., Jakes, C. & Bones, J. Comparative elucidation of cetuximab heterogeneity on the intact protein level by cation exchange chromatography and capillary electrophoresis coupled to mass spectrometry. Anal. Chem. 92, 5431–5438 (2020).
[19] Dai, J., Lamp, J., Xia, Q. & Zhang, Y. Capillary Isoelectric focusing-mass spectrometry method for the separation and online characterization of intact monoclonal antibody charge variants. Anal. Chem. 90, 2246–2254 (2018).
[20] Dai, J. & Zhang, Y. A middle-up approach with online capillary isoelectric focusing/mass spectrometry for in-depth characterization of cetuximab charge heterogeneity. Anal. Chem. 90, 14527–14534 (2018).
[21] Wang, L. et al. High resolution capillary isoelectric focusing mass spectrometry analysis of peptides, proteins, and monoclonal antibodies with a flow-through microvial interface. Anal. Chem. 90, 9495–9503 (2018).
[22] Shi, R.L., Xiao, G., Dillon, T.M., Ricci, M.S. & Bondarenko, P.V. Characterization of therapeutic proteins by cation exchange chromatography-mass spectrometry and top-down analysis. mAbs 12, 1739825 (2020).
[23] Schachner, L. et al. Characterization of chain pairing variants of bispecific IgG expressed in a single host cell by high-resolution native and denaturing mass spectrometry. Anal. Chem. 88, 12122–12127 (2016).
[24] Yin, Y. et al. Precise quantification of mixtures of bispecific IgG produced in single host cells by liquid chromatography-Orbitrap high-resolution mass spectrometry. mAbs 8, 1467–1476 (2016).
[25] Phung, W. et al. Data on charge separation of bispecific and mispaired IgGs using native charge-variant mass spectrometry. Data Brief 30, 105435 (2020).
[26] Doneanu, C.E. et al. Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. mAbs 4, 24–44 (2012).
[27] Yang, F. et al. A 2D LC-MS/MS Strategy for Reliable Detection of 10-ppm Level Residual Host Cell Proteins in Therapeutic Antibodies. Anal. Chem. 90, 13365–13372 (2018).
[28] Aboulaich, N. et al. A novel approach to monitor clearance of host cell proteins associated with monoclonal antibodies. Biotechnol. Prog. 30, 1114–1124 (2014).
[29] Madsen, J.A. et al. Toward the complete characterization of host cell proteins in biotherapeutics via affinity depletions, LC-MS/MS, and multivariate analysis. mAbs 7, 1128–1137 (2015).
[30] Johnson, R.O., Greer, T., Cejkov, M., Zheng, X. & Li, N. Combination of FAIMS, protein a depletion, and native digest conditions enables deep proteomic profiling of host cell proteins in monoclonal antibodies. Anal. Chem. (2020).

[31] Huang, L. et al. A novel sample preparation for shotgun proteomics characterization of HCPs in antibodies. Anal. Chem. 89, 5436–5444 (2017).
[32] Ecker, D.J. et al. Ibis T5000: a universal biosensor approach for microbiology. Nat. Rev. Microbiol. 6, 553–558 (2008).
[33] Wang, Q. et al. Enhancing host-cell protein detection in protein therapeutics using HILIC enrichment and proteomic analysis. Anal. Chem. (2020).
[34] Chen, I.H., Xiao, H., Daly, T. & Li, N. Improved host cell protein analysis in monoclonal antibody products through molecular weight cutoff enrichment. Anal. Chem. 92, 3751–3757 (2020).
[35] Barone, P.W. et al. Viral contamination in biologic manufacture and implications for emerging therapies. Nat. Biotechnol. 38, 563–572 (2020).
[36] Hollak, C.E. et al. Force majeure: therapeutic measures in response to restricted supply of imiglucerase (Cerezyme) for patients with Gaucher disease. Blood Cells Mol. Dis. 44, 41–47 (2010).
[37] Köster, H. et al. A strategy for rapid and efficient DNA sequencing by mass spectrometry. Nat. Biotechnol. 14, 1123–1128 (1996).
[38] Null, A.P. & Muddiman, D.C. Perspectives on the use of electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for short tandem repeat genotyping in the post-genome era. J. Mass Spectrom. 36, 589–606 (2001).
[39] Koonin, E.V., Makarova, K.S. & Wolf, Y.I. Evolution of microbial genomics: conceptual shifts over a quarter century. Trends Microbiol. (2021).
[40] Wolk, D.M., Kaleta, E.J. & Wysocki, V.H. PCR-electrospray ionization mass spectrometry: the potential to change infectious disease diagnostics in clinical and public health laboratories. J. Mol. Diagn. 14, 295–304 (2012).
[41] Sampath, R. et al. Rapid identification of emerging infectious agents using PCR and electrospray ionization mass spectrometry. Ann. N. Y. Acad. Sci. 1102, 109–120 (2007).
[42] Deyde, V.M., Sampath, R. & Gubareva, L.V. RT-PCR/electrospray ionization mass spectrometry approach in detection and characterization of influenza viruses. Expert. Rev. Mol. Diagn. 11, 41–52 (2011).
[43] Özenci, V., Patel, R., Ullberg, M. & Strålin, K. Demise of polymerase chain reaction/electrospray ionization-mass spectrometry as an infectious diseases diagnostic tool. Clin. Infect. Dis. 66, 452–455 (2018).
[44] Özenci, V. & Strålin, K. Clinical implementation of molecular methods in detection of microorganisms from blood with a special focus on PCR electrospray ionization mass spectrometry. Expert Rev. Mol. Diagn. 19, 389–395 (2019).
[45] Demirev, P.A. & Fenselau, C. Mass spectrometry for rapid characterization of microorganisms. Annu. Rev. Anal. Chem. 1, 71–93 (2008).
[46] Mishra, R.K., Robert-Peillard, F., Ravier, S., Coulomb, B. & Boudenne, J.L. Beta-hydroxymyristic acid as a chemical marker to detect endotoxins in dialysis water. Anal. Biochem. 470, 71–77 (2015).
[47] Gutsmann, T., Schromm, A.B. & Brandenburg, K. The physicochemistry of endotoxins in relation to bioactivity. Int. J. Med. Microbiol. 297, 341–352 (2007).
[48] Williams, K. The emerging view of endotoxin as an IIRMI. Biopharm. Int. 29, 24–31 (2016).
[49] Kilár, A., Dörnyei, Á. & Kocsis, B. Structural characterization of bacterial lipopolysaccharides with mass spectrometry and on- and off-line separation techniques. Mass Spectrom. Rev. 32, 90–117 (2013).
[50] Jackie, J., Lau, W.K., Feng, H.T. & Li, S.F.Y. Detection of endotoxins: from inferring the responses of biological hosts to the direct chemical analysis of lipopolysaccharides. Crit. Rev. Anal. Chem. 49, 126–137 (2019).
[51] Schneier, M., Razdan, S., Miller, A.M., Briceno, M.E. & Barua, S. Current technologies to endotoxin detection and removal for biopharmaceutical purification. Biotechnol. Bioeng. 117, 2588–2609 (2020).

[52] Schwarz, H. et al. Biological activity of masked endotoxin. Sci. Rep. 7, 44750 (2017).
[53] Reich, J., Tamura, H., Nagaoka, I. & Motschmann, H. Investigation of the kinetics and mechanism of low endotoxin recovery in a matrix for biopharmaceutical drug products. Biologicals 53, 1–9 (2018).
[54] Chiominto, A., Marcelloni, A.M., Tranfo, G., Paba, E. & Paci, E. Validation of a high performance liquid chromatography-tandem mass spectrometry method for beta-hydroxy fatty acids as environmental markers of lipopolysaccharide. J. Chromatogr. A 1353, 65–70 (2014).
[55] Uhlig, S. et al. Profiling of 3-hydroxy fatty acids as environmental markers of endotoxin using liquid chromatography coupled to tandem mass spectrometry. J. Chromatogr. A 1434, 119–126 (2016).
[56] Haishima, Y. et al. Chemical and biological evaluation of endotoxin contamination on natural rubber latex products. J. Biomed. Mater. Res. 55, 424–432 (2001).
[57] Cody, R.B., Laramée, J.A. & Durst, H.D. Versatile new ion source for the analysis of materials in open air under ambient conditions. Anal. Chem. 77, 2297–2302 (2005).
[58] Li, H., Hitchins, V.M. & Wickramasekara, S. Rapid detection of bacterial endotoxins in ophthalmic viscosurgical device materials by direct analysis in real time mass spectrometry. Anal. Chim. Acta. 943, 98–105 (2016).
[59] Zdravkovic, S.A. Understanding leaching from stoppers into lyophilized drugs. Biopharm. Int. 32, 21–41 (2019).
[60] Jiang, Y. et al. Tungsten-induced protein aggregation: solution behavior. J. Pharm. Sci. 98, 4695–4710 (2009).
[61] Wang, W., Ignatius, A.A. & Thakkar, S.V. Impact of residual impurities and contaminants on protein stability. J. Pharm. Sci. 103, 1315–1330 (2014).
[62] Pang, J. et al. Recognition and identification of UV-absorbing leachables in EPREX pre-filled syringes: an unexpected occurrence at a formulation-component interface. PDA J. Pharm. Sci. Technol. 61, 423–432 (2007).
[63] Macdougall, I.C. Pure red cell aplasia with anti-erythropoietin antibodies occurs more commonly with one formulation of epoetin alfa than another. Curr. Med. Res. Opin. 20, 83–86 (2004).
[64] Wooltorton, E. Epoetin alfa (Eprex): reports of pure red blood cell aplasia. Can. Med. Assoc. J. 166, 480 (2002).
[65] Hermeling, S., Schellekens, H., Crommelin, D.J. & Jiskoot, W. Micelle-associated protein in epoetin formulations: aA risk factor for immunogenicity? Pharm. Res. 20, 1903–1907 (2003).
[66] Villalobos, A.P., Gunturi, S.R. & Heavner, G.A. Interaction of polysorbate 80 with erythropoietin: a case study in protein-surfactant interactions. Pharm. Res. 22, 1186–1194 (2005).
[67] Ryan, M.H. et al. An in vivo model to assess factors that may stimulate the generation of an immune reaction to erythropoietin. Int. Immunopharmacol. 6, 647–655 (2006).
[68] Pramanik, B.N., Lee, M.S. & Chen, G. (eds.) Characterization of Impurities and Degradants using Mass Spectrometry. (John Wiley, Hoboken, N.J.; 2011).
[69] Nageswara Rao, R. & Talluri, M.V. An overview of recent applications of inductively coupled plasma-mass spectrometry (ICP-MS) in determination of inorganic impurities in drugs and pharmaceuticals. J. Pharm. Biomed. Anal. 43, 1–13 (2007).
[70] Barin, J.S., Mello, P.A., Mesko, M.F., Duarte, F.A. & Flores, E.M. Determination of elemental impurities in pharmaceutical products and related matrices by ICP-based methods: a review. Anal. Bioanal. Chem. 408, 4547–4566 (2016).
[71] Huang, J. et al. The application of inductively coupled plasma mass spectrometry in pharmaceutical and biomedical analysis. J. Pharm. Biomed. Anal. 40, 227–234 (2006).
[72] Jin, C. Clean chemistry for elemental impurities analysis of pharmaceuticals in compliance with USP 232. AAPS Pharm. Sci. Tech. 17, 1141–1149 (2016).

Chapter 7
Quantitation of protein therapeutics in biological samples

During protein therapeutics development, assays that can accurately, reproducibly, and sensitively quantitate protein targets in various biological samples are required to support a wide range of studies, such as toxicokinetic (TK), pharmacokinetic (PK), and biotransformation studies. Ligand-binding assays (LBA) are long regarded as the gold standard for such analyses, which first utilize capture reagents to pull down target molecules from biological matrices and then apply detection reagents for target quantitation. However, mass spectrometry (MS) has also enjoyed a tremendous growth in popularity as a protein quantitation tool, a process that was largely fueled by the dramatic progress of proteomics in the past two decades. Many of these tools proved useful in the field of biopharmaceutical quantitation, and provide unique advantages over the classic assays. In this chapter, we provide a brief overview of the most common MS-based quantitation tools employed in this field, discuss a few promising emerging approaches, and present several specific examples of their uses.

7.1 What motivates the development of MS-based quantitation strategies for protein therapeutics?

LBAs were traditionally considered indispensable protein quantitation tools in PK applications, as well as other areas where reliable measurements of the amounts of specific protein drugs in complex biological matrices are a necessity. Indeed, they exhibit excellent sensitivity and can be applied in a high-throughput format using 96, 384, or even 1,536-well plates, making these methods ideally suited for large sample sets from animal and clinical studies. However, development of LBAs can often be a time-consuming and resource-demanding process, largely due to the difficulty in sourcing and developing suitable capture/detection reagents. In addition, LBAs are generally limited in multiplexing capability (e.g., analyzing multiple targets in the same assay), due to the requirement for highly specific capture and detection reagents to avoid cross reactivity. Therefore, application of these methods for co-formulated or co-administered protein drugs can be challenging and requires even longer development time. Other limitations include limited dynamic range, as well as a possibility of endogenous antibodies interfering with the immunoassays (an interested reader is referred to an excellent review on this subject [3]). Therefore, it is not surprising that MS-based protein quantitation methods are frequently viewed as an attractive alternative and increasingly applied at both preclinical and clinical stages of the protein drug development. Compared to LBAs, MS-based methods exhibit advantages in short development time, minimal requirement for reagents, high specificity, and capability in multiplexing, which makes them ideally suited for programs on tight timelines or with co-formulated or co-administered protein drugs. However, compared

https://doi.org/10.1515/9783110546187-007

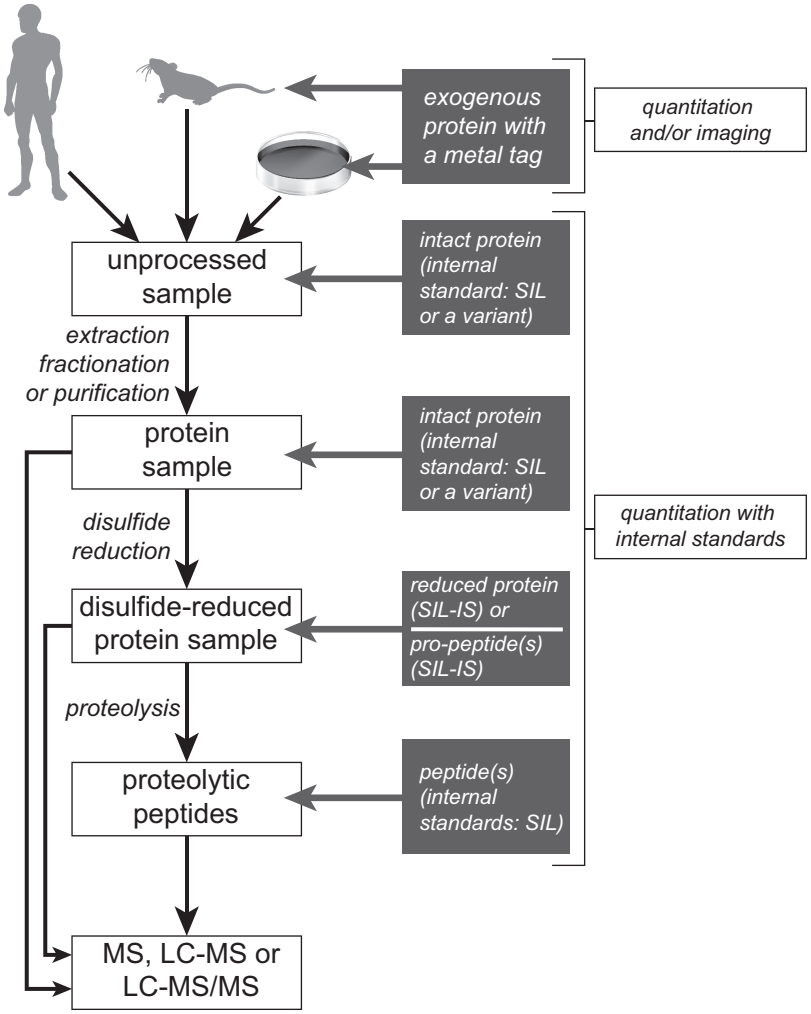


Figure 7.1: A schematic diagram of possible quantitation scenarios and uses of internal standards for protein therapeutics in complex biological matrices. Quantitation not relying on internal standards can be done at any stage.

to LBAs, MS-based quantitation methods suffer from shortcomings such as lower sensitivity and throughput. In the following paragraphs, different MS-based protein quantitation strategies will be discussed with an emphasis on limitations as well as opportunities for further development.

7.2 Quantitation with and without internal standards

Generally speaking, the vast arsenal of protein quantitation strategies in the tool chest of modern proteomics can be divided into two large groups: global (or system-wide, multicomponent) strategies and targeted approaches [4]. While the former can be used at the drug design/development stage (e.g., biomarker discovery and assessment of the treatment's impact on the host's proteome), it is the second group of methods that are most relevant to the quantitation tasks in PK studies. The targeted (single-component or several-component) quantitation strategies focus on one or a few proteins or proteoforms that are selectively isolated from the sample prior to the quantitation step [4]. The latter could be achieved either directly or with the help of an internal standard (IS), which can be executed in a variety of formats (Figure 7.1). Although direct quantitation of specific proteins in complex biological matrices is feasible, application of ISs almost always improves the quantitation performance, and our discussion will be mostly focused on these methods.

Selection of an IS for protein quantitation is a very important step in the development of the quantitation protocol. Since multiple options are usually available (Figure 7.1), pros and cons of each should be carefully considered prior to making the final selection. An ideal IS would have physical and chemical characteristics that are nearly identical to those of the target protein, yet allow distinct signals to be generated for the target protein and the IS. For example, stable isotope labeling (SIL) can be used to generate variants of the target protein that have identical covalent structure (both amino acid sequence and PTMs), but specific amino acids will be isotopically labeled. This can be achieved by modifying the growth media such that one of the amino acids is replaced with its isotopically labeled analog (^{13}C, combined $^{13}C/^{15}N$ or ^{18}O labeling is strongly preferred to ^{2}H labeling, as the latter results in noticeable changes of many physical and chemical characteristics, and frequently affects LC retention). Adding such "whole protein" SIL-IS to the unprocessed sample will expose it to the same complete sequence of sampling handling and processing events as the target protein (Figure 7.1). Since the physical and (bio)chemical properties of the two proteins are nearly indistinguishable from each other, any losses encountered throughout the entire downstream procedure (e.g., due to imperfect recovery during the purification step, or incomplete proteolysis) would affect them to the same extent. Therefore, relating the MS signal of the target protein to that of the SIL-IS at the final (MS measurement) step would allow accurate numbers to be obtained for the target protein quantity present in the original unprocessed sample, even if only a fraction of the protein molecules "makes it" throughout the entire procedure.

The major downside of the strategy outlined in the preceding paragraph is the cost/time of the whole protein SIL-IS production. A cost-effective alternative uses synthetic peptides incorporating one isotopically labeled amino acid. The peptide sequence is selected such that it matches the sequence of one of the proteolytic fragments of the target protein generated prior to the LC-MS or LC-MS/MS analysis

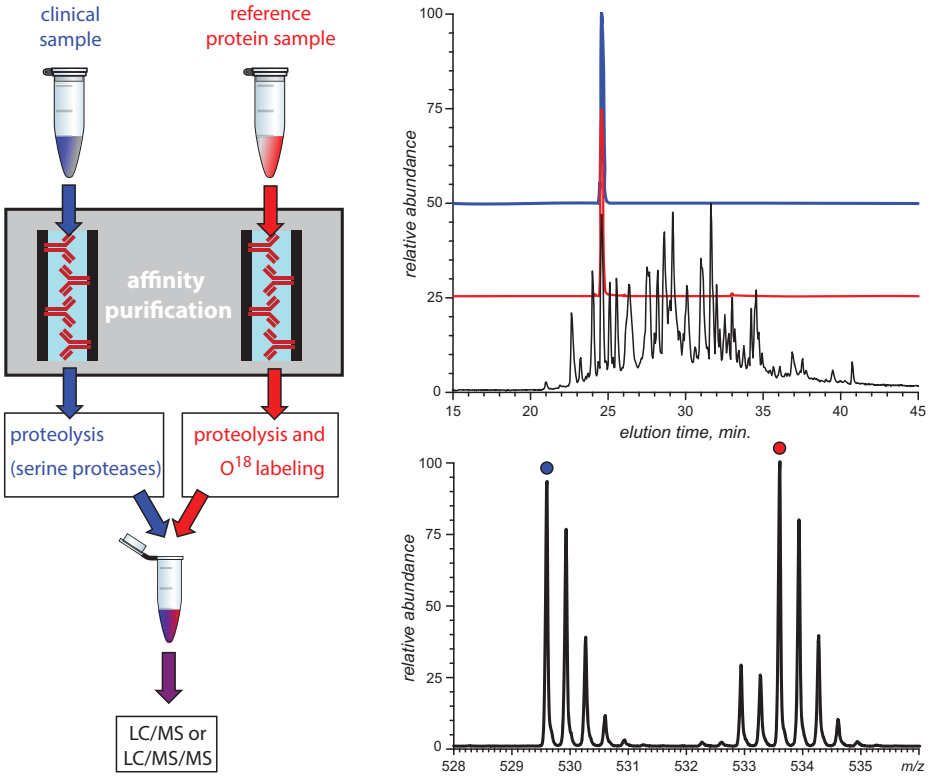


Figure 7.2: A schematic diagram showing generation of ^{18}O-labeled internal standard for protein quantitation in complex biological matrices using a surrogate peptide approach. Left: affinity purification is used to extract the target protein from the biological sample, followed by its proteolysis and LC-MS or LC-MS/MS detection. Proteolysis of the internal standard is carried out in $H_2{}^{18}O$, which introduces an isotopic tag within each fragment peptide (by replacing ^{16}O atoms with ^{18}O at their C-termini) without changing its LC elution profile [5]. Right: a total ion chromatogram of a tryptic digest of exogenous (recombinant) human transferrin extracted from serum (black trace); the blue and red traces correspond to extracted ion chromatograms for ions representing tryptic fragment peptides KPVEEYANCHAR derived from the target protein (blue) and the internal standard (red). The mass spectrum at the bottom of the diagram shows isotopic clusters of these two peptides (charge state +3). To avoid any signal interference, the spacing between the two isotopic clusters was increased by acid-catalyzed ^{18}O-labeling of IS tryptic fragments (which places ^{18}O atoms on acidic side chains [6]) prior to LC-MS analysis. Adapted from Wang et al. [7] with permission from Elsevier.

(Figure 7.1). This proteolytic fragment (frequently referred to as the surrogate peptide) is then used as a single reporter of the entire protein. Obviously, the sequence of the surrogate peptide must be specific to the target protein (e.g., a surrogate peptide cannot be derived from the constant region of a mAb molecule, or else there would be significant interferences from endogenous antibodies). The most significant downside of this

approach stems from the fact that this SIL-IS is introduced very late (following the completion of the sample work-up and processing), and would not compensate for the losses of the target protein that have occurred upstream. However, some improvements are possible, and will be discussed in detail in the following sections; we will also consider some "compromise" strategies, in which IS introduction "mid-stream" allows some (but not all) factors resulting in the target protein loss to be accounted for.

7.3 Protein quantitation using surrogate peptides

In this approach, a unique signature peptide, generated by enzymatic digestion of protein target, is used as the surrogate for quantitation. Subsequently, a targeted LC-MS method is applied to monitor the surrogate peptide at either the intact peptide ion level or its fragmentation product ion levels. One cost-effective strategy that is particularly useful in proteomic applications, but can also be applied for quantitation of therapeutic proteins relies on peptide standards that are produced enzymatically. Carrying out protein digestion with serine proteases (such as trypsin and Lys-C) in $H_2{}^{18}O$ for an extended period of time results in isotopic exchange of both carboxylate oxygen atoms at the C-terminus of each proteolytic fragment [5]. This produces a mass tag of 4 Da without any chemical modification of the peptide. The extent of the exchange (and the magnitude of the resulting mass shift) can be increased even more by involving carboxylate groups of the aspartic acid and the glutamic acid side chains [6]. The $^{16}O/^{18}O$ exchange has minimal impact on the physical and chemical properties of the proteolytic fragments apart from the mass change. Importantly, it does not affect the chromatographic elution time, allowing both the unlabeled peptide fragment derived from the target protein and the SIL-IS peptide to be co-eluted, while their ionic signals in MS are clearly distinct from each other due to the mass difference (Figure 7.2).

By far, stable isotope-labeled synthetic peptides are most frequently used due to their relatively low cost and wide availability. However, as peptide IS can only be added to the protein sample after the digestion step, variations occurring at any of the prior sample processing steps cannot be accounted for and corrected. In particular, the digestion efficiency, which cannot be normalized using a peptide IS, might contribute significantly to the variability of quantitation. One particularly effective mitigation strategy to address this issue involves the use of a synthetically prepared and isotopically labeled "pro-peptide," whose sequence is extended at both N- and C-termini by a few amino acid residues compared to the sequence of the surrogate peptide representing a specific proteolytic fragment [9]. This SIL-IS pro-peptide is introduced into the protein sample prior to the digestion step, and is proteolytically processed concurrently with the target protein, thereby reducing the measurement errors caused by incomplete digestion.

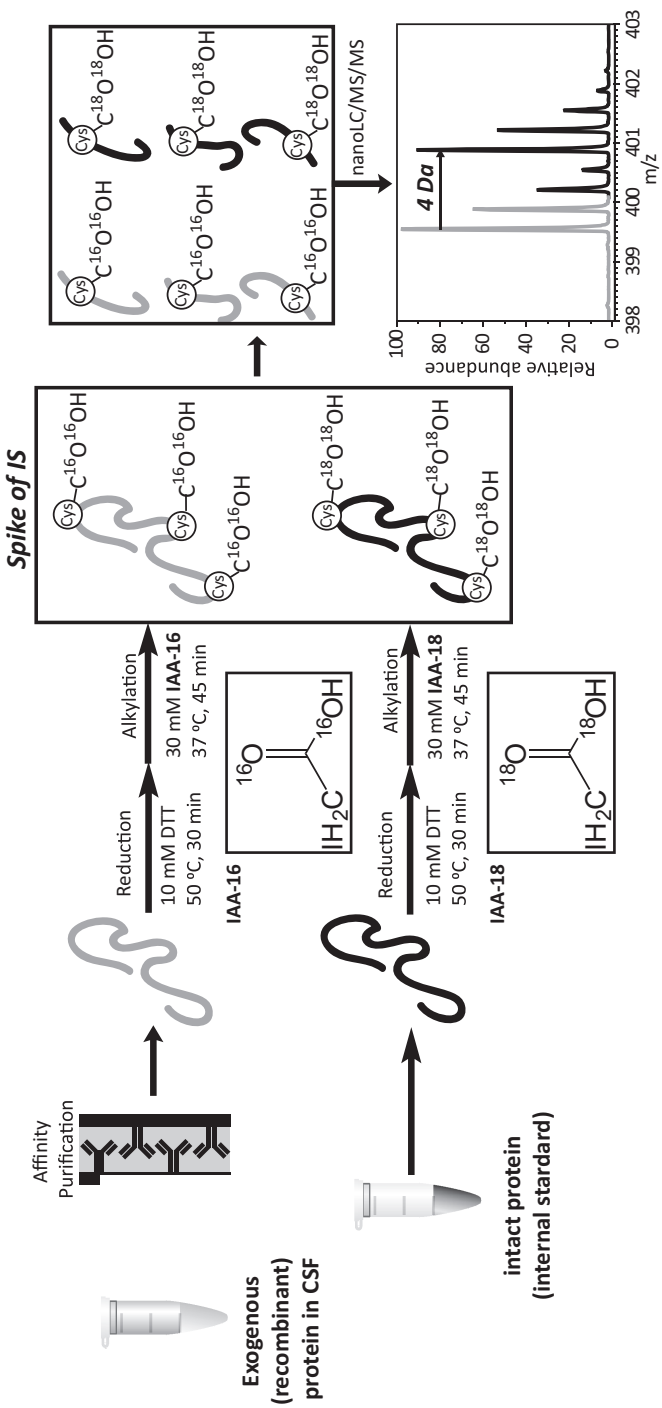


Figure 7.3: A schematic diagram of a protein quantitation workflow using internal standards generated by alkylating cysteine residues of an intact disulfide-reduced protein with isotopically labeled iodoacetic acid. Adapted with permission from Wang et al. [8].

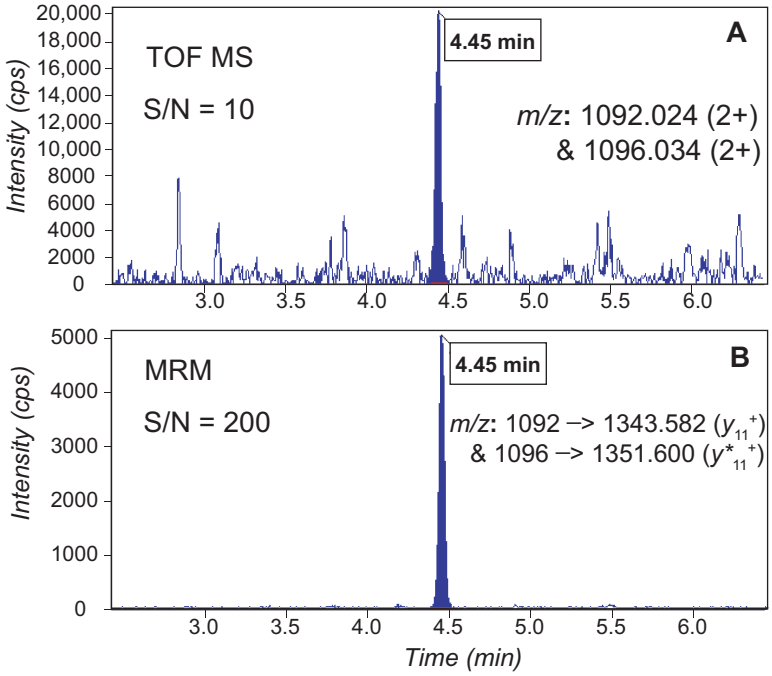


Figure 7.4: A comparison of protein quantitation based on signal measurements for a surrogate peptide using a high-resolution LC-MS (A) and multiple reaction monitoring with LC-MS/MS (B). The internal standard (a surrogate peptide) representing a tryptic fragment of the heavy chain of Rituximab™ GLEWIGAIYPGNGDTSYNQK was synthesized as a pre-peptide (PGR**GLEWIGAIYPGNGDTSYNQK***FAG, where K* denotes a $^{13}C_6$-, $^{15}N_2$-labeled lysine residue) and added to the protein sample prior to the proteolytic step to account for possible biases due to incomplete digestion. Adapted from Mekhssian et al. [9]; permission conveyed through Copyright Clearance Center, Inc.

Other modifications of the protein quantitation strategies based on surrogate peptides that move the SIL-IS introduction point further upstream are also possible. For example, a mass tag can be introduced at the intact protein level by chemically modifying it with a stable isotope-labeled reagent. For example, the cysteine alkylation step, which is frequently done prior to the enzymatic digestion, provides an excellent opportunity to introduce a stable isotope-labeled mass tag for the quantitation purposes. Such reagents can be readily obtained from commercial sources, such as irreversible isobaric iodoacetyl Cys-reactive tandem mass tag (iodoTMT) [10] reagents and a variety of stable isotope-labeled iodoacetamide molecules. In addition, by incubating iodoacetic acid (IAA) in ^{18}O-enriched water, ^{18}O-labeled iodoacetic acid (IAA-18) can be readily prepared in house at low cost and used for protein quantitation (Figure 7.3) [11]. In a workflow of LC-MS-based quantitation of a recombinant form of an 80 kDa protein human transferrin (hTf) in cerebrospinal fluid of an animal model, the intact protein was alkylated with ^{18}O-labeled iodoacetic acid

(IAA-18), and subsequently used as a protein IS to quantify exogenous hTf that was subsequently alkylated with regular iodoacetic acid (IAA-16) during sample preparation. The two reduction protocols were otherwise identical, and the IS was added to the target protein sample prior to the tryptic digestion step; a Cys-containing peptide was selected as the surrogate for protein quantitation with LC-MS [12]. Although the choice of surrogate peptides using this strategy is limited to Cys-containing peptides, the frequent occurrence of Cys in therapeutic proteins makes this approach a valuable alternative for protein quantitation.

By far, multiple reaction monitoring (MRM) assay conducted with a triple quadrupole MS system has been most commonly and widely applied for protein quantitation due to its excellent specificity and robustness. In this assay, the precursor ion of the surrogate peptide is first isolated in the first quadrupole (Q1), fragmented within Q2, and a predefined product ion is then monitored in Q3. The two ion selection steps, along with the highly reproducible retention time from a modern LC system, make MRM assays highly specific. Further improvements can be provided by employing high-resolution mass spectrometers capable of executing MRM routines, such as TOF/TOF MS. This is illustrated in Figure 7.4, which shows quantitation of Rituximab™ in human plasma using a surrogate peptide GLEWIGAIYPGNGDTSYNQK (the SIL-IS was introduced to the target protein sample prior to its digestion with trypsin as a pro-peptide PGR**GLEWIGAIYPGNGDTSYNQK***FAG with the $^{13}C_6$-, $^{15}N_2$-tag at the lysine residue). The high resolving power of the TOF mass analyzer allows the extracted ion chromatogram to be plotted using a narrow ion selection window (0.020–0.030 m/z unit); nevertheless, signal-to-noise ratios better than 10 cannot be obtained in the LC-MS mode due to the presence of chemical noise (Figure 7.4A). In contrast, switching to the LC-MS/MS mode and using MRM for quantitation (e.g., by simultaneously monitoring the $[GLEWIGAIYPGNGDTSYNQK+H]^{2+} \rightarrow y_{11}^{+}$ and $[GLEWIGAIYPGNGDTSYNQK^{*} + H]^{2+} \rightarrow y^{*}_{11}{}^{+}$ transitions) results in a 20-fold improvement of the signal-to-noise ratio despite the apparent loss of the absolute signal intensity (Figure 7.4B).

Compared to LBAs, MRM-based methods are often less sensitive and offer significantly lower throughput. The sensitivity of MRM assays can be significantly improved by adopting a hybrid assay format, where an immunoprecipitation step is introduced into the workflow to enrich the target protein. However, the throughput of MRM assays remains the bottleneck that limits their broader applicability in preclinical and clinical studies. Unlike LBAs, LC-MS-based protein quantitation can only be performed one sample at a time. In addition, a lengthy LC separation step (>10 min) is frequently required in order to reduce the background of interfering species that co-elute with the surrogate peptides. Furthermore, the run-to-run carryover is another concern for LC-MS-based protein quantitation that often requires blank injections between sample runs, leading to even longer cycle time. To improve the throughput, an attractive strategy is to eliminate the LC separation step. Instead, the sample delivery to MS may be achieved by various high-speed sample introduction techniques, such as RapidFire by Agilent

and Echo MS by Sciex. However, without LC separation step, the method sensitivity may be compromised due to increased matrix complexity. In a proof-of-concept study, this issue was overcome by a combination of offline fractionation of surrogate peptide with reversed-phase solid-phase extraction (RP SPE) and high-resolution MS/MS analysis (Figure 7.5) [1]. By applying optimized wash and elution conditions, SPE fractionation of the surrogate peptide reduced the complexity of the sample matrix and therefore improved the sensitivity. In addition, SPE fractionation can be readily performed in a 96 well-plate and therefore significantly improve the throughput. The MS/MS data acquisition step can use a range of high-resolution MS instrumentation (the high resolving power of the MS analyzer is essential, as it allows the surrogate peptide signal to be resolved from interferences). This proof-of-concept study successfully demonstrated that high-speed quantitation of human mAb in monkey serum (1 min/sample) can be achieved by eliminating the LC step and reach lower limits of quantitation (LLOQ) of 1 µg/mL, comparable to conventional LC-MS/MS based methods [13]. Nevertheless, the throughput will remain in the near future one of the most significant challenges for MS-based protein quantitation methods.

Another important factor that needs to be taken into account when considering surrogate peptides as a means of protein quantitation is the sensitivity of this approach to post-production PTMs that occur within the protein segment represented by the surrogate peptide, and insensitivity to PTMs occurring elsewhere in the sequence. In the example of Rituximab^{TM} quantitation considered earlier in this section, the surrogate peptide GLEWIGAIYPG**N**GDTSYNQK contains a likely site of deamidation; however, should this PTM affect a fraction of mAb molecules either *in vivo* or during any of the sample handling steps prior to the protein digestion, none of the

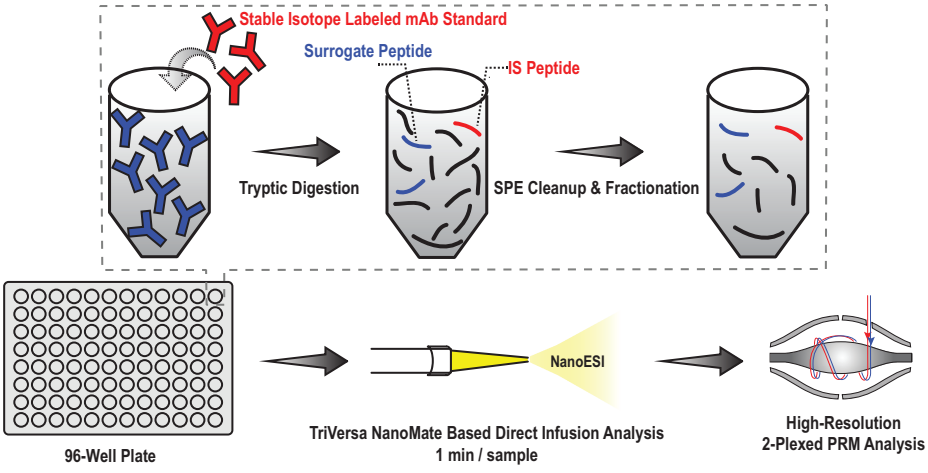


Figure 7.5: Workflow of an LC-free approach for high-throughput bioanalysis of mAbs. Adapted with permission from Zhang et al. [1].

affected protein molecules will be counted, as the corresponding tryptic fragment (GLEWIGAIYPG**D**GDTSYNQK) will have different elution time and mass compared to the surrogate peptide. Only deamidation events occurring during the proteolytic step will be accounted for, since this particular IS is introduced into the protein sample as a pro-peptide prior to the digestion step (vide supra). On the flip side, any PTMs occurring outside of the protein segment represented by the surrogate peptide will go undetected and unreported. Likewise, any information on the possible differences among the PK profiles exhibited by various proteoforms (e.g., mAb glycoforms) will be lost when the surrogate peptide approach is used for quantitation.

7.4 Protein quantitation at intact and subunit levels

Historically, surrogate peptide-based strategy has been the method of choice for MS-based quantitation of protein drugs in complex biological samples. It is largely due to the fact that MS detection of small peptides is much more sensitive than that of large proteins [14]. During MS analysis, the ionization, transmission and detection processes of relatively small peptide ions are generally more efficient compared to large proteins. In addition, large proteins often exhibit a wide charge state envelope that further dilutes peak intensity at any given charge state. The intrinsic mass heterogeneity of intact proteins (e.g., glycan heterogeneity, N- and C-terminal heterogeneity) can further complicate the quantitation and significantly limit the detection capability. On the other hand, despite being significantly more sensitive, surrogate peptide-based approaches require lengthy and complicated sample preparation. More importantly, valuable information on the structural integrity of protein drugs is often lost due to the proteolytic digestion. For example, quantitation of an antibody–drug conjugate (ADC) in PK studies using the surrogate peptide approach is not able to reveal the possible changes in drug-to-antibody ratio (DAR), which is a critically important attribute that can only be assessed at the intact protein level. Therefore, it is not surprising that MS-based protein drug quantitation has been increasingly carried out at the intact protein level, particularly during the preclinical stage of the drug development. To compensate the reduced sensitivity compared to peptide-level analysis, MS-based quantitation of intact proteins in complex biological samples is almost always performed in combination with an immunoprecipitation step, where the target proteins are isolated and enriched prior to MS analysis [15]. An example of such analyses is shown in Figure 7.6, where magnetic streptavidin beads bound with a biotinylated goat anti-human IgG Fc polyclonal antibody were used to capture both the “analyte” mAb and its isotopically labeled version employed as an IS, allowing a quantifiable signal to be observed for mAb concentrations in serum as low as 1 μg/mL.

A careful examination of the data presented in Figure 7.6 makes it apparent that the quantitation efforts at the intact molecule level are negatively affected by structural heterogeneity of the protein, i.e., the presence of multiple glycoforms results in

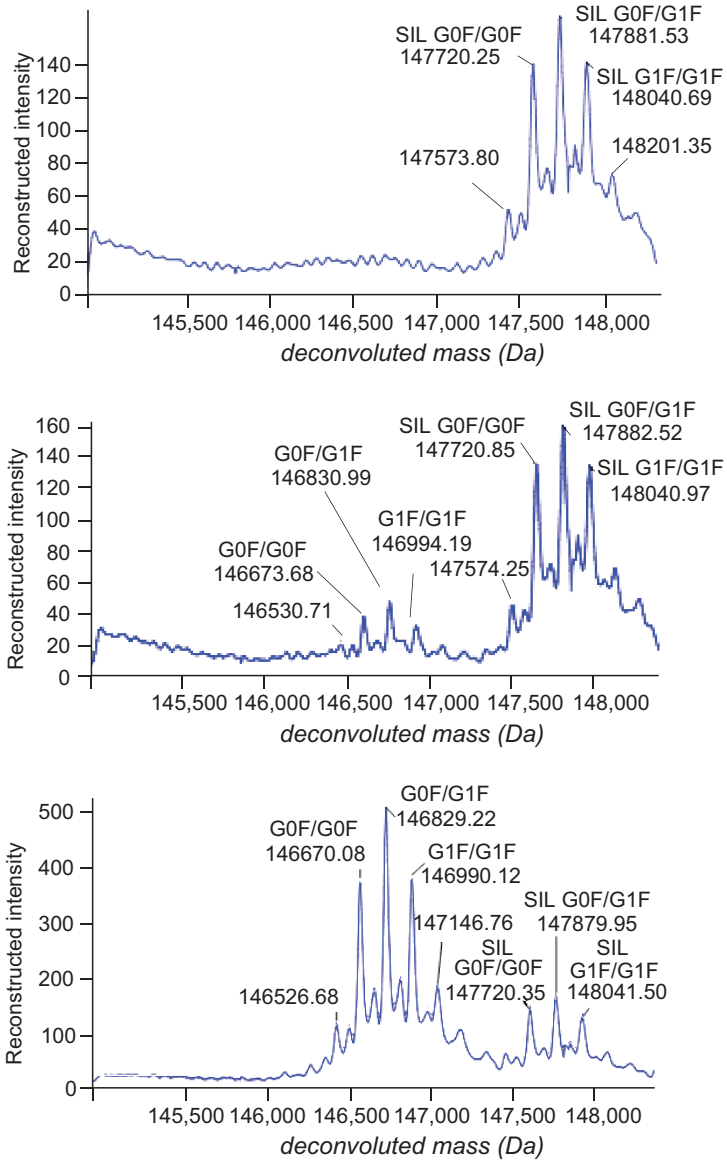


Figure 7.6: Deconvoluted mass spectra of antibodies extracted from mouse serum spiked with varying quantities of a SILu^{TM}Lite mAb (***top***: blank; middle: 1 µg/mL; bottom: 15 µg/mL) and ($^{13}C_6$-, $^{15}N_4$-arginine/$^{13}C_6$-, $^{15}N_2$-lysine)-labeled SILu mAb used as an internal standard. The analyte and the SIL-IS were captured using magnetic streptavidin beads bound with a biotinylated goat anti-human IgG Fc polyclonal antibody, followed by RPLC-MS (full scan) detection. Adapted from Jian et al. [15]; permission conveyed through Copyright Clearance Center, Inc.

the overall signal intensity being distributed among several protein species with different masses. This can be alleviated to a certain degree by employing various sample treatment strategies that aim to reduce either the mass heterogeneity or the size of protein analytes (or both), thereby improving the sensitivity of MS-based quantitation. For example, removing the Fc N-glycosylation from therapeutic mAbs by PNGase F is a simple and effective way to improve their quantitation by reducing the mass heterogeneity. Similarly, compared to intact mAbs, quantitation of Fab_2 (or Fab) fragments after IdeS (or limited Lys-C) digestion is an attractive alternative, as it not only reduces the mass heterogeneity (e.g., due to Fc N-glycosylation and incomplete C-terminal Lys processing) but also reduces the size of the analytes (Figure 7.7). These sample treatment steps (e.g., deglycosylation and IdeS digestion), along with the immunoprecipitation step, can now be readily implemented using various automated platforms in a high-throughput manner [16].

While both examples discussed in the preceding paragraph utilized the stable isotope-labeled ISs, quantitation of intact protein analytes can also be achieved directly, thanks to the continuous improvement in performance stability of LC-MS systems. The latter would eliminate the need for ISs, which would result in a significant simplification of the quantitation protocol. This has been demonstrated by several groups in studies of quantitating mAbs in matrices as complex as serum at both intact mAb level [15] and its fragments that are generated by disulfide reduction and IdeS digestion [18]. In a typical protocol, antibodies are extracted from the biological sample by means of affinity capture, followed by their minimal processing and LC-MS analysis. Extracted ion chromatograms for the target mAb-specific ions are used to obtain an absolute measure of the protein concentration (Figure 7.8). Although it is possible to bypass the protein processing step altogether [15], the resulting sensitivity is modest due to the signal split among multiple protein glycoforms (vide supra). Minimal processing of the extracted mAb molecules (disulfide reduction and digestion with IdeS to generate large antibody fragments, such as the intact light chain, as well as the Fc/2 and Fd segments) results in a 10-fold sensitivity improvement (when using either the light chain or the Fd segment for quantitation). At the same time, it still provides an option of making measurements in a glycoform-specific fashion (by monitoring the signal of Fc/2, an IgG segment incorporating its glycosylation site). Currently, the LLOQ values for the quantitation of intact mAb stand at 50 ng/mL [19], and this figure of merit is likely to be improved due to the continuous technological advances in both MS hardware and front-end protein handling methods.

By far, reversed-phase chromatography (RPLC) is most commonly used in conjunction with MS for protein drug quantitation at intact levels. Although RPLC is highly compatible with MS detection, limited separation efficiency is generally expected for large protein analytes. In fact, for protein quantitation purpose, RPLC methods frequently use very fast gradients and are employed solely for online desalting prior to the MS detection. When developing RPLC-MS based protein quantitation methods, it is important to find a balance between MS sensitivity and chromatographic

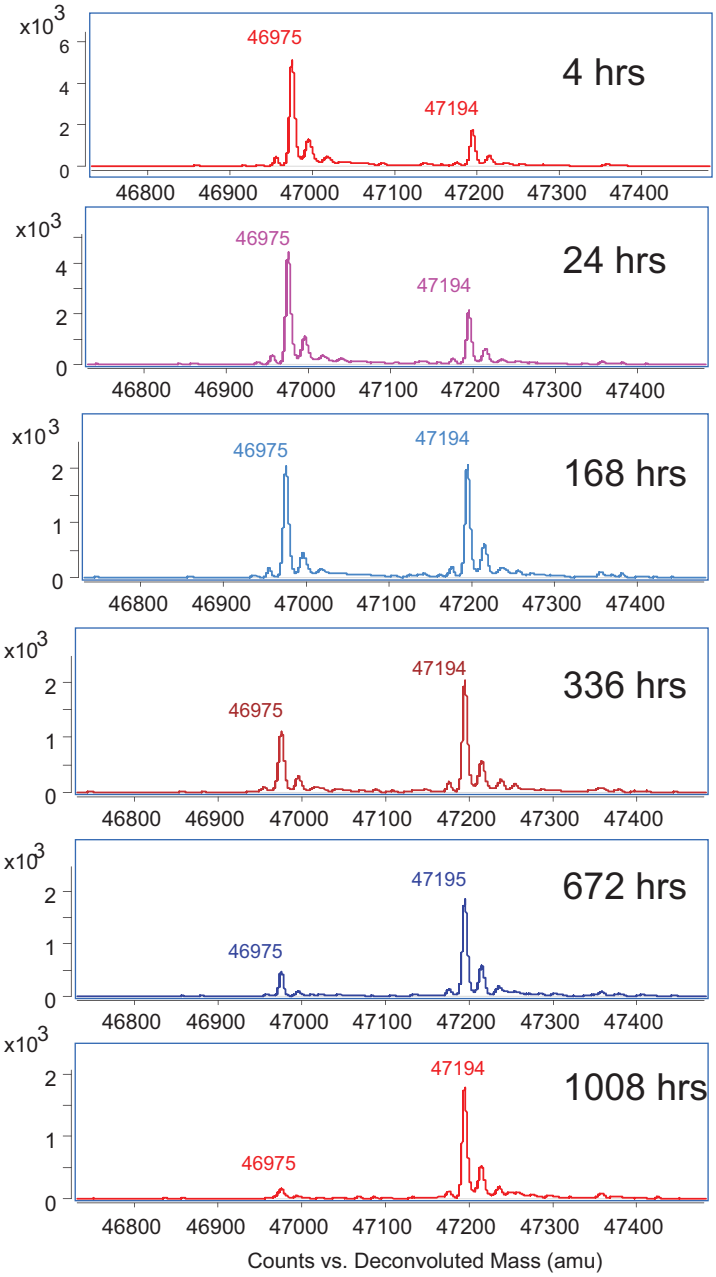


Figure 7.7: Mass spectra of Fab fragments produced by limited Lys-C digestion of mAb molecules extracted from serum samples collected 4, 24, 168, 336, 672, and 1,008 h after administration and spiked with a fixed amount of the stable isotope-labeled antibody. Reproduced from Liu et al. [17] with permission from Elsevier.

performance. For example, formic acid (FA) is widely accepted as an ideal mobile phase additive for RPLC-MS analysis, as it yields excellent MS sensitivity. However, FA-based RPLC methods often exhibit broad elution profiles and poor recovery of large protein analytes, even at elevated column temperatures. In contrast, trifluoroacetic acid (TFA) is a much more effective ion-pairing reagent that can significantly improve both the peak shape and recovery of protein analytes. Unfortunately, TFA is also notorious for its ionization suppression effect in ESI-MS, and therefore compromising MS sensitivity. To overcome this dilemma, a common strategy is to simply use a mixture of the two acids (e.g., 0.1% FA and 0.02% TFA) as mobile-phase additives, although the exact recipe needs to be optimized on a case-by-case basis [20]. An interesting alternative, as recently demonstrated in ADC analysis, is to use

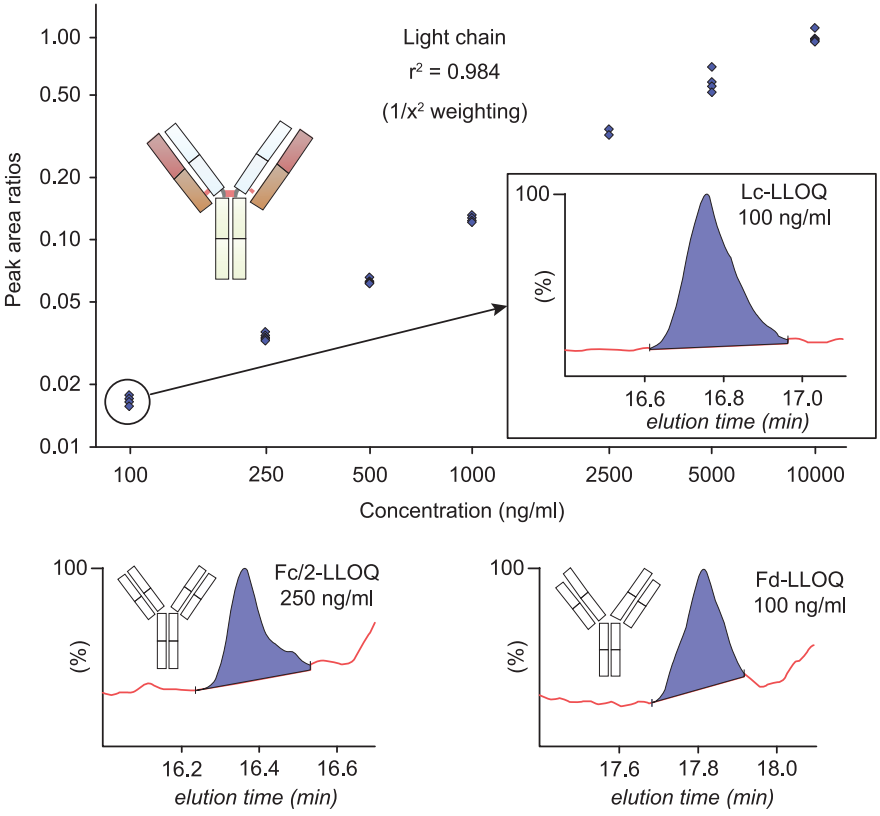


Figure 7.8: Quantitation of mAb in serum without using IS illustrated by constructing an external calibration curve based on LC-MS measurements of antibodies following their affinity capture from human serum and IdeS digestion. Top: the standard curve for the light chain is shown along with the light chain LLOQ LC-MS peak (inset). Bottom: the LLOQs peaks from the Fc/2 (left) and the Fd (right) regions for the biotherapeutic mAb, respectively. Reproduced with permission from Kellie et al. [18] Copyright 2020 American Chemical Society.

difluoroacetic acid (DFA) as the mobile phase additive. In this study, DFA was shown to increase the MS sensitivity by threefold while maintaining LC resolution comparable to TFA-based analysis [21]. Capillary electrophoresis coupled with MS (CE-MS) is another attractive alternative to quantitate intact protein drugs in biological samples. Recent advances in both methodology and instrumentation of CE-MS have led to significantly improved sensitivity and robustness that are critical for protein drug bioanalysis. Compared to RPLC-MS, the limited loading capacity (nanoliter range) of CE-MS has both advantages and disadvantages. First, the extremely low sample consumption afforded by CE-MS allows replicate analyses to be readily implemented. On the flipside, it also impacts the limit of quantitation of protein targets in complex biological samples, as only a small fraction of the sample can be utilized. To overcome this issue, different strategies that aim at increasing the loading capacity in CE-MS can be applied, such as the transient capillary isotachophoresis/capillary zone electrophoresis (CITP/CZE, or t-ITP) technique [22] and the dynamic pH junction technique [23]. It is expected that novel strategies to increase the sample loading in CE-MS will remain one of the key development areas to make this technique better suited for protein bioanalysis.

By far, intact mass analysis-based quantitation of protein drugs is most frequently carried out using conventional ESI-MS under denaturing conditions (low pH solvent with high organic co-solvent content applied either as a mobile phase in RPLC-MS or as sheath liquid in CE-MS). Indeed, intact mass analysis under denaturing conditions is a much more established and widely applied technique in biopharmaceutical laboratories compared to native MS, which has only recently started to gain popularity in the biopharma sector (see Chapter 4 for more details). Comparing to conventional ESI-MS, native MS generates far fewer charge states that in theory should be beneficial for the method sensitivity, as the analyte signal is distributed among fewer charge states. However, it is also well known that ionization of large protein molecules in aqueous and neutral pH solution, as well as the transmission and detection of low charge-density (high m/z) ions generated by native MS are generally far less efficient compared to the analysis of highly charged ions in conventional ESI-MS. Furthermore, more extensive adduct formation in native MS frequently interferes with the detection of multiple proteoforms. Therefore, at the moment, conventional ESI-MS is almost exclusively employed in intact mass analysis-based protein drug quantitation workflow due to its superior sensitivity and availability. Nevertheless, recent advances in both instrumentation (e.g., commercial instruments designed for native MS applications) and methodology have brought an exciting opportunity of employing native MS in protein quantitation. For instance, using an online buffer exchange native MS workflow, van Aernum et al. have recorded a detection limit of 39 ng of NISTmAb [24]. Although the achieved sensitivity was still decidedly lower compared to the conventional RPLC-MS-based approaches (e.g., LLOD ranging from picograms to low nanograms of neat mAbs) [19], several unique advantages could be potentially offered by the native MS-based workflow. First, although RPLC-MS analysis of intact proteins can be completed

in a considerably shorter run time compared to analysis of digested peptides, a typical RPLC method can still easily exceed 10 min per sample, considering the time needed for column equilibration, protein elution and column cleaning. In addition, to minimize protein carryover (which is a common issue for all RPLC methods, and especially those targeting intact proteins), blank injections between sample runs are frequently required, which further increases the cycle time. Conversely, native MS analysis of protein analytes can be achieved in a significantly reduced run time with negligible concerns of protein carryover. This is demonstrated in Figure 7.9, where Yan et al. performed online native MS analysis of mAbs on an SEC guard column (4.6 mm × 30 mm) using a range of analytical flow rates [2]. At 0.8 mL/min, each analysis could be completed within 42 s with excellent peak shape, highlighting potential applications in high-throughput protein quantitation. Furthermore, native MS-based protein quantitation methods might be particularly advantageous for bioanalysis of Cys-linked ADC molecules, where the structural integrity of the analytes can be maintained during MS analysis, and therefore

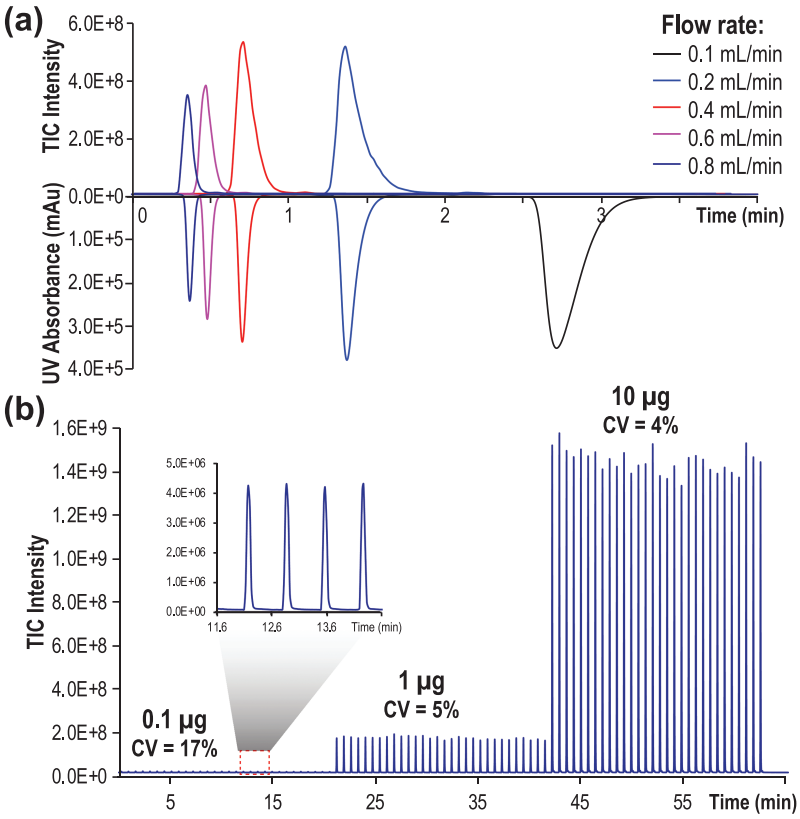


Figure 7.9: High-throughput online native MS analysis of a therapeutic mAb carried out with an SEC guard column at a range of flow rates. Adapted from Yan et al. [2]; permission conveyed through Copyright Clearance Center, Inc.

possible changes in the drug–antibody ratio (DAR) can be readily determined at the intact molecule level (as opposed to monitoring the heavy and light chain levels separately by the RPLC-MS methods). Finally, because of the reduced number of charge states in native MS analysis, it is possible that protein quantitation could be carried out by monitoring a few selected charge states or their fragment ions, using strategies similar to the surrogate peptide analysis (e.g., selected ion monitoring or selected reaction monitoring). The low charge density of protein ions produced in native MS has a negative impact on fragmentation efficiency, but it can be improved by using the supercharging phenomenon [25, 26]. Such strategies are expected to significantly improve both sensitivity and specificity, which has already been demonstrated in an RPLC-based top-down intact mass workflow for protein quantitation [27]. In sum, although native MS is an emerging technique in industry, we believe it presents an exciting opportunity for protein drug quantitation in biological samples. A more detailed discussion of the current practices, limitation, and future trends in quantitative bioanalysis carried out at the intact molecule level can found in a recent review articles [28, 29].

7.5 Protein quantitation using metal tags

A very different approach to quantitation of proteins in complex biological matrices takes advantage of the ability of inductively coupled plasma (ICP) MS to measure concentration of a range of metals with very high precision and accuracy, and at very low levels (down to the parts per trillion level). Attaching a metal tag to a protein enables its detection by ICP MS at very low levels (as long as this metal is not endogenously present in the biological matrix), eliminating the need for protein extraction [30, 31]. However, labeling any protein with a metal tag invariably affect its structure and, therefore, will have an impact on its biodistribution. It follows that successful utilization of ICP MS as a tool for absolute quantitation of therapeutic proteins in biological samples hinges on the availability of methods to place a metal tag on a protein in a manner that would be minimally disruptive to its higher order structure and the ability to interact with its physiological partners. Although this task may seem trivial for metalloproteins (where substitution of the cognate metal with an exogenous one will do the trick [32, 33]), such a strategy cannot be applied to the vast majority of therapeutic proteins. It may be possible to circumvent this problem by taking advantage of the His-tag segments in recombinant proteins, as such poly-histidine sequences can be selectively and irreversibly labeled with ruthenium, enabling their detection with ICP MS [34]. While this approach offers sensitivity that is far greater than what can be achieved with the LC-MS-based absolute quantitation methods discussed earlier in this section, it suffers from two significant drawbacks. First, the eventual catabolism of the protein by the organism will generate a population of metal atoms (e.g., ruthenium) no longer associated with the protein. Second, even

though the presence of a His-tag is not expected to have a significant effect on the protein higher order structure and its biodistribution, it has the potential to alter the PK profile and, therefore, compromise the results of the quantitation work. Lastly, the reliance on a single metal label as a means of protein detection will not allow the same level of specificity to be achieved that is offered by the intact-molecule LC-MS analyses (vide supra). It will be interesting to see if these shortcomings could be addressed in the near future, paving the way for the ICP MS inclusion into the analytical toolbox in biopharma.

7.6 Biodistribution study of protein drugs by mass spectrometry imaging

Understanding the biodistribution of protein drugs in tissue after administration plays a vital role during their discovery and development. Such knowledge not only facilitates the understanding of pharmacological responses but also helps to identify off-target binding-induced toxicity. Biodistribution study of protein drugs often relies on molecular imaging techniques, where the reporter labels (e.g., radionuclides or near-infrared fluorescent tags) are first introduced to the protein drug of interest and used as the surrogate for subsequent detection [35, 36]. Because of the high sensitivity and dynamic range, these approaches have been the industry gold standards to study tissue distribution during drug discovery and development. However, as the detection is performed indirectly and relied on the presence of molecular labels, several shortcomings are inherent to these approaches. First, the labeling process can potentially perturb the structure and function of a protein drug molecule [37]. Therefore, the labeled drug might not fully represent its unmodified counterpart. Second, the labels, due to the added properties (e.g., size, charge, and hydrophobicity), might also change the protein biodistribution in certain tissues [38]. Finally, as the protein drugs break down via catabolism, the remaining labels can still be detected and can confound the data interpretation, as they no longer represent the location and quantity of active drugs.

The ICP MS technique discussed in the preceding paragraph can be interfaced with laser ablation (LA), offering an opportunity to obtain 2-D distributions of metal-tagged proteins in tissue cross-sections [40]. While the use of LA-ICP MS imaging of metal-labeled recombinant (exogenous) proteins within tissue sections of model animals has been reported [32], a more popular approach to obtaining 2-D distribution maps of various proteins within tissue sections uses a different strategy, termed mass spectrometry-immunohistochemistry [39]. In this approach (schematically illustrated in Figure 7.10), detection of a particular protein is enabled by its interaction with an antibody tagged with a rare earth metal. Antibodies targeting different proteins can be used in a single assay, and as long as each antibody is labeled with a unique metal

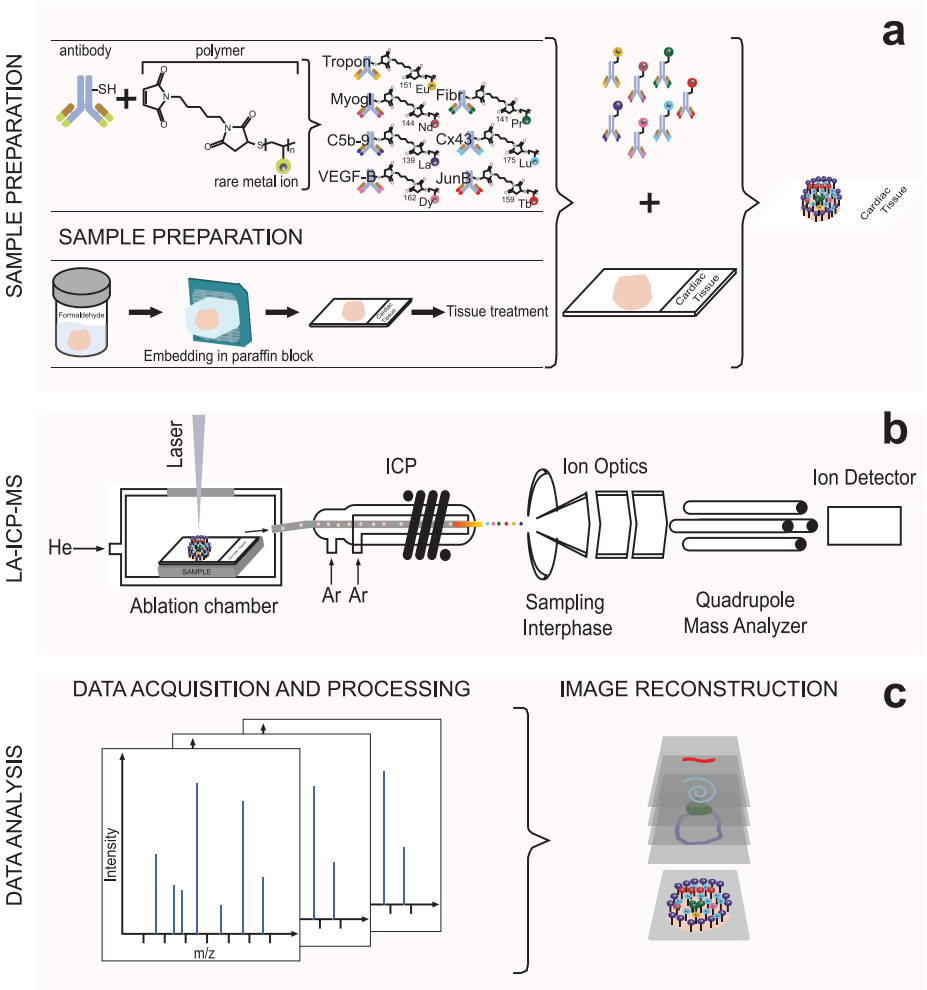


Figure 7.10: A Workflow for quantitative multiplex imaging using a combination of rare-metal-isotope-tagged antibodies and LA-ICP-MS. Sample preparation (A): the primary antibodies are tagged with a complex consisting of a polymer chelated with a rare metal. The tissue sample fixed in formaldehyde is embedded in paraffin, sectioned, placed onto a slide, and processed (deparaffinization, epitope retrieval, etc.). Multiplex hybridization is then performed: tissue section is incubated with a mixture of uniquely labeled antibodies. Data acquisition (B): the sample is placed into the laser ablation chamber and the sample surface is scanned using focused laser beam (line by line). The laser ablation process generates particles, which are transported to the ICP. The released metal tags are detected and recorded as mass spectra. Data analysis (C): the acquired data set is analyzed to create ion image reconstruction and obtain quantitative information on localization of different proteins within the tissue section. Reproduced by permission from Springer Nature Customer Service Centre GmbH: Ajakna et al. [39].

tag, highly selective protein detection and localization can be achieved. While this technique is rapidly gaining popularity in the field of pathology (as well as other areas where immunohistochemistry is a commonly used tool), its applications as an imaging tool for biodistribution studies of protein drugs are lagging. One of the likely reasons is its reliance on highly specific antibodies, for example, those being able to distinguish between exogenous antibodies (mAbs) and those produced by the host organism.

Another emerging MS imaging (MSI) technique – MALDI-MSI [41] – also shows a tremendous promise and a potential to overcome some of the limitations besetting the "classical" imaging techniques. Because the MS detection in MALDI MSI is directly performed on the protein or its proteolytic digests, no labeling is required. In addition, if performed at intact protein levels, it is possible that MALDI MSI can even discriminate between the parent protein drug and its catabolites. Furthermore, the multiplexing capability afforded by MS analysis makes it possible to trace many different analytes simultaneously. Despite the great promise and notable success in the field of small-molecule drugs [42], MSI is yet to demonstrate its true potential in the biodistribution studies of intact protein drugs [43]. The slow progress is largely attributed to the significant challenge of detecting and identifying large protein molecules that are present in small quantities within complex sample matrices. To date, only one successful MALDI MSI study of a protein drug biodistribution *ex vivo* has been published [44]. In this work, unique fragments of Bevacizumab (Avastin^{TM}) produced by the top-down in-source decay (ISD) of MALDI-generated protein ions were used as a means of detection of this monoclonal antibody across brain sections and demonstration of its effective localization within the tumor areas. A similar approach (reliance on the ISD fragment ions as a means of mAb detection) was used to map another growth factor-targeting antibody (Cetuximab) in colon cancer cell cultures [45]. MALDI MSI has also been used as a small-molecule imaging tool to detect and localize a cytotoxin (monomethyl auristatin E, MMAE) released from an ADC (anti-tissue factor monoclonal antibody conjugated with MMAE) in tumor tissues [46].

Despite the modest record of MALDI-MSI applications in the field of protein drug biodistribution studies, it is expected that the role of this technique will grow in the coming years. One promising approach to overcoming the obstacles associated with the high mass of protein therapeutics is to perform on-tissue enzymatic digestion prior to MSI analysis, as peptides are much easier to detect and identify compared to proteins. Although this technique is currently developed and used in the setting of biomarker discovery and clinical investigation [47, 48], we envision it can also be applied in a more targeted workflow for protein drug biodistribution studies. Further advances are also expected in the field of intact protein MALDI-MSI analysis [49, 50]; and the recent emergence of other ionization techniques suitable for the intact-protein MSI work [51–53] may further catalyze the progress in this field.

References

[1] Zhang, Z., Yan, Y., Wang, S. & Li, N. Development of a chromatography-free method for high-throughput MS-based bioanalysis of therapeutic mAbs. Bioanalysis 13, 725–735 (2021).

[2] Yan, Y., Xing, T., Wang, S. & Li, N. Versatile, sensitive, and robust native LC-MS platform for intact mass analysis of protein drugs. J. Am. Soc. Mass Spectrom. 31, 2171–2179 (2020).

[3] Hoofnagle, A.N. & Wener, M.H. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. J. Immunol. Methods 347, 3–11 (2009).

[4] Duncan, M.W., Yergey, A.L. & Gale, P.J. Quantifying proteins by mass spectrometry. Spectroscopy 30, 42–58 (2015).

[5] Fenselau, C. & Yao, X. $^{18}O_2$-Labeling in quantitative proteomic strategies: a status report. J. Proteome Res. 8, 2140–2143 (2009).

[6] Niles, R. et al. Acid-catalyzed oxygen-18 labeling of peptides. Anal. Chem. 81, 2804–2809 (2009).

[7] Kaltashov, I.A. et al. Advances and challenges in analytical characterization of biotechnology products: mass spectrometry-based approaches to study properties and behavior of protein therapeutics. Biotechnol. Adv. 30, 210–222 (2012).

[8] Wang, S., Bobst, C. & Kaltashov, I. A new liquid chromatography–mass spectrometry-based method to quantitate exogenous recombinant transferrin in cerebrospinal fluid: a potential approach for pharmacokinetic studies of transferrin-based therapeutics in the central nervous systems. Eur. J. Mass Spectrom. 21, 369–376 (2015).

[9] Mekhssian, K., Mess, J.N. & Garofolo, F. Application of high-resolution MS in the quantification of a therapeutic monoclonal antibody in human plasma. Bioanalysis 6, 1767–1779 (2014).

[10] Pan, K.T. et al. Mass spectrometry-based quantitative proteomics for dissecting multiplexed redox cysteine modifications in nitric oxide-protected cardiomyocyte under hypoxia. Antioxid. Redox Signal. 20, 1365–1381 (2014).

[11] Wang, S. & Kaltashov, I.A. A new strategy of using O18-labeled iodoacetic acid for mass spectrometry-based protein quantitation. J. Am. Soc. Mass Spectrom. 23, 1293–1297 (2012).

[12] Wang, S., Bobst, C.E. & Kaltashov, I.A. A new liquid chromatography-mass spectrometry-based method to quantitate exogenous recombinant transferrin in cerebrospinal fluid: a potential approach for pharmacokinetic studies of transferrin-based therapeutics in the central nervous systems. Eur. J. Mass Spectrom. (Chichester) 21, 369–376 (2015).

[13] An, B., Zhang, M. & Qu, J. Toward sensitive and accurate analysis of antibody biotherapeutics by liquid chromatography coupled with mass spectrometry. Drug Metab. Dispos. 42, 1858–1866 (2014).

[14] Blackburn, M. Advances in the quantitation of therapeutic insulin analogues by LC-MS/MS. Bioanalysis 5, 2933–2946 (2013).

[15] Jian, W., Kang, L., Burton, L. & Weng, N. A workflow for absolute quantitation of large therapeutic proteins in biological samples at intact level using LC-HRMS. Bioanalysis 8, 1679–1691 (2016).

[16] Vasicek, L.A., Zhu, X., Spellman, D.S. & Bateman, K.P. Direct quantitation of therapeutic antibodies for pharmacokinetic studies using immuno-purification and intact mass analysis. Bioanalysis 11, 203–213 (2019).

[17] Liu, H., Manuilov, A.V., Chumsae, C., Babineau, M.L. & Tarcsa, E. Quantitation of a recombinant monoclonal antibody in monkey serum by liquid chromatography-mass spectrometry. Anal. Biochem. 414, 147–153 (2011).

[18] Kellie, J.F., Kehler, J.R., Mencken, T.J., Snell, R.J. & Hottenstein, C.S. A whole-molecule immunocapture LC-MS approach for the in vivo quantitation of biotherapeutics. Bioanalysis 8, 2103–2114 (2016).
[19] Wong, D.L., Han, M., Barnaby, O. & Yang, Y. An Integrated workflow for sensitive intact protein quantitation of monoclonal antibodies from biological matrix. Appl. Note Agilent 5994-1249EN (2019).
[20] Kellie, J.F. et al. Intact mAb LC-MS for drug concentration from pre-clinical studies: bioanalytical method performance and in-life samples. Bioanalysis 12, 1389–1403 (2020).
[21] Nguyen, J.M., Smith, J., Rzewuski, S., Legido-Quigley, C. & Lauber, M.A. High sensitivity LC-MS profiling of antibody-drug conjugates with difluoroacetic acid ion pairing. mAbs 11, 1358–1366 (2019).
[22] Wang, C., Lee, C.S., Smith, R.D. & Tang, K. Ultrasensitive sample quantitation via selected reaction monitoring using CITP/CZE-ESI-triple quadrupole MS. Anal. Chem. 84, 10395–10403 (2012).
[23] Imami, K., Monton, M.R., Ishihama, Y. & Terabe, S. Simple on-line sample preconcentration technique for peptides based on dynamic pH junction in capillary electrophoresis-mass spectrometry. J. Chromatogr. A 1148, 250–255 (2007).
[24] VanAernum, Z.L. et al. Rapid online buffer exchange for screening of proteins, protein complexes and cell lysates by native mass spectrometry. Nat. Protoc. 15, 1132–1157 (2020).
[25] Iavarone, A.T., Jurchen, J.C. & Williams, E.R. Supercharged protein and peptide ions formed by electrospray ionization. Anal. Chem. 73, 1455–1460 (2001).
[26] Yang, Y., Niu, C., Bobst, C.E. & Kaltashov, I.A. Charge manipulation using solution and gas-phase chemistry to facilitate analysis of highly heterogeneous protein complexes in native mass spectrometry. Anal. Chem. 93, 3337–3342 (2021).
[27] Chen, Y., Mao, P. & Wang, D. Quantitation of intact proteins in human plasma using top-down parallel reaction monitoring-MS. Anal. Chem. 90, 10650–10653 (2018).
[28] Bults, P., Spanov, B., Olaleye, O., Van De Merbel, N.C. & Bischoff, R. Intact protein bioanalysis by liquid chromatography – High-resolution mass spectrometry. J. Chromatogr. B 1110-1111, 155–167 (2019).
[29] Kellie, J.F., Kehler, J.R., Karlinsey, M.Z. & Summerfield, S.G. Toward best practices in data processing and analysis for intact biotherapeutics by MS in quantitative bioanalysis. Bioanalysis 9, 1883–1893 (2017).
[30] Sanz-Medel, A., Montes-Bayon, M., Bettmer, J., Fernandez-Sanchez, M.L. & Encinar, J.R. ICP-MS for absolute quantification of proteins for heteroatom-tagged, targeted proteomics. TRAC Trends Anal. Chem. 40, 52–63 (2012).
[31] Cid-Barrio, L. et al. Advances in absolute protein quantification and quantitative protein mapping using ICP-MS. Trends Analyt. Chem. 104, 148–159 (2018).
[32] Zhao, H., Wang, S., Nguyen, S., Elci, S.G. & Kaltashov, I. Evaluation of nonferrous metals as potential in vivo tracers of transferrin-based therapeutics. J. Am. Soc. Mass Spectrom. 27, 211–219 (2016).
[33] Xu, S. & Kaltashov, I.A. Evaluation of gallium as a tracer of exogenous hemoglobin–haptoglobin complexes for targeted drug delivery applications. J. Am. Soc. Mass Spectrom. 27, 2025–2032 (2016).
[34] Ren, C., Bobst, C.E. & Kaltashov, I.A. Exploiting his-tags for absolute quantitation of exogenous recombinant proteins in biological matrices: ruthenium as a protein tracer. Anal. Chem. 91, 7189–7198 (2019).
[35] Sevick-Muraca, E.M. Translation of near-infrared fluorescence imaging technologies: emerging clinical applications. Annu. Rev. Med. 63, 217–231 (2012).

[36] Solon, E.G. Autoradiography: high-resolution molecular imaging in pharmaceutical discovery and development. Expert Opin. Drug Discov. 2, 503–514 (2007).
[37] Eisenhut, M. & Haberkorn, U. Molecular position of radiolabels and its impact on functional integrity of proteins. J. Nucl. Med. 47, 1400–1402 (2006).
[38] Boswell, C.A. et al. Effects of charge on antibody tissue distribution and pharmacokinetics. Bioconjug. Chem. 21, 2153–2163 (2010).
[39] Aljakna, A. et al. Multiplex quantitative imaging of human myocardial infarction by mass spectrometry-immunohistochemistry. Int. J. Legal Med. 132, 1675–1684 (2018).
[40] Sussulini, A., Becker, J.S. & Becker, J.S. Laser ablation ICP-MS: application in biomedical research. Mass Spectrom. Rev. 36, 47–57 (2017).
[41] Gessel, M.M., Norris, J.L. & Caprioli, R.M. MALDI imaging mass spectrometry: spatial molecular analysis to enable a new age of discovery. J. Proteomics. 107, 71–82 (2014).
[42] Schulz, S., Becker, M., Groseclose, M.R., Schadt, S. & Hopf, C. Advanced MALDI mass spectrometry imaging in pharmaceutical research and drug development. Curr. Opin. Biotechnol. 55, 51–59 (2019).
[43] Cobice, D.F. et al. Future technology insight: mass spectrometry imaging as a tool in drug research and development. Br. J. Pharmacol. 172, 3266–3283 (2015).
[44] Ait-Belkacem, R. et al. Monitoring therapeutic monoclonal antibodies in brain tumor. mAbs 6, 1385–1393 (2014).
[45] Liu, X. et al. MALDI-MSI of immunotherapy: mapping the EGFR-targeting antibody cetuximab in 3D Colon-Cancer Cell Cultures. Anal. Chem. 90, 14156–14164 (2018).
[46] Fujiwara, Y. et al. Imaging mass spectrometry for the precise design of antibody-drug conjugates. Sci. Rep. 6, 24954 (2016).
[47] Heijs, B., Tolner, E.A., Bovee, J.V., Van Den Maagdenberg, A.M. & McDonnell, L.A. Brain region-specific dynamics of on-tissue protein digestion using MALDI mass spectrometry imaging. J. Proteome Res. 14, 5348–5354 (2015).
[48] Andersson, M., Groseclose, M.R., Deutch, A.Y. & Caprioli, R.M. Imaging mass spectrometry of proteins and peptides: 3D volume reconstruction. Nat. Methods 5, 101–108 (2008).
[49] Burnum, K.E. et al. Imaging mass spectrometry reveals unique protein profiles during embryo implantation. Endocrinology 149, 3274–3278 (2008).
[50] Cazares, L.H. et al. Imaging mass spectrometry of a specific fragment of mitogen-activated protein kinase/extracellular signal-regulated kinase kinase kinase 2 discriminates cancer from uninvolved prostate tissue. Clin. Cancer Res. 15, 5541–5551 (2009).
[51] Griffiths, R.L., Creese, A.J., Race, A.M., Bunch, J. & Cooper, H.J. LESA FAIMS mass spectrometry for the spatial profiling of proteins from tissue. Anal. Chem. 88, 6758–6766 (2016).
[52] Griffiths, R.L., Randall, E.C., Race, A.M., Bunch, J. & Cooper, H.J. Raster-mode continuous-flow liquid microjunction mass spectrometry imaging of proteins in thin tissue sections. Anal. Chem. 89, 5683–5687 (2017).
[53] Hsu, C.C., Chou, P.T. & Zare, R.N. Imaging of proteins in tissue samples using nanospray desorption electrospray ionization mass spectrometry. Anal. Chem. 87, 11171–11175 (2015).

Chapter 8
Non-protein biopharmaceuticals and related macromolecular drugs

Proteins are not the only class of modern medicines that are produced using the means of biotechnology, nor are they the only type of macromolecules/biopolymers used as therapeutics. While the nucleic acid–based drugs are a relatively new addition to the armamentarium of modern medicine, the therapeutic use of certain polysaccharides (heparin and its derivatives) long predates the introduction of protein drugs into clinical practice. Even though neither nucleic acid–based medicines nor heparin are biopharmaceuticals in the strict sense of this word (the former are usually produced synthetically, while the latter is a natural product), there is a great deal of similarity between them and protein-based drugs vis-à-vis analytical characterization, which warrants their inclusion in this book. Furthermore, the issues related to heterogeneity of heparin extracted from animal sources continue to put significant pressure on the drug companies to develop biotechnological processes to produce this polysaccharide. In fact, it appears likely that in the next decade we will be witnessing a shift toward biotechnological production of heparin with better controlled "on-demand" structural features and biological properties, which will place this medicine in the realm of biopharmaceuticals (a transition which is not dissimilar to the fate of several early protein drugs, such as insulin and interferons). A switch from chemical synthesis to biological production for the majority of existing nucleic acid–based drugs seems unlikely, but biotechnology is already playing a major role in production of delivery vehicles for these highly labile biopolymers. Finally, nucleic acids are a major component of the newest class of biopharmaceuticals, gene delivery products. Analytical characterization of all of these classes of therapeutics will be considered in this chapter.

8.1 Nucleic acid–based therapeutics: MS characterization of small nucleic acids (antisense therapeutics and aptamers)

A promise of oligonucleotides as potentially powerful therapeutics has been recognized for a long time due to not only their ability to modulate gene expression, but also a variety of other unique properties that can be exploited for therapeutic interventions [2]. The first oligonucleotide approved as a drug was fomivirsen, a 21-nucleotide phosphorothioate oligonucleotide developed for treating cytomegalovirus (CMV) retinitis in immunocompromised patients [3]. It inhibits replication of human CMV by binding to complementary sequences on the messenger RNA transcribed from the transcriptional unit of the virus [4]. This mode of action is common to a class of oligonucleotide therapeutics that are referred to as antisense oligonucleotides (short strands of modified nucleotides that target RNA in a sequence-specific manner, inducing targeted protein knockdown or restoration) [5]. Another class of oligonucleotide therapeutics is aptamers, short single-stranded oligonucleotides that bind to specific target molecules (usually proteins) with high affinity and are used as analogs of antibodies [6]. Several other classes of therapeutic oligonucleotides are currently in development (Figure 8.1), but their detailed consideration is beyond the scope of this

https://doi.org/10.1515/9783110546187-008

chapter (interested readers are referred to a recent comprehensive review [1]). The molecular weight of therapeutic oligonucleotides is typically in the range of several kilodaltons, but can exceed 50 kDa, as is the case for PEGylated oligonucleotide products (such as pegaptanib [7]). Nevertheless, they are considered "small" when compared to the gene delivery products, which will be considered in Section 8.3.

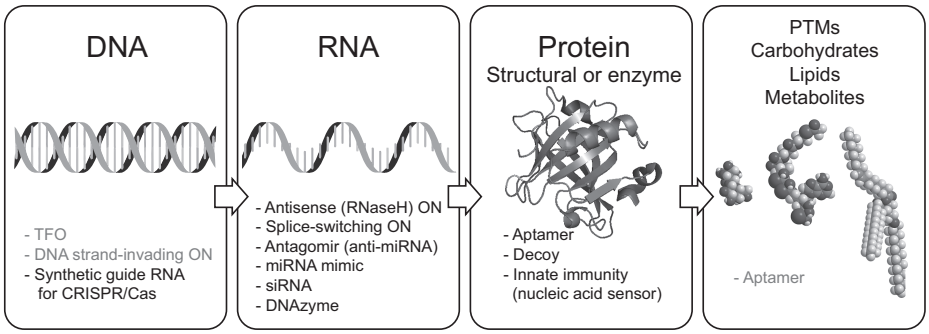


Figure 8.1: A schematic representation of various targeting strategies for oligonucleotides shown in the context of the extended central dogma that includes the effects of downstream enzymatic events. Strategies that have not yet reached the clinic are shown in gray. Reproduced from Smith and Zain [1].

As is the case for protein therapeutics, intact mass analysis of oligonucleotides typically provides the most straightforward way of establishing the consistency of the oligonucleotide mass and its presumed structure [8]. Isotopic resolution can be readily achieved in many cases for smaller oligonucleotides (below 20 kDa, which covers the majority of the currently approved products) using ESI as a means of ionization and high-resolution mass analyzers (Figure 8.2), enabling mass measurements with high precision and accuracy. Although MALDI MS can also be used for intact mass measurements of therapeutic oligonucleotides [9, 10], it produces mostly singly charged ions, which populate high-m/z regions of mass spectra, making it nearly impossible to achieve isotopic resolution for all but the smallest oligonucleotides (unless FT-ICR is used as the mass analyzer [11]). One distinct feature of oligonucleotides that clearly sets them apart from peptide and protein therapeutics vis-à-vis MS analysis is their polyanionic nature, which is frequently cited as a reason why the measurements should be carried out in the negative ion mode. In practice, abundant ionic signals of oligonucleotides can be readily generated in both positive and negative ion modes, although the extent of multiple charging in ESI MS tends to be lower for positive ions. Another important practical consequence of the polyanionic nature of oligonucleotides is their acting as "cation sponges" upon transition from the condensed phase to the vacuum, leading to extensive formation of Na^+, K^+, and other metal cation adducts. This behavior is evident in both positive and negative ion modes, although the extent of cationic adduct formation is obviously higher in the positive-ion ESI

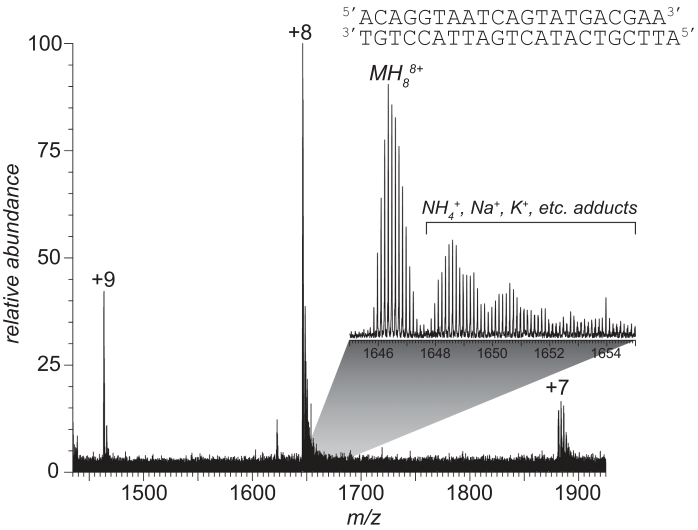


Figure 8.2: Intact mass measurement (ESI MS) of a short synthetic double-stranded DNA (10 µM in aqueous 150 mM ammonium acetate, pH 6.8). Acquired by G. Wang at UMass-Amherst.

mass spectra. Alkali metal adduct formation affects even relatively small oligonucleotides, but becomes particularly problematic in larger ones comprising tens or hundreds of bases. While the extent of adduct formation can be reduced in many instances by employing robust ion desolvation conditions in the ESI interface, the resulting oligonucleotide ions still contain dozens of Na and K atoms, as shown in Figure 8.3 using a transfer RNA as an example. Furthermore, excessive collisional activation can (and frequently does) lead to ion fragmentation in the gas phase, a process that obviously compromises the reliability of the mass measurements. The most reliable way of reducing/eliminating the signal of metal cation adducts in the mass spectra of oligonucleotides takes advantage of a range of extensive desalting procedures [8], which aim at replacing Na^+ and K^+ with NH_4^+. The latter readily dissociates from oligonucleotides in the gas phase in the form of a neutral ammonia molecule, giving rise to ions in which the negative charge excess is compensated almost exclusively by protons.

Direct-injection ESI MS analysis of synthetic oligonucleotides is suitable for quality control and detection of major impurities in both drug discovery and process development samples; however, combining ESI MS with online LC separation frequently enables more thorough characterization [8]. The majority of LC-MS analyses of oligonucleotides are carried out with reversed-phase LC, although the use of other chromatographic modalities, such as mixed-mode LC and HILIC, has also been reported. HILIC provides an advantage of allowing the LC-MS analyses to be carried out without ion pairing reagents [13], while the commonly used reversed-phase LC typically relies on strong organic bases (such as trimethylamine and dibutylamine)

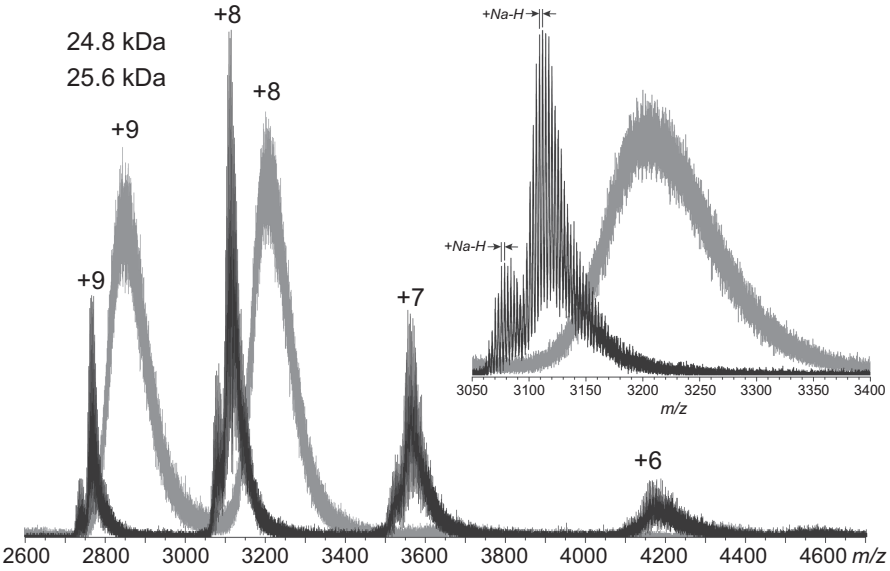


Figure 8.3: Intact mass measurements (ESI MS) of $tRNA^{His}$ with low collisional activation of ESI-generated ions (mild desolvation conditions, gray trace) and high collisional activation (robust desolvation, black trace). Data courtesy of Dr. Cedric E. Bobst (UMass-Amherst).

to achieve optimal chromatographic resolution via ion-pairing effect [14]. The presence of ion pairing reagents is frequently detrimental to the quality of MS data, mainly due to the ion signal suppression; however, in some cases their presence in the eluate may prove beneficial, as they may reduce the extent of alkali metal adduct formation [15].

Fragmentation of oligonucleotide ions can be readily triggered in the gas phase using a variety of ion activation techniques, the most common being CAD. The fragmentation mechanisms have been studied in detail for chemically unmodified oligonucleotides, and the fragment ion nomenclature was introduced nearly three decades ago [16]. However, the complexity of the fragmentation pathways (which frequently include base elimination in addition to the phosphodiester bond cleavage) limits the use of tandem MS methods to relatively short sequences. Furthermore, the majority of oligonucleotide therapeutics incorporate extensive structural modifications (Figure 8.4) that are introduced to enhance their stability and pharmacokinetic characteristics [12]. Structural modifications of oligonucleotides are known to exert a significance influence on the fragmentation pathways [17], but such correlations remain poorly characterized for the majority of modifications (indeed, systematic studies of the mechanisms of oligonucleotide ions fragmentation in the gas phase remain rare [18]). This certainly limits the practical utility of tandem mass spectrometry as a means of examining the structure of therapeutic oligonucleotides; however, if the fragmentation pathways are known for a specific class of oligonucleotide therapeutics, the

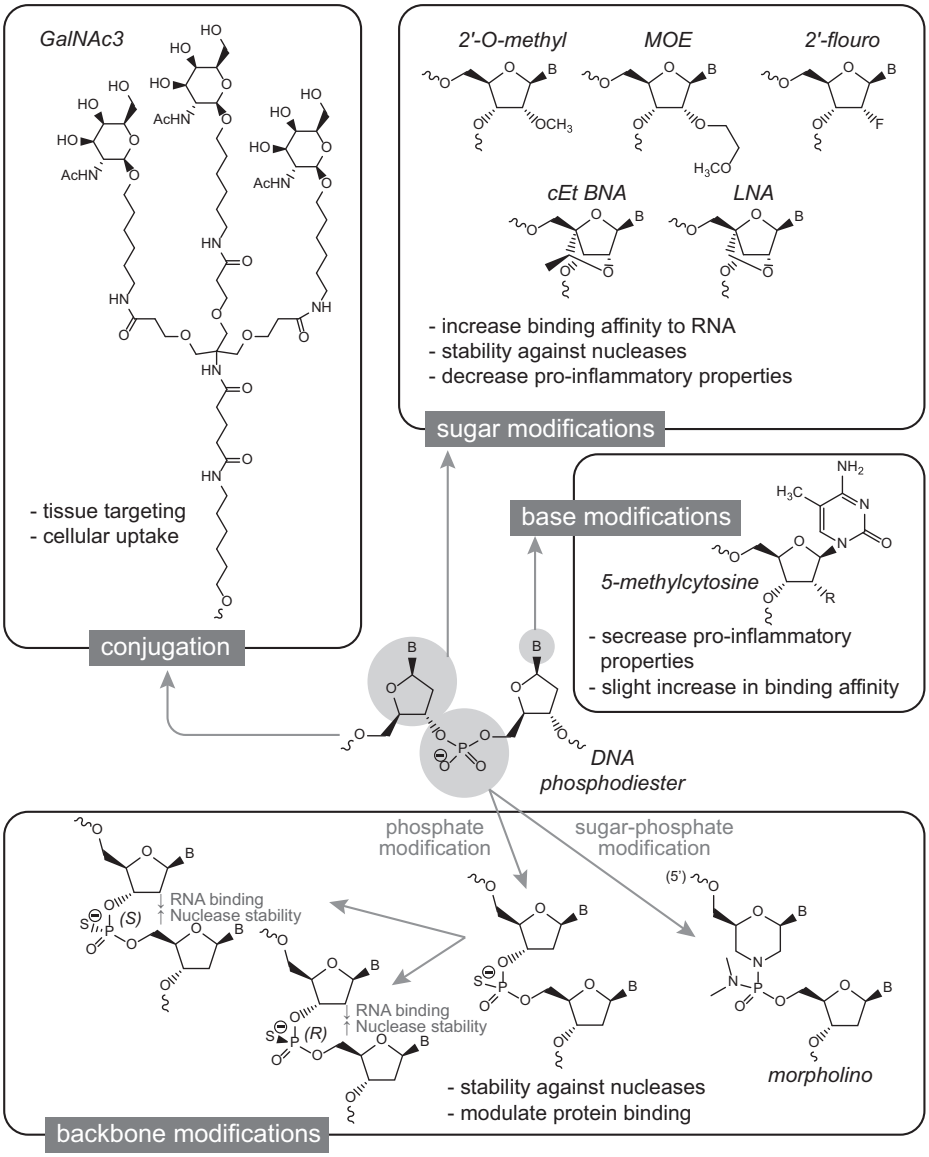


Figure 8.4: Common chemical modifications used in antisense drugs and their properties. A dinucleotide structure is shown, identifying positions where oligonucleotides are commonly modified. These positions include the phosphate backbone, the 2′ position on the sugar, the 5 position of pyrimidine bases, and addition of targeting ligands such as GalNAc sugars. Abbreviations: cEt BNA, (*S*)-constrained ethyl bicyclic nucleic acid; LNA, locked nucleic acid; MOE, 2′-*O*-methoxyethyl. Reproduced from Bennett [12].

detection selectivity and sensitivity of oligonucleotide ions can be greatly enhanced in LC-MS measurements by deploying the selected reaction monitoring mode of analysis [19]. Lastly, several software packages and online tools are available to facilitate structural analysis of both modified and unmodified oligonucleotides based on MS/MS data (reviewed by Sutton et al. [20]).

The development of MS characterization of oligonucleotides was inspired in large measure by the successes enjoyed by mass spectrometry in the fields of protein and polypeptide structural analysis. Therefore, many approaches that had been initially developed as a means of obtaining information on various aspects of the protein higher order structure (as discussed in detail in Chapter 4) had later been tested vis-à-vis their utility to probe conformation of oligonucleotides. Unfortunately, several MS-based tools that proved extremely useful in the field of protein analysis did not meet with much success when applied to oligonucleotides. For example, earlier work on the ionic charge state distributions of oligonucleotide ions yielded a paradoxical conclusion that the linear structures exhibit lower charge state distribution compared to the hairpin strands of the same composition [22]. As far as HDX MS, the majority of labile hydrogen atoms within oligonucleotides undergo H/D exchange reactions in solution on a timescale that is too short to allow analytical exploitation with MS as a means of probing conformation of these biopolymers. Other challenges include the inability to "freeze" any group of labile hydrogen atoms within oligonucleotides to afford their enzymatic processing prior to MS analysis (in a manner similar to quenching the exchange of the backbone amides in polypeptides, which enables their digestion with acidic proteases without a significant loss of the deuterium label, as discussed in Chapter 4). As a result, until recently the oligonucleotide HDX MS measurements were carried out almost exclusively in the gas phase, sometimes supplemented with ion fragmentation to evaluate spatial distribution of the exchanged hydrogen atoms within the oligonucleotide sequence [23] (inevitably raising questions related to the possibility of the gas-phase hydrogen scrambling affecting the outcome of the measurements).

Recently, there was a renewed interest in using the solution-phase H/D exchange reactions as a means of characterizing the higher order structure of therapeutic oligonucleotides. A systematic study by Largy and Gabelica [21] explored solution-phase HDX of small DNA quadruplexes and evaluated the influence of the ESI process (which was used to generate oligonucleotide ions for MS analysis) on the outcome of the measurements (Figure 8.5). It was concluded that the bases (but not the phosphates) were isotopically exchanged proportionally to the solution deuterium content without in-source back-exchange. Importantly, the measured exchange rates depended on the hydrogen bonding status of nucleobases, suggesting that this technique may in fact allow the nucleic acid conformation to be probed by HDX MS [21]. However, further studies in this field would be needed prior to HDX MS becoming a reliable tool for characterizing the higher order structure of therapeutic oligonucleotides.

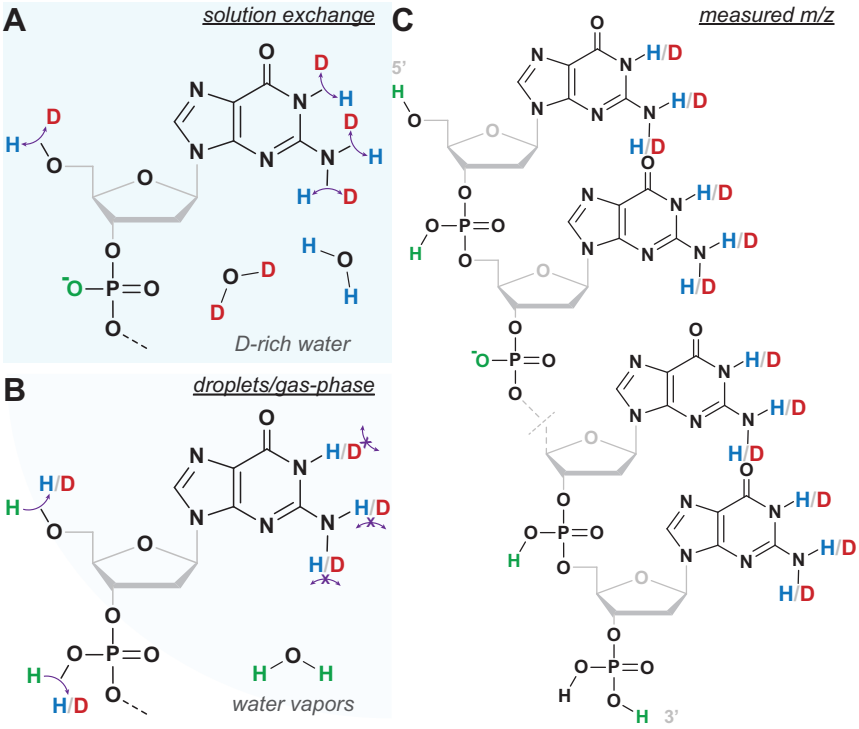


Figure 8.5: Exchangeable sites in native HDX MS of trimethylammonium acetate-solutions of DNA oligonucleotides. (A) Unpaired nucleobases exchange to reflect the isotopic solvent composition, whereas phosphates remain ionized until (B) the electrospray process during which some (depending on the charge) are neutralized by either H or D. Ultimately they are back-exchanged by the protons from the ambient H_2O vapors in the gas phase and/or droplets. Nucleobases do not back-exchange, leading to (C) the final analyte in which the nucleobases reflect the solution exchange and the phosphate are either ionized or protonated. Reproduced with permission from Largy and Gabelica [21]. Copyright 2020 American Chemical Society.

Unlike the charge state distribution analysis and HDX MS, native MS of oligonucleotide non-covalent complexes has been used successfully in a variety applications ranging from tasks as mundane as observing short DNA duplexes (see Figure 8.2) to detection and characterization of macromolecular associations formed by oligonucleotide drugs and their therapeutic targets. The stability of DNA duplexes under ESI conditions that was initially noticed nearly three decades ago [25] is usually ascribed to the nature of non-covalent interactions (hydrogen bonds) that effectively maintain the integrity of such complexes both in solution and in the solvent-free environment. Hydrogen bonds and ionic interactions also maintain the integrity of oligonucleotide complexes with proteins, allowing ESI MS to be used as a means of detecting and probing the composition of such assemblies. An example of using native MS for this

task is shown in Figure 8.6, where association of a therapeutic oligonucleotide (thrombin-binding aptamer) with its target (thrombin) is readily observed [24]. The remarkable stability of such non-covalent associations in the gas phase allows fast-heating ion fragmentation techniques to be used as a means of localizing the oligonucleotide-binding interface on the protein surface. This is done by comparing the top-down fragmentation patterns of the protein backbone generated by ECD of ions representing the free protein and its complex with the oligonucleotide [24]. The use of ECD or another fast-heating method of ion activation (ETD or UVPD) is essential, as slow-heating methods (such as CAD or IRMPD) result in a gradual increase of the ionic internal energy, which results in the ligand dissociation from the protein prior to fission of the covalent bonds.

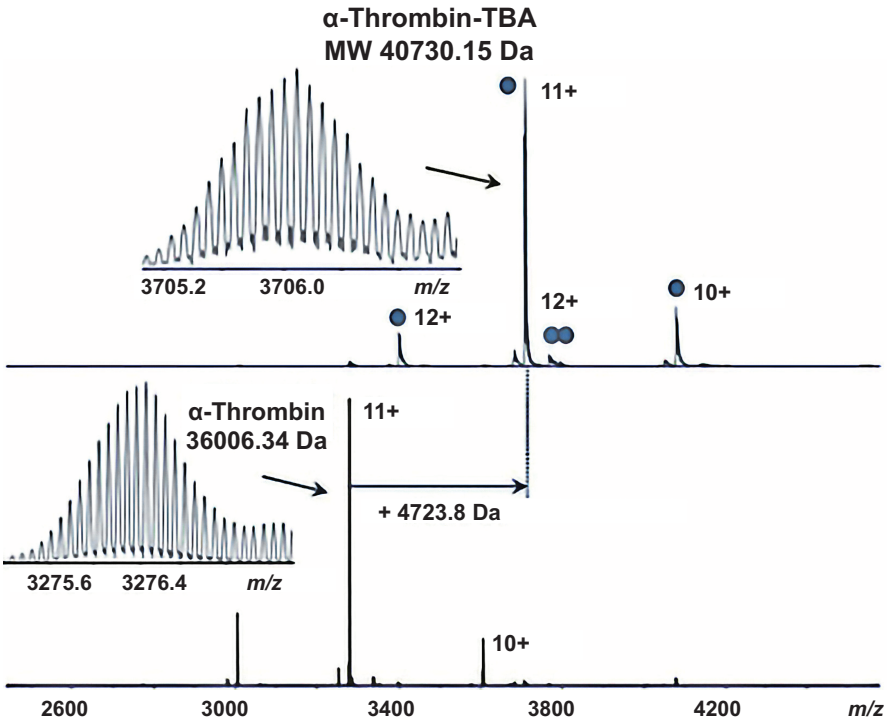


Figure 8.6: A high-resolution native mass spectrum of human α-thrombin/thrombin-binding aptamer (TBA) complex acquired in a 200 mM ammonium acetate solution. The mass difference shows the association of a single TBA molecule with the protein. The insets show accurate mass measurements for the 11+ ions. The blue circles indicate the number of TBA molecules in each complex. Adapted with permission from Zhang et al. [24].

8.2 Macromolecular natural products: heparin and related medicines

Heparin is arguably the oldest macromolecular drug (discovered in the late 1910s [26] and widely accepted in clinical practice as an effective anticoagulant within two decades [27, 28], it still remains one of the most frequently used blood clotting prevention agents). Heparin possesses a range of other biological activities that are extensively investigated with the hope of extending the therapeutic uses of this biopolymer beyond coagulopathy [29–31]. These efforts have been intensified as a result of the COVID-19 pandemic, following the realization that the antithrombotic properties of heparin work synergistically with its anti-inflammatory and anti-viral activities, bringing it to the forefront of the fight against the novel coronavirus [32, 33]. Heparin is a highly sulfated linear polysaccharide (Figure 8.7) that belongs to the family of glycosaminoglycans (GAGs). Although all medicinal heparin is currently harvested from animals, there is a concerted push toward replacing this animal-sourced natural product with the bioengineered heparin [34, 35], which is primarily driven by two factors. First, the need to extract the raw material for heparin mass-scale production from animals has resulted in two major crises in the past three decades, one associated with the emergence and proliferation of bovine spongiform encephalopathy [36], which culminated in a withdrawal of all bovine heparin from the market in 1990s, and another one associated with the catastrophic instance of a massive contamination of porcine-derived medicinal heparin with oversulfated chondroitin sulfate in 2000s [37, 38]. The second significant stimulus for the development of biotechnological production of heparin relates to the enormous degree of structural heterogeneity exhibited by this biopolymer obtained from animal sources (both the tremendous diversity of the O- and N-sulfation and N-acetylation patterns, and the chain length polydispersity, see Figure 8.7A). While this unmatched degree of structural diversity endows this biopolymer with the ability to interact with an impressive range of proteins [39, 40], many of which are high-value therapeutic targets, it also means that a randomly selected heparin molecule is unlikely to possess the structure optimized for interactions with a specific target. For example, the canonical antithrombin-binding segment of heparin comprises a pentasaccharide sequence [41] (Figure 8.7C) which appears to be relatively rare and is found in only about one-third of the heparin chains in commercial preparations [42]. The ability to bioengineer and produce heparin [43–46] that incorporates high-affinity "recognition elements" for a specific therapeutic target and lacks structural segments that can trigger interactions causing undesirable physiological effects would undoubtedly allow both efficacy and safety of this century-old drug to be dramatically enhanced. In our opinion, emergence of bioengineered heparin and/or related GAGs approved for clinical use is only a matter of time, which warrants consideration of MS-based methods of analysis of heparin and heparin-like biopolymers in this chapter.

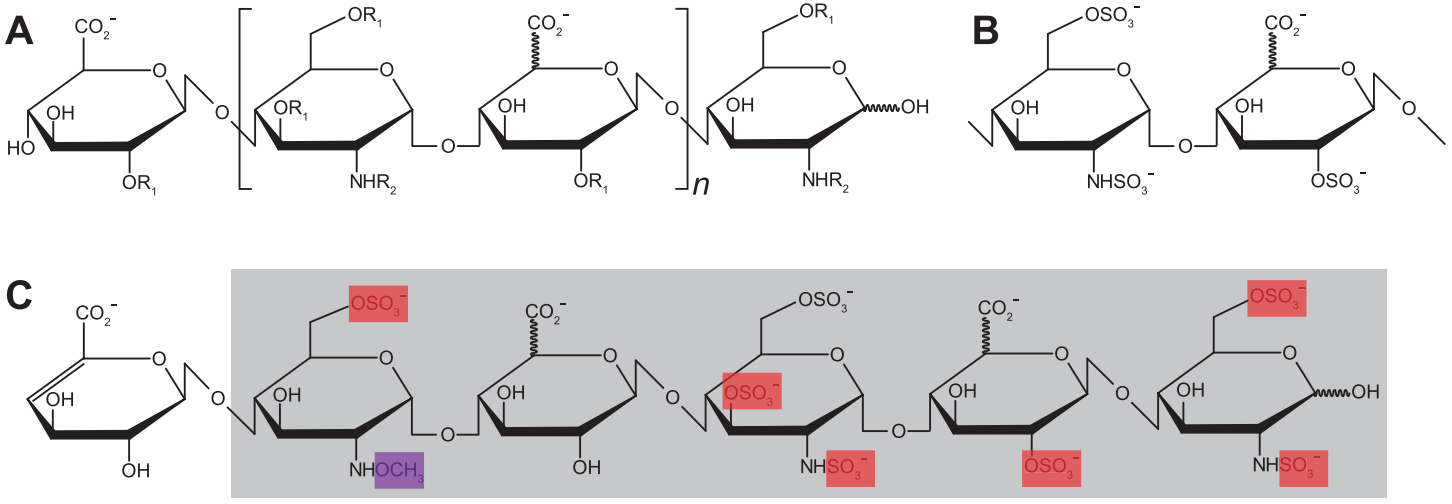


Figure 8.7: (A) The template of heparin and heparan sulfate chemical structure (R_1 = H or SO_3, and R_2 = H, OCH_3 or SO_3). (B) The most common disaccharide building block of heparin. (C) The canonical antithrombin-binding hexasaccharide element of heparin (ΔHex-GlcNAc,6S-GlcA(IdoA)-GlcNS, 3S,6S-IdoA2S-GlcNS,6S). The sulfate and acetylate groups of the hexamer highlighted in red and blue, respectively, are essential for antithrombin binding.

The physical size of heparin is relatively modest in comparison with the majority of protein therapeutics (the average molecular weight is only 15 kDa for unfractionated heparin, and does not exceed 8 kDa for low-molecular-weight heparin). Nevertheless, intact mass measurements that have become routine analyses for protein therapeutics as large as mAbs (see Chapter 3 for more details) are not feasible for heparin. The reason for this is the enormous structural heterogeneity exhibited by heparin, with the structure of each chain being the result of a concerted action of a number of enzymes as opposed to an imprint off a genetic template. While the chain length is the single most significant source of heparin polydispersity, variations in the extent of sulfation and the number of incorporated N-acetyl groups give rise to a significant number of distinct species with unique masses even in the case of relatively short fixed-length heparin oligomers (Figure 8.8). While ESI MS alone can readily detect all these species for each fixed-length heparin oligomer (at least up to eicosamers), a mixture of even a small number of such oligomers is going to result in a very crowded mass spectrum populated with overlapping ionic signals from different heparin species. Using MALDI instead of ESI might seem advantageous, as it would mostly generate singly charged ions, spreading the signal of heparin oligomers of different lengths across the entire m/z scale (as opposed to concentrating it within a relatively narrow m/z range, as is the case for multiply charged ions) and thereby avoiding signal overlap. However, the lability of the sulfate groups results in their facile cleavage in MALDI, giving rise to abundant fragment ion signal [48]. The loss of sulfate groups in the gas phase would obviously invalidate the results of the intact mass measurements, and even though a judicious use of matrices may minimize the occurrence of this phenomenon, there is still no consensus as to what conditions provide optimal results. For example, a crystalline matrix norharmane and an ionic liquid 1-methylimidazolium α-cyano-4-hydroxycinnamate had been reported to cause minimal fragmentation [49], while another study concluded that using norharmane as a matrix results in significant sulfate loss, and recommended the use of dihydroxy benzoic acid [50]. It appears that any protocol to be employed for MALDI MS analysis of heparin and related products should be tested using a homogeneous heparin mimetic fondaparinux, a synthetic pentasaccharide incorporating eight sulfate groups, in order to determine conditions that eliminate the specter of ion fragmentation in the gas phase. As a side note, it should also be mentioned that the sulfate loss may also occur when ions are produced by ESI followed by excessively aggressive desolvation using collisional activation in the ESI interface region. In such situations acquisition of a fondaparinux mass spectrum to prove that robust ion desolvation does not trigger sulfate shedding in the gas phase also seems prudent.

An alternative option for intact mass analysis of heparin samples involves the use of LC-MS [52]. Both SEC [47, 53–55] and reversed-phase LC [51, 56–58] have been particularly popular choices vis-à-vis front-end separation. The use of SEC-MS for the analysis of heparin products is relatively straightforward as long as the mobile

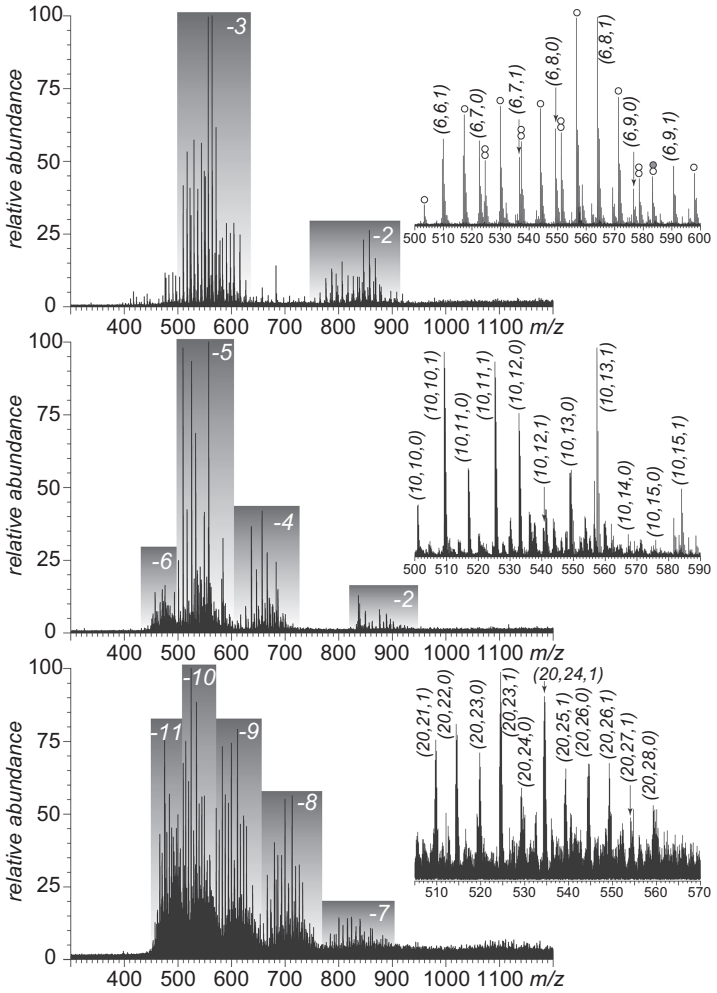


Figure 8.8: Intact mass measurements of fixed-length heparin oligomers (hexamers, top; decamers, middle; and eicosamers, bottom) carried out with ESI FT ICR MS. Shaded boxes identify ionic species with a specific charge state, as shown on the graph. The insets show zoomed views of m/z regions corresponding to the most abundant charge state in each mass spectrum. Henriksen's nomenclature [47] is used to label all detected heparin oligomers (each heparin species is identified based on a number of saccharide units in the chain, a number of sulfate groups and a number of N-acetyl groups, which are shown as a triad for each species). Data courtesy of Yang Yang, a research assistant at UMass-Amherst.

phase does not include non-volatile salts and buffers (the restrictions that have been already discussed in Chapter 4), although in some instances an online removal of excess ammonium from the eluate is required to eliminate ammonium adducts of heparin anions and improve the ionic signal in MS [53]. Successful uses of reversed-phase

chromatography in LC-MS analyses of heparin and related glycosaminoglycans almost always utilize ion pairing (e.g., tripropyl ammonium acetate). Other (less frequently used) options include anion exchange LC [59], HILIC [60], and capillary electrophoresis [61] with online MS detection.

The appearance of total ion chromatograms is typically very convoluted, reflecting the presence of dozens (and possibly hundreds) of different molecular species with unique masses (an example is shown in Figure 8.9). While the majority of these species can be readily identified based on mass measurements, any conclusions vis-à-vis their relative abundance should be taken *cum grano salis* due to the very likely influence of the heparin oligomer size, the number of sulfate groups, mobile phase composition (if a gradient elution is employed), as well as the presence of other species in the eluate that may be competing for charge during the ESI process.

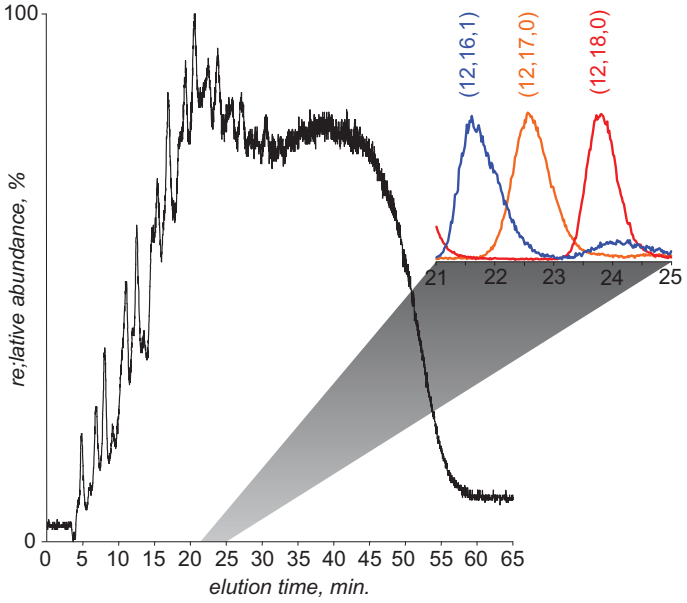


Figure 8.9: Ion-pairing reversed-phase LC-MS chromatogram (TIC) of a medium molecular weight heparin (prepared by limited digestion of intact unfractionated heparin). Mobile phases were H_2O (A) and 1:9 H_2O/CH_3OH (B); both phases contained 25 mM $N(C_3H_7)_3$ and 30 mM CH_3CO_2H. The gradient was 40–50% B in 15 min followed by 50–65% (B) in 50 min. The inset shows extracted ion chromatograms corresponding to ionic species (charge state –6) representing three different dodecameric species. Adapted from Henriksen et al. [51] with permission from Elsevier.

Intact mass measurements, if carried out at sufficient resolution, provide information on the chain length (the number of saccharide units), as well as the extent of sulfation and the number of N-acetyl groups within each molecule. In most cases these three numbers are sufficient for identification of heparin oligomers with unique

masses, providing the basis for Henriksen's nomenclature [47], which identifies all species with a unique mass using a triad of integer numbers (a number of saccharide units in the chain, a number of sulfate groups, and a number of N-acetyl groups, as shown in Figure 8.8). Each of these species, however, comprises a large number of isomers, which may display similar physico-chemical properties, but differ in their biological activities. Precise localization of the sulfate and N-acetyl groups within a heparin chain cannot be determined simply by "sequencing" heparin (in a fashion similar to protein sequencing) due to the enormous degree of inter-chain structural variability. Instead, the initial step in structure evaluation usually involves heparin chain lysis with heparinase (an enzyme that cleaves the glycosidic linkage between hexosamines and uronic acids) down to the di- and tetra-saccharide level. The LC-MS analysis of the resulting small heparin fragments allows the overall sulfation level to be evaluated (e.g., see Figure 8.10, which shows that the majority of the disaccharide fragments produced by lysis of the eicosa-saccharide heparin chains are tri-sulfated disaccharides, although disaccharides incorporating fewer sulfate groups are also present and can be quantified).

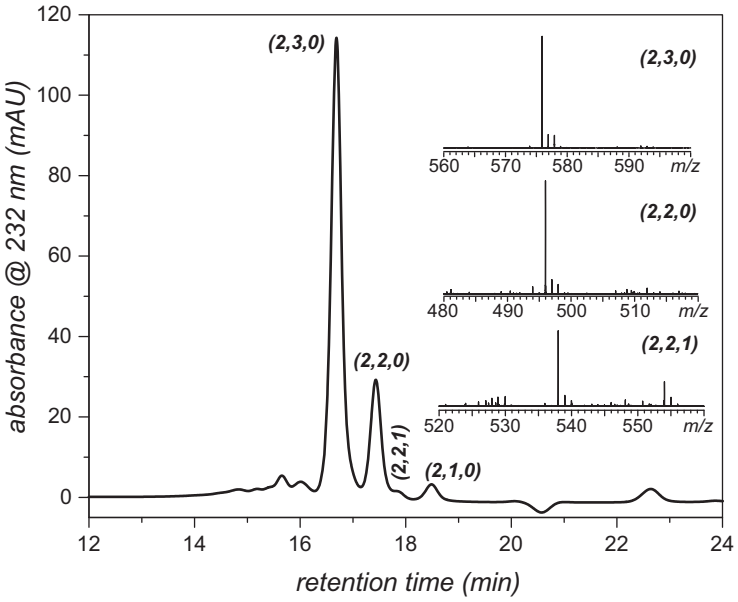


Figure 8.10: SEC-MS analysis of completely digested heparin chains dp20. In a 16-h digestion, dp20 was mostly digested to disaccharides containing tri-sulfated (16.5 min, 77.2%), di-sulfated (non-acetylated at 17.5 min and acetylated at 17.8 min, 29.7%) and mono-sulfated (18.5 min, 2.1%) species. Relative abundance (%) was calculated based on the integrated chromatographic peak area. Small amount of tetrasaccharide (15.5–16.2 min) were resistant to enzymatic digestion. The inset shows mass spectra of collected fractions of (2,3,0), (2,2,0), and (2,2,1). Adapted with permission from Niu et al. [55]. Copyright 2020 American Chemical Society.

If the heparin oligomer samples are extensively fractionated (usually by a combination of size exclusion and strong anion exchange LC), MS/MS analysis of such relatively homogeneous samples can be carried out aiming at localization of the sulfate and N-acetyl groups within the polysaccharide. One significant complication that is frequently encountered in such measurements is the lability of sulfate groups in the gas phase following collisional activation [62]. The extent of sulfate shedding can be minimized by selecting precursor ions in which all acidic hydrogen atoms are substituted with sodium atoms (which is achieved by adding mM-level NaOH to the heparin oligomer solution). Such Na^+ adducts exhibit enhanced stability vis-à-vis sulfate groups upon collisional activation, while producing abundant structurally informative fragment ions resulting from both glycosidic and cross-ring cleavages [63]. Sulfate shedding can also be minimized by using electron-based ion fragmentation techniques, such as electron detachment dissociation (EDD) [64] and negative ion electron transfer dissociation (NETD) [65]; the recently introduced ultraviolet photodissociation (UVPD) also shows a promise in this regard [66]. The presence of cross-ring cleavages in the tandem mass spectra of heparin oligomers and related glycosaminoglycans allows in many cases precise localization of the sulfate groups to be achieved, but at the same time it gives rise to convoluted mass spectra which may be difficult to analyze and interpret manually. Spectral interpretation can be greatly assisted by specialized software packages, such as GAGfinder [67]. Despite all these advances, MS/MS analysis of glycosaminoglycans is currently limited to relatively short oligomers, even when supplemented with online LC separation [68] or ion mobility [69] to resolve isobaric species prior to the precursor ion selection.

Unlike proteins and nucleic acids, heparin and related glycosaminoglycans do not appear to have stable higher order structures, unless associated with proteins. Therefore, native MS analysis of these biopolymers is usually focused on characterizing their non-covalent complexes. While detection and characterization of such complexes is relatively straightforward in the case of short heparin oligomers [71–75], application of this technique to heparin chains of even a moderate size (let alone intact heparin) is challenging due to their extreme structural heterogeneity. As has been already discussed in Chapter 4, polydispersity gives rise to broad ionic peaks in ESI mass spectra, which in the case of intact heparin becomes a continuum signal with very few discernable features (e.g., see the black trace in Figure 8.11 representing a mass spectrum of a heparin/antithrombin mixture). Nevertheless, meaningful interpretation of such mass spectra can be made possible by supplementing native MS measurements with limited charge reduction, a technique that had been already discussed in detail in Chapter 4. This is illustrated in Figure 8.11, where charge reduction of ionic populations isolated from different m/z regions of a nearly continuum ion signal in the mass spectrum of antithrombin/unfractionated heparin mixture reveals the presence of heparin/protein complexes with different binding stoichiometries.

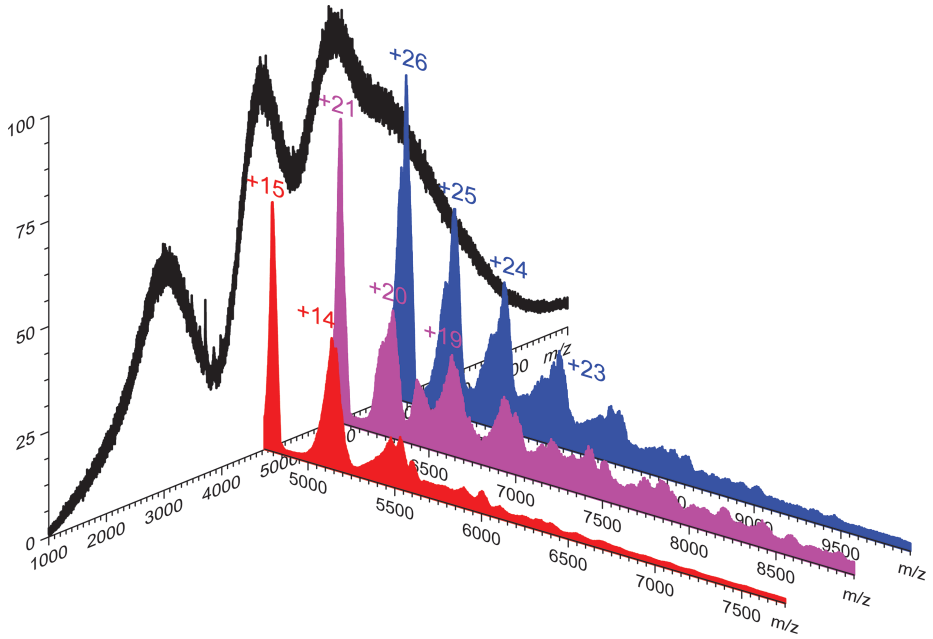


Figure 8.11: Native ESI mass spectrum of the antithrombin/ heparin mixture (black trace) and three representative mass spectra of ions produced via limited charge reduction of precursor ions whose *m*/*z* values fall within 4,450–4,550 (red), 5,950–6,050 (purple), and 6,850–7,050 (blue) *m*/*z* regions revealing the presence of 1:1, 1:2, and 1:3 heparin/protein complexes, respectively. Reproduced with permission from Zhao et al. [70]. Copyright 2016 American Chemical Society.

8.3 MS in the analytical support of gene therapies

Gene therapy products are the latest-generation biopharmaceuticals, which are used to modify or introduce genes into a patient's body aiming at preventing or treating a disease. When performed in vivo, the therapeutic genes are delivered to cells inside the patient's body using viral vectors (with exosome-based vectors also being actively pursued, but not yet approved). Ex vivo gene therapy entails inserting therapeutic gene into cells outside of the body followed by their (re)introduction into the body, which makes it a form of the so-called cell therapy (in which new or modified cells are introduced into a patient's body aiming at treating a disease or repairing damaged tissues, which is the subject of the following section).

The first gene therapies were not approved in the USA until 2017, but the current list of approved biopharmaceuticals of this type already includes over a dozen entries,[1]

1 https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/approved-cellular-and-gene-therapy-products

with several hundred more currently in development [77–79]. According to the FDA's Office of Tissues and Advanced Therapies, product testing and characterization is an integral part of successful development of quality cell and gene therapy products.[2] However, the complexity of these products requires new analytical approaches, as a large number of tools developed for characterization of protein therapeutics no longer meet the new demands. MS-based tools have been used very actively to address various aspects of analytical characterization of the gene therapy products, and in this section we will highlight both traditional MS-based approaches and novel methods that have been developed recently to address specific challenges posed by this type of biopharmaceuticals. The scope of our discussion will be mostly limited to the gene therapy products that use an adeno-associated virus (AAV) [76], the most common vector at present.

AAV is a nonpathogenic parvovirus composed of a 4.7 kb single-stranded DNA genome within a nonenveloped, icosahedral capsid. In a recombinant version of AAV, a gene of interest is inserted between the inverted terminal repeats (which function as the viral origin of replication and the packaging signal), replacing the elements encoding nonstructural proteins that have roles in viral replication, transcriptional regulation, genomic integration, and virion assembly, as well as three structural proteins (VP1, VP2, and VP3) that make up the 60-meic viral capsid (Figure 8.12). There are several naturally occurring serotypes of AAV (commonly referred to as AAV1 through AAV12) and over a hundred variants, which exhibit notable differences in amino acid sequence of the three capsid proteins (mostly within the so-called hypervariable regions, see Figure 8.13). In addition to the amino acid sequence variations, the capsid proteins are also affected by a range of PTMs (vide infra), and many of these structural features are important determinants of the virus tissue tropism, making the control of the VP1-VP3 structure an important task.

Besides the identity test of the AAV serotype that is recommended for the AAV vector product release (especially those manufactured in multi-product/multi-serotype facilities), complete characterization of the constituent viral capsid proteins of AAV vectors (both amino acid sequences and PTMs) is highly recommended to ensure AAV product quality and consistency [80]. This task can be readily accomplished by carrying out an LC-MS analysis of the intact viral capsid proteins (intact mass analysis, see Chapter 3 for more details) and an LC-MS/MS analysis of their enzymatic digests. The latter allows complete sequence confirmation of all capsid proteins to be achieved, as well as PTMs to be detected and identified [80], such as the N-terminal acetylation (as illustrated in Figure 8.14 for N-terminal proteolytic fragments derived from VP1-VP3 of AAV2). Peptide mapping can also be used to profile other PTMs within the capsid proteins, such as N-glycosylation. [81]. In addition to the N-terminal acetylation and

2 https://www.fda.gov/vaccines-blood-biologics/news-events-biologics/advanced-topics-successful-development-quality-cell-and-gene-therapy-products

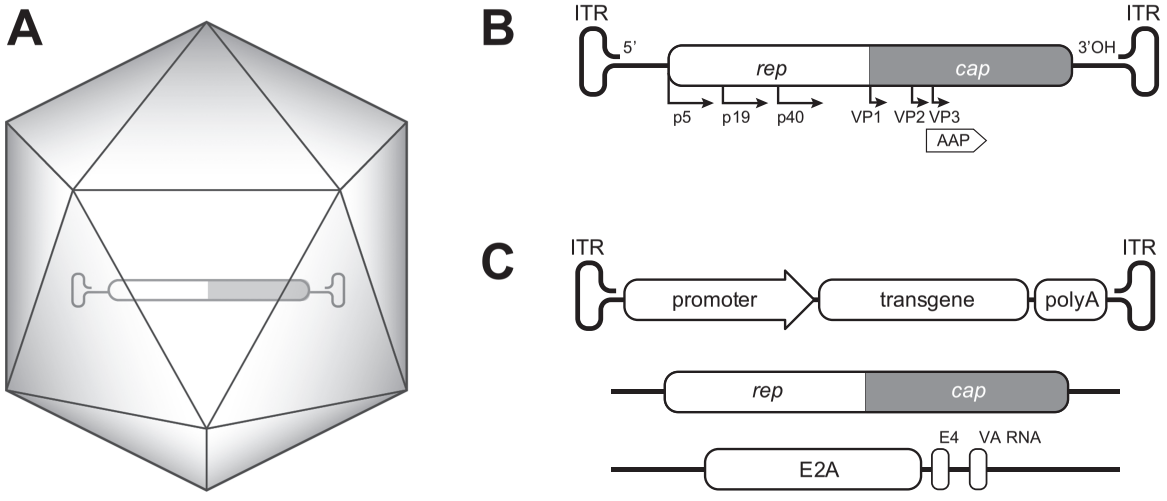


Figure 8.12: AAV and its vector. (A) Schematic of AAV. A single-stranded genome is surrounded by the protein capsid. (B) The DNA genome of AAV. The 4.7 kb genome encoding the *rep*, *cap*, and *aap* open reading frames is flanked by ITRs. (C) Genetic components of AAV vectors. The therapeutic transgene, along with associated promoter and polyadenylation sequences, is inserted between the viral ITRs. AAV *rep* and *cap*, and adenovirus E2A, E4, and VA RNA sequences are supplied in *trans* to produce the vector. Abbreviations: AAP, assembly-activating protein; AAV, adeno-associated virus; ITR, inverted terminal repeat; PolyA, polyadenylation sequence; VA, viral associated. Adapted with permission from Kotterman et al. [76].

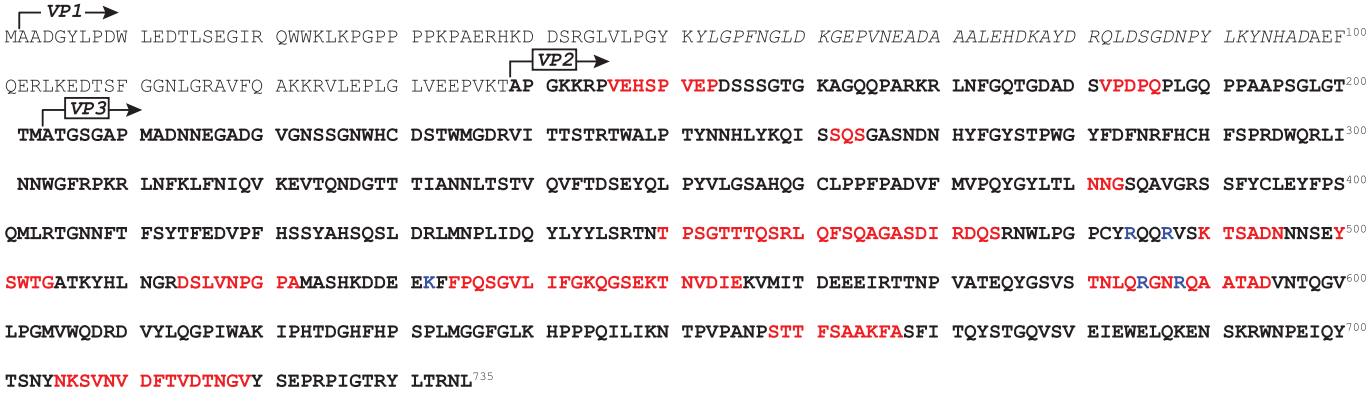


Figure 8.13: Amino acid sequence of the AAV2 coat proteins (UniProt P03135). The 12 hypervariable regions (HVRs) are highlighted in red.

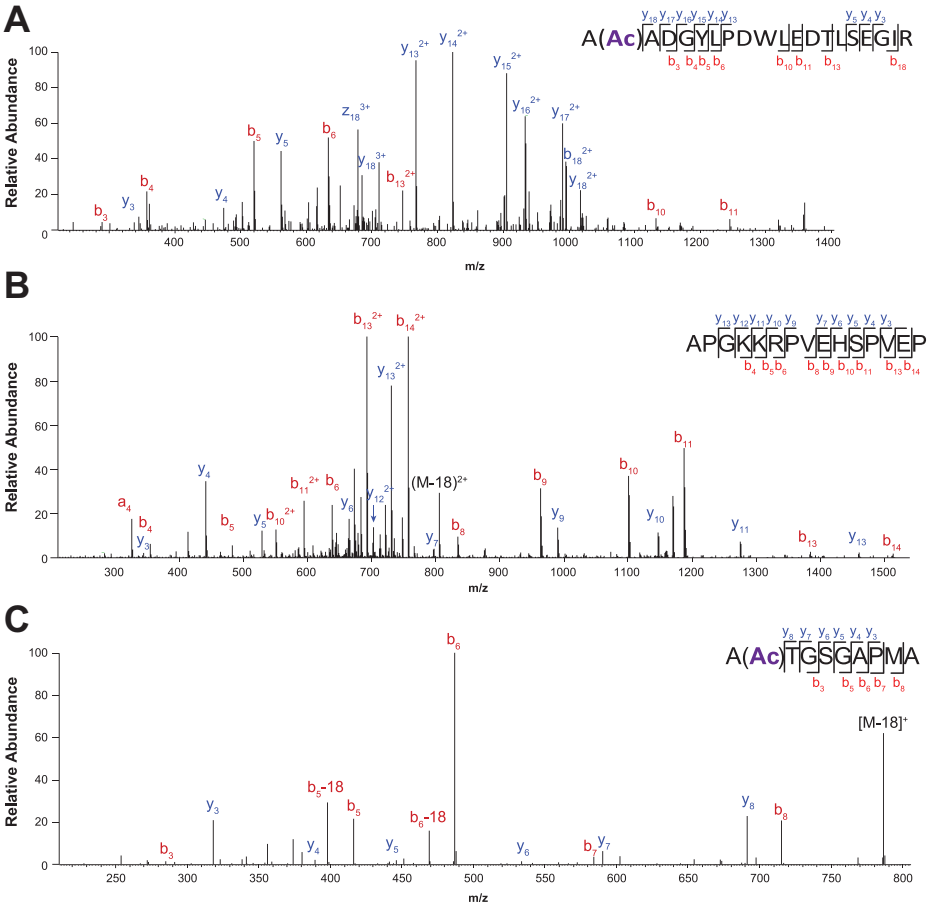


Figure 8.14: Tandem mass spectra of the N-terminal peptides of AAV2 VPs. (A) N-terminal tryptic peptide of VP1; (B) N-terminal Asp-N peptide of VP2; and (C) N-terminal Asp-N peptide of VP3. Adapted with permission from Jin et al. [80].

N-glycosylation, a range of other PTMs can affect the structure of the capsid proteins, with 1 recent study reporting over 50 modifications in AAV2–AAVrh10 capsids, with at least 4 AAV serotypes exhibiting over 7 PTMs in their capsid proteins [82].

Above and beyond the qualitative analysis of the capsid proteins, quantitation of their stoichiometric levels is an important task in the viral particle characterization. The capsid proteins are typically present in a viral particle at a ca. 1:1:10 (VP1/VP2/VP3) ratio. While VP3 is the most abundant protein, the viral particle loses its functionality when it is composed exclusively of VP3 (note that a number of functionally important elements, such as the phospholipase 2 domain – residues 52–97 – which is critical for the virus release from the endosome, are not present in VP3, see Figure 8.13). A relatively straightforward way to obtain relative abundance of VP1, VP2,

and VP3 relies on ^{18}O labeling [83], a technique that has been already considered in detail in Chapter 7. One unique challenge that arises when ^{18}O labeling is applied to quantitation of the capsid proteins relates to the fact that regardless of the specific protease employed for digestion of these proteins, there is always a single peptide (the N-terminal proteolytic fragment) that is unique to each capsid protein. While this is not a significant problem as far as stoichiometric measurements, it unfortunately rules out a possibility of quantitating the protein isoforms in which PTMs fall outside of the N-terminal proteolytic fragments. An elegant approach to addressing this problem that has been reported recently takes advantage of intact mass analysis using LC with online detection by MS (for identification/assignment of the capsid proteins and their isoforms) and by optical spectroscopy methods (for the straightforward quantitation). For this type of analysis, hydrophilic interaction chromatography (HILIC) offers a significant advantage over the traditional reversed-phase LC, as it allows improved separation of VPs from a variety of AAV serotypes to be achieved using a generic method prior to online MS detection [84]. HILIC is also ideally suited for capsid heterogeneity characterization due to its remarkable sensitivity vis-à-vis capsid protein variants resulting from different PTMs, as well as protein backbone clippings. One unique challenge in using intact mass LC-MS analyses of capsid proteins is the low protein concentrations in AAV samples; this can be circumvented by increasing the HILIC column loading. Other issues that have to be dealt with in these analyses are common in HILIC-MS characterization of proteins, for example, ion signal suppression by trifluoroacetic acid (TFA); one possible way to address this problem is incorporation of a desolvation gas modification device in the experimental workflow [84]. Once all these issues are resolved, HILIC-MS method can be applied to characterize a wide variety of AAV serotype samples at low concentrations without any sample treatment to achieve unambiguous serotype identification, PTM characterization (as illustrated in Figure 8.15), and stoichiometry assessment [84]. Importantly, quantitation can be carried out using online detection methods that are complementary to MS (e.g., highly sensitive fluorescence detection [84]).

A typical payload for an AAV-based gene therapy is a single-stranded DNA consisting of ca. 3,500 nucleotides (~ 1 MDa). It is quite heterogeneous (can be as large as 4,300 nucleotides, and the genome size affects the potency). While the biopolymers of such size and heterogeneity presently remain beyond the reach of traditional MS-based methods, exciting opportunities have arisen due to a recent revival of a technique introduced a quarter-century ago [85]. This technique combines MS measurements with ionic charge measurements, thereby enabling extraction of mass distributions for large and heterogeneous ions generated by ESI [85], and provides an opportunity to assess both the average size and the extent of heterogeneity of large nucleic acids [86, 87] (an example is shown in Figure 8.16).

The ability of the AAV capsid proteins to assemble into a capsid hinges upon their correct folding; at the same time there is a flexible domain within VP1, which plays a role of a pH-sensing element during the endosomal uptake of the viral particle.

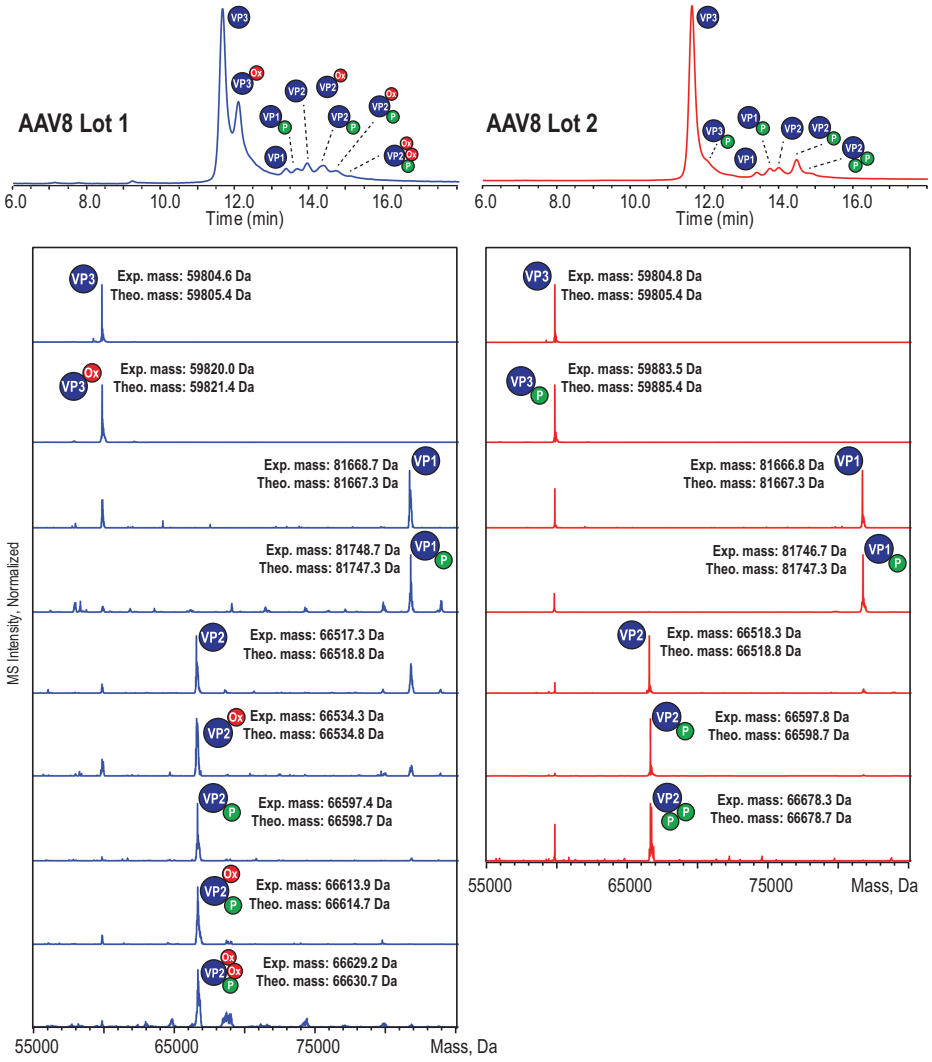


Figure 8.15: Top: Comparison of HILIC chromatograms with fluorescence detection (using λ_{ex} = 280 nm and λ_{em} = 348 nm) of AAV8 from Lots 1 and 2 (blue and red traces, respectively). The peak identities were confirmed by accurate intact mass measurements. Red and green small circles indicate the presence of oxidation and phosphorylation, respectively. Bottom: Deconvoluted mass spectra from the online MS analysis of the two lots of capsid proteins. Adapted from Liu et al. [84] with permission from Elsevier.

HDX MS is now a mature technology commonly used to assess the integrity of the protein higher order structure (see Chapter 4 for more details), and in fact is commonly used to study conformational dynamics of a range of viruses [88]. However, this technique has not yet become a part of the analytical armamentarium supporting

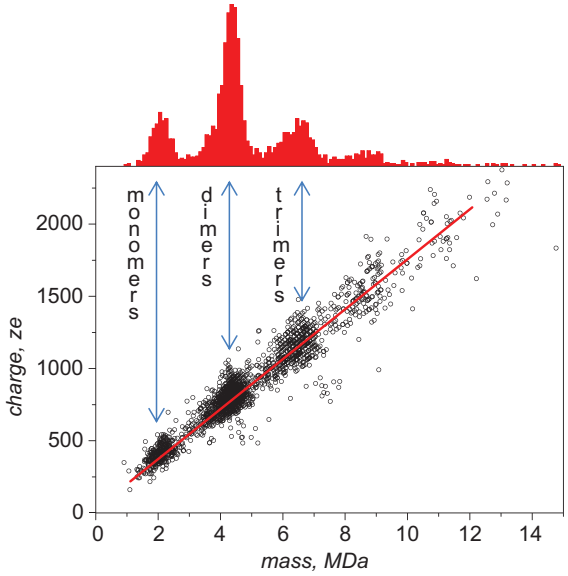


Figure 8.16: Charge detection MS scatter plots of charge versus mass for a single-stranded DNA plasmid M13mp18. Each point on the graph corresponds to an individual measurement. Projection of the scatter plot on the mass axis provides the mass histogram (bin size = 0.1 MDa). Solid red line is representative of the charge state $m/z = 5{,}800$. Adapted with permission from Doussineau et al. [87].

development and production of the gene therapy products, and the emphasis is currently placed on close monitoring the primary structure of the capsid proteins. In contrast, the quaternary structure of the viral particles is a critical quality attribute that needs to be closely monitored, necessitating intact virus characterization. This entails addressing a range of questions, such as the fraction of the empty capsids in the final product, as well as the presence and the level of clumps and aggregates. Empty capsids are usually present in significant quantities and their level must be quantified. Likewise, aggregation is a major limiting factor vis-à-vis concentration of viral particles in the final product (which rarely exceeds 30 µg/mL, corresponding to only 0.7 nM for 4.6 MDa particles).

Despite the large size of the viral capsids, several successful attempts at using native MS to assess their quaternary structure have been reported in the course of the past two decades [89, 90]. However, the usual limitations of native MS, especially those related to the structural heterogeneity of the object of the intact mass measurements (see Chapter 4 for more details), limit the utility of this technique to relatively homogeneous viral capsids, such as those illustrated in Figure 8.17 (it is also noteworthy that the capsid concentration used in that work exceeded what is typically recommended for AAV by two orders of magnitude). Unfortunately, the presence of three different capsid proteins (VP1, VP2, and VP3 in AAV viral particles) of varying stoichiometries, as well as the frequent occurrence of PTMs within these

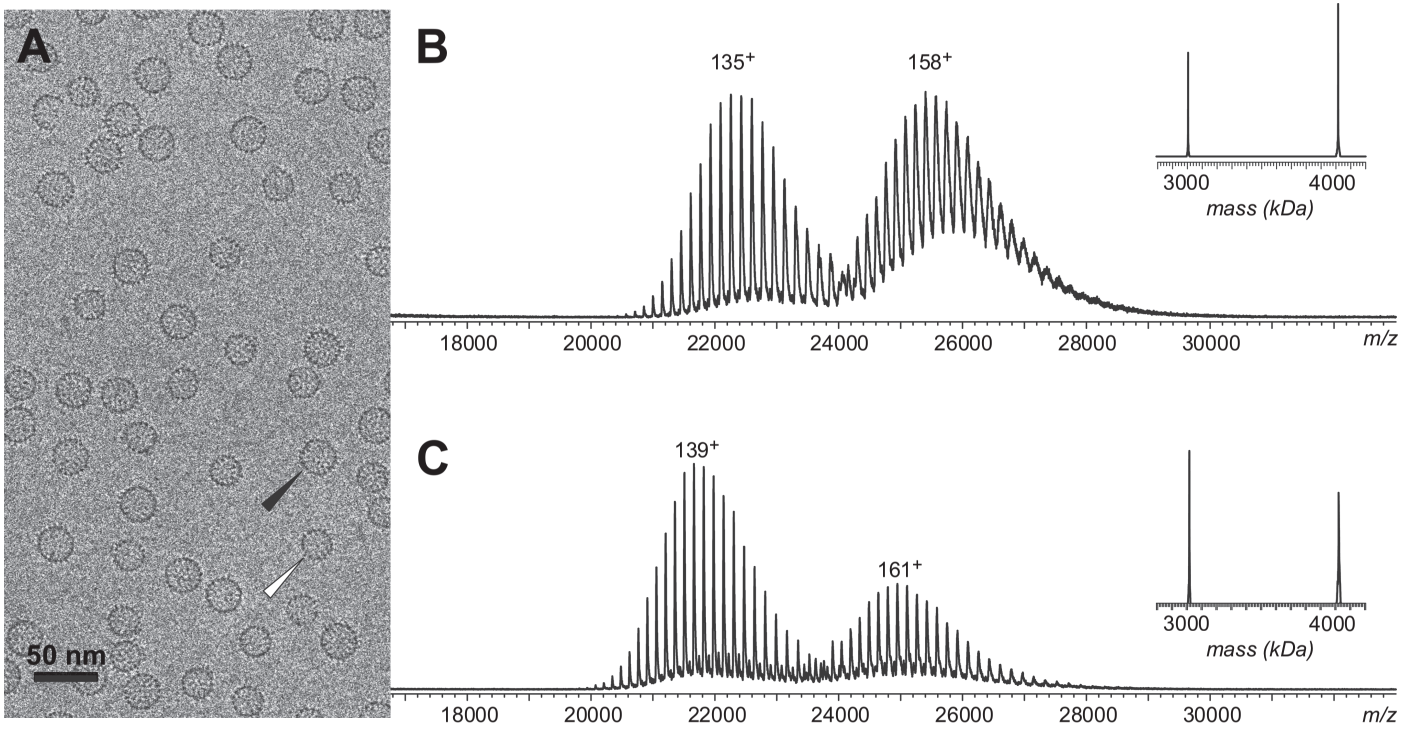


Figure 8.17: Electron microscopy (A) and native MS (B, C) of the hepatitis B viral capsids. (A) Capsids of two sizes are observed: large ones composed of nominally 120 dimers with $T = 4$ geometry (white arrowhead) and small ones composed of nominally 90 dimers with $T = 3$ geometry (black arrowhead). (B) A mass spectrum of cp149 3C3A capsids (composed of dimers without an intramolecular disulfide bond). (C) A mass spectrum of cp149 61 C capsids (with an intramolecular disulfide bond). Both samples were at ca. 0.04 µM capsids (ca. 8 µM monomer of the capsid protein) in 200 mM ammonium acetate (pH 6.8). The distributions of peaks around m/z 22,000 and 25,000 represent $T = 3$ and $T = 4$ capsids, respectively. Adapted with permission from Uetrecht et al. [90]. Copyright (2008) National Academy of Sciences, U.S.A.

proteins (*vise supra*) gives rise to an extremely heterogeneous mixture of capsids (let alone viral particles armed with the payload), which makes extraction of meaningful data from native MS measurements next to impossible. The only feasible way to circumvent this problem is offered by supplementing the native MS measurements with ionic charge measurements [91], a technique that has been mentioned earlier in this section. Charge detection MS (CDMS) has been successfully applied to characterize a range of heterogeneous viral particles, including AAV-based gene therapy products, for which CDMS can quantitatively characterize diverse AAV particle populations including particles packing the complete genome, empty particles, particles packing partial genomes, as well as particles with impurities [92]. This is illustrated in Figure 8.18, which clearly shows the presence of several distinct ionic clusters corresponding to the empty, partial, and full capsids, all having similar charges (the gray box in Figure 8.18A). This indicates that both the complete genome and the partial genome are packaged inside the capsid, as the extent of multiple charging in ESI is determined by the physical size of the macromolecules in solution, and is most closely correlated with the solvent-accessible surface area [93, 94]. Protrusion of the genome's parts outside of the capsid should therefore result in an increase of the number of charges accommodated by the viral particles, explaining the charge distribution tailing observed for the loaded viral particles toward higher z values, but not for empty capsids.

Until recently CDMS measurements required specialized custom-built instrumentation [95–98]; however, at least one manufacturer is now actively engaged in implementing charge detection capability on high-end commercial MS [99]. Above and beyond assessing the integrity of the viral particles, CDMS provides an opportunity to monitor their encounters with physiological partners, such as antibodies, as was recently demonstrated for a canine parvovirus interaction with multiple (up to 60 per single capsid) Fab fragments [100]. There is little doubt that the scope of this technique will continue to expand in the near future to include interactions of viral particles employed in gene therapy with their therapeutic targets (cell surface receptors) as well.

One disadvantage of CDMS as a single-ion measurement technique is the relatively low speed of analysis. An alternative approach to characterizing viral particles and related nano-objects is offered by gas-phase electrophoretic mobility molecular analyzer (GEMMA) coupled to an ESI source [101–103]. GEMMA is based on separation of solvent-free analyte molecules at ambient pressure [104]. Singly charged large macromolecules and/or nano-objects are separated according to their electrophoretic mobility diameter (which has a straightforward correlation with the particle's mass in the case of spherical analytes). In addition to exceeding CDMS in the speed of analysis, GEMMA provides an additional benefit of being less dependent on the sample purity [101]. A disadvantage of GEMMA is its reliance on singly charged ions; the presence of doubly and triply charged species in the ionic population results in crowding of the mobility spectra and makes their interpretation less straightforward. Nevertheless, continuous improvements of this technique are likely to make it a powerful complement to CDMS.

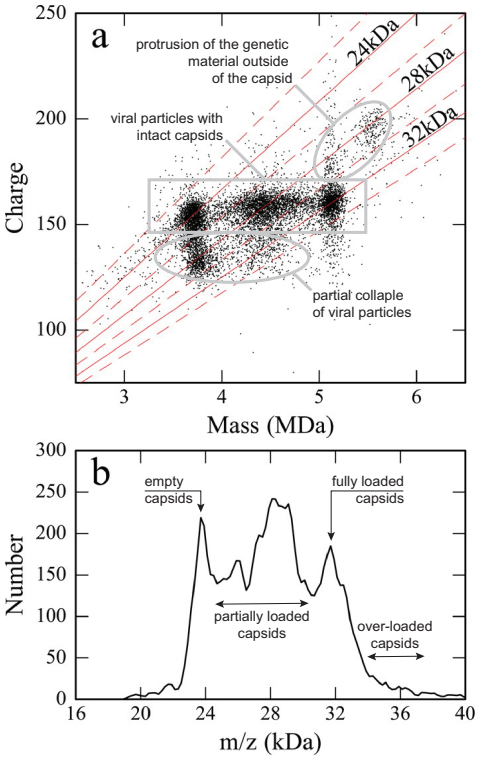


Figure 8.18: (A) A charge versus mass scatter plot for AAV8 packaging an scDNA genome (formed by joining two sequence-inverted vectors by a hairpin, so the effective length of the unique transgene sequence is halved). The red diagonal lines are lines of constant m/z. (B) An m/z histogram for AAV8 packaging an scDNA genome. Adapted with permission from Pierson et al. [92]. Copyright 2016 American Chemical Society.

So far, the discussion of the analytical methods supporting gene therapy product development and production mostly focused on their structural aspects. However, a range of other questions have to be addressed during various stages of the development process, such as the fraction of the dosed product present in circulation and in the target tissue, as well as the duration of time during which the product remains there. A variety of quantitative MS-based approaches can be used to address these questions; since they have been already covered in detail in Chapter 7, no discussion will be provided here to avoid redundancy. These methods can also be used to address another critical question related to the success or failure of a specific gene therapy product, namely the change of the target protein expression level. Lastly, MS-base quantitation can also be used to assess the possibility of the vector genome being shed (released) from the patient, which is now falling under increased scrutiny due to the possibility of the vector's transfer to other humans.

8.4 MS in the analytical support of cell-based therapies

Cell therapies are another class of novel/emerging therapeutic approaches in which new or modified cells are introduced into a patient's body aiming at treating a disease or repairing damaged tissues. While the chimeric antigen receptors (CAR) T-cell therapies remain the best known (and so far the most successful) type of therapies of this type, other cell-based immunotherapeutic strategies are being actively pursued as well (T-cell receptor therapies, natural killer cell therapies, tumor infiltrating lymphocytes, and marrow-derived lymphocytes to name just a few) [105]. Analytical methods supporting development and production of cell-based therapies have to address a range of questions related to both distribution and persistence of the product. While the specific set of analytical methods depends on a particular system that needs to be tested, the common tools are flow cytometry, quantitative PCR and immunogenicity measurements. Mass spectrometry has a tremendous potential in this field, and is already beginning to play a very visible role in analytical support of various stages of design, development and production of cell-based therapies [106, 107]. For example, the success of any cell-based therapeutic strategy is determined by the ability of the transferred cells to survive, expand and function properly in vivo. Therefore, expression/production of the appropriate biomarkers [108–111] can be used as a means of evaluating the status of the transferred cells. The levels of these biomarkers can be effectively measured using a vast arsenal of MS-based methods developed in the past two decades for proteomics and metabolomics applications. While a detailed discussion of these approaches is certainly beyond the scope of our book, an interested reader is referred to several excellent reviews [107, 112–115], and books [116–118] on this subject. Furthermore, the dramatic progress made in recent years in the field of single-cell MS [119, 120], provides reasonable assurance that the MS-based methods of analysis will soon become standard analytical tools supporting both optimization and quality control in the production of a range of cell-based therapies.

8.5 MS in characterization of modern vaccines

A significant number of modern vaccines are produced using biotechnological tools, and their fraction continues to expand [122]. These can be classified into several categories, including protein subunit-based vaccines, virus-like particles vaccines, virus vectored vaccines, as well as mRNA and DNA vaccines (see Figure 8.19, which uses COVID-19 vaccines as an example). Many of these vaccine types fall into the categories that have been already considered in detail in this chapter. For example, mRNA vaccines utilize a synthetically produced nucleic acid strand, analysis of which is considered in Section 8.1, while both virus-like particle vaccines and virus-vectored vaccines resemble gene therapy products in their structural organization (MS-based methods of their analysis were considered in Section 8.3).

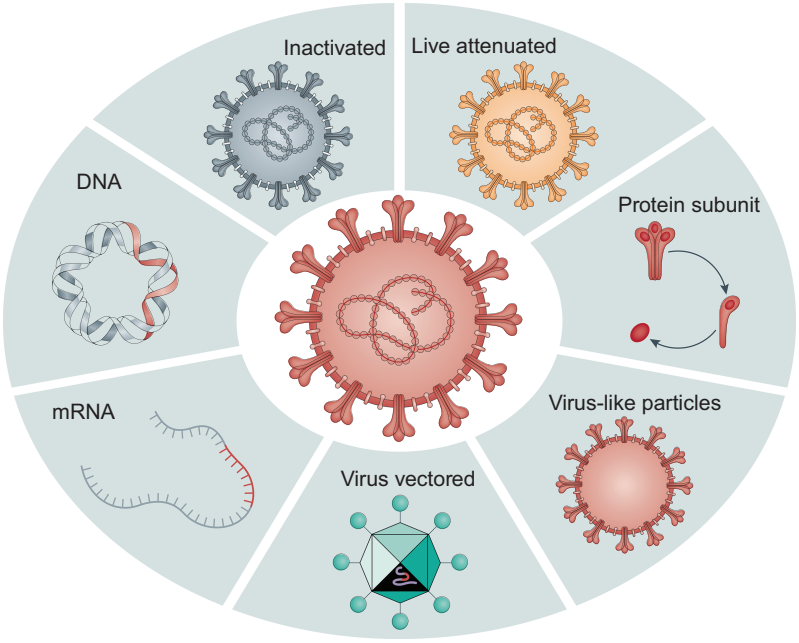


Figure 8.19: Vaccine strategies for SARS-CoV-2 (clockwise from the top). **Inactivated**: the viruses are physically or chemically inactivated but the integrity of the virus particle is preserved, which serves as the immunogen. **Live-attenuated**: the virus is attenuated by in vitro or in vivo passage, or reverse-genetic mutagenesis. It becomes non-pathogenic or weakly pathogenic but retains immunogenicity by mimicking the live virus infection. **Protein subunit**: only key viral proteins or peptides are expressed and serve as immunogens. **Virus-like particle**: non-infectious particles resembling real virions but lacking the virus genome act as the immunogens. **Virus-vectored**: Gene (s) encoding pathogen antigen(s) are cloned into non-replicating or replicating virus vectors (e.g., adenovirus). The antigen(s) are produced by transduced host cells after immunization. **mRNA vaccines**: synthesized by in vitro transcription; produce viral antigen(s) in the cytoplasm through direct protein translation in vivo. **DNA vaccines**: viral antigen(s) encoded by a recombinant DNA plasmid are produced in the host cells. Adopted by permission from Springer Nature Customer Service Centre GmbH: Dai and Gao [121].

Another approach to creating potent vaccines utilizes either synthetically or recombinantly produced antigens (e.g., protein subunits) that are chemically conjugated to a recombinant carrier protein, which is typically chosen to elicit a strong T-cell response [123] (such as a non-toxic mutant of a diphtheria toxin CRM197 [124]). Presentation of multiple copies of the antigen on the carrier protein surface causes the B-cell immune receptors to cluster together, triggering a specific immune response to the antigen. In many cases, this allows an immune response to be developed to molecular entities – haptens – that are too small to be immunogenic on their own (e.g., nicotine in smoking-secession vaccines [125]). Haptenation (conjugation of antigens to a carrier protein) [126] gives rise to highly heterogeneous products due to the

variation of the extent of conjugation (the hapten load), as well as the presence of the by-products representing non-productive (dead-ended) conjugation side-reactions. Technically speaking, haptenated carrier proteins fall into the protein/small-molecule conjugate category. Characterization of such products has been considered in detail in Chapter 3 (covalent structure) and Chapter 4 (higher order structure), and will not be repeated here. A more in-depth coverage of various MS-based methods used in vaccine development can be found in a recent review by Sharma et al. [127].

8.6 MS in characterization of nanomedicines

The term "nanomedicines" is somewhat loosely defined, and is frequently applied to a variety of objects that are grouped together solely on the base of their size (the nanometer scale and up), and includes a variety of gene therapy products and functionalized nanoparticles. A narrower definition excludes the objects that are produced by means of biotechnology alone (such as AAV-based gene therapy products considered earlier in this chapter). Strictly speaking, the nanometer-sized particles that are not produced with the biotechnology tools are nanomaterials (a class of nano-objects that includes nanoparticles of various types, nano-tubes, liposomes, DNA pyramids, and micelles, to name a few). Many of these objects possess unique physical properties that are actively exploited for the drug delivery purposes, including targeted delivery of a range of biopharmaceuticals. Because of the diversity of the nanomaterial-based or nanomaterial-containing products, analytical characterization of each subclass should be approached individually, targeting the product-specific set of CQAs. Perhaps only one characteristic of this class of medicines – the particle size distribution – is a CQA that is common to all such products. While this task is usually accomplished by light scattering-based or high-resolution imaging techniques [128], it seems that certain MS-based techniques considered earlier in this chapter, such as CDMS and GEMMA, are likely to be included in the analytical armamentarium of nanomaterials characterization as well (in fact, the potential of CDMS in characterizing heterogeneous exosomes has been already demonstrated [129]). Another area of nanomedicine characterization, where the role of mass spectrometry is already indisputable, is ex vivo imaging. MS-based imaging allows the spatial distribution of nanomedicines within specific organs and tissues to be determined, and their impact on the biochemical microenvironment to be evaluated. A detailed discussion of this field is beyond the scope of our book, but the interested readers are referred to several excellent recent reviews on this subject [130, 131].

References

[1] Smith, C.I.E. & Zain, R. Therapeutic oligonucleotides: state of the art. Annu. Rev. Pharmacol. Toxicol. 59, 605–630 (2019).

[2] Offord, C. Waiting for oligonucleotide therapeutics. Scientist 30, 58–61 (2016).

[3] De Smet, M.D., Meenken, C.J. & Van Den Horn, G.J. Fomivirsen – a phosphorothioate oligonucleotide for the treatment of CMV retinitis. Ocul. Immunol. Inflamm. 7, 189–198 (1999).

[4] Perry, C.M. & Balfour, J.A. Fomivirsen. Drugs 57, 375–380 (1999).

[5] Kuijper, E.C., Bergsma, A.J., Pijnappel, W. & Aartsma-Rus, A. Opportunities and challenges for antisense oligonucleotide therapies. J. Inherit. Metab. Dis. (2020).

[6] Adachi, T. & Nakamura, Y. Aptamers: a review of their chemical properties and modifications for therapeutic application. Molecules 24, 4229 (2019).

[7] Siddiqui, M.A. & Keating, G.M. Pegaptanib: in exudative age-related macular degeneration. Drugs 65, 1571–1577 (2005).

[8] Pourshahian, S. Therapeutic oligonucleotides, impurities, degradants, and their characterization by mass spectrometry. Mass Spectrom. Rev.40, 75–109 (2021),

[9] Yokoi, H., Kasahara, Y., Obika, S., Doi, T. & Kamada, H. Development of a detection method for antisense oligonucleotides in mouse kidneys by matrix-assisted laser desorption/ ionization imaging mass spectrometry. Rapid Commun. Mass Spectrom. 32, 1984–1990 (2018).

[10] Herkt, M., Foinquinos, A., Batkai, S., Thum, T. & Pich, A. Pharmacokinetic studies of antisense oligonucleotides using MALDI-TOF mass spectrometry. Front. Pharmacol. 11, 220 (2020).

[11] Li, Y., Tang, K., Little, D.P., Köster, H., Hunter, R.L. & McIver, R.T. Jr. High-resolution MALDI Fourier transform mass spectrometry of oligonucleotides. Anal. Chem. 68, 2090–2096 (1996).

[12] Bennett, C.F. Therapeutic antisense oligonucleotides are coming of age. Annu. Rev. Med. 70, 307–321 (2019).

[13] Lobue, P.A., Jora, M., Addepalli, B. & Limbach, P.A. Oligonucleotide analysis by hydrophilic interaction liquid chromatography-mass spectrometry in the absence of ion-pair reagents. J. Chromatogr. A 1595, 39–48 (2019).

[14] Goyon, A., Yehl, P. & Zhang, K. Characterization of therapeutic oligonucleotides by liquid chromatography. J. Pharm. Biomed. Anal. 182, 113105 (2020).

[15] Lin, Z.J., Li, W. & Dai, G. Application of LC-MS for quantitative analysis and metabolite identification of therapeutic oligonucleotides. J. Pharm. Biomed. Anal. 44, 330–341 (2007).

[16] McLuckey, S.A., Van Berkel, G.J. & Glish, G.L. Tandem mass spectrometry of small, multiply charged oligonucleotides. J. Am. Soc. Mass Spectrom. 3, 60–70 (1992).

[17] Sannes-Lowery, K.A. & Hofstadler, S.A. Sequence confirmation of modified oligonucleotides using IRMPD in the external ion reservoir of an electrospray ionization Fourier transform ion cyclotron mass spectrometer. J. Am. Soc. Mass Spectrom. 14, 825–833 (2003).

[18] Nyakas, A., Eberle, R.P., Stucki, S.R. & Schürch, S. More than charged base loss – revisiting the fragmentation of highly charged oligonucleotides. J. Am. Soc. Mass Spectrom. 25, 1155–1166 (2014).

[19] Van Dongen, W.D. & Niessen, W.M. Bioanalytical LC-MS of therapeutic oligonucleotides. Bioanalysis 3, 541–564 (2011).

[20] Sutton, J.M., Kim, J., El Zahar, N.M. & Bartlett, M.G. Bioanalysis and biotransformation of oligonucleotide therapeutics by liquid chromatography-mass spectrometry. Mass Spectrom. Rev. 40, 334–358 (2021)

[21] Largy, E. & Gabelica, V. Native Hydrogen/deuterium exchange mass spectrometry of structured DNA oligonucleotides. Anal. Chem. 92, 4402–4410 (2020).
[22] Guo, X., Bruist, M.F., Davis, D.L. & Bentzley, C.M. Secondary structural characterization of oligonucleotide strands using electrospray ionization mass spectrometry. Nucl. Acids Res. 33, 3659–3666 (2005).
[23] Steven, A., Hofstadler, K.A.S.-L. & Griffey R.H. Enhanced gas-phase hydrogen-deuterium exchange of oligonucleotide and protein ions stored in an external multipole ion reservoir. J. Mass Spectrom. 35, 62–70 (2000).
[24] Zhang, J., Loo, R.R.O. & Loo, J.A. Structural characterization of a thrombin-aptamer complex by high resolution native top-down mass spectrometry. J. Am. Soc. Mass Spectrom. 28, 1815–1822 (2017).
[25] Ganem, B., Li, Y.-T. & Henion, J.D. Detection of oligonucleotide duplex forms by ion-spray mass spectrometry. Tetrahedron Lett. 34, 1445–1448 (1993).
[26] Howell, W.H. & Holt, E. Two new factors in blood coagulation – heparin and pro-antithrombin. Am. J. Physiol. 47, 328–341 (1918).
[27] Murray, G.D. & Best, C.H. The use of heparin in thrombosis. Ann. Surg. 108, 163–177 (1938).
[28] Hedenius, P. & Wilander., O. The influence of intravenous injections of heparin in man on the time of coagulation. Acta Med. Scandinav. 88, 443–449 (1936).
[29] Oduah, E.I., Linhardt, R.J. & Sharfstein, S.T. Heparin: past, present, and future. Pharmaceuticals 9 (2016).
[30] Lindahl, U. What else can 'Heparin' do? Haemostasis 29, Suppl S1, 38–47 (1999).
[31] Lima, M., Rudd, T. & Yates, E. New applications of Heparin and other Glycosaminoglycans. Molecules 22 (2017).
[32] Lindahl, U. & Li, J.P. Heparin – an old drug with multiple potential targets in COVID-19 therapy. J. Thromb. Haemost. 18, 2422–2424 (2020).
[33] Thachil, J. The versatile heparin in COVID-19. J. Thromb. Haemost. 18, 1020–1022 (2020).
[34] Fu, L., Suflita, M. & Linhardt, R.J. Bioengineered heparins and heparan sulfates. Adv. Drug Deliv. Rev. 97, 237–249 (2016).
[35] Farrugia, B.L., Lord, M.S., Melrose, J. & Whitelock, J.M. Can we produce heparin/heparan sulfate biomimetics using "mother-nature" as the gold standard? Molecules 20, 4254–4276 (2015).
[36] Collee, J.G. Bovine spongiform encephalopathy. Lancet 336, 1300–1303 (1990).
[37] Kishimoto, T.K., Viswanathan, K., Ganguly, T., Elankumaran, S., Smith, S., Pelzer, K., Lansing, J.C., Sriranganathan, N., Zhao, G., Galcheva-Gargova, Z., Al-Hakim, A., Bailey, G.S., Fraser, B., Roy, S., Rogers-Cotrone, T., Buhse, L., Whary, M., Fox, J., Nasr, M., Dal Pan, G.J. et al. Contaminated heparin associated with adverse clinical events and activation of the contact system. N. Engl. J. Med. 358, 2457–2467 (2008).
[38] Szajek, A.Y., Chess, E., Johansen, K., Gratzl, G., Gray, E., Keire, D., Linhardt, R.J., Liu, J., Morris, T., Mulloy, B., Nasr, M., Shriver, Z., Torralba, P., Viskov, C., Williams, R., Woodcock, J., Workman, W. & Al-Hakim, A. The US regulatory and pharmacopeia response to the global heparin contamination crisis. Nat. Biotechnol. 34, 625–630 (2016).
[39] Ori, A., Wilkinson, M.C. & Fernig, D.G. A systems biology approach for the investigation of the heparin/heparan sulfate interactome. J. Biol. Chem. 286, 19892–19904 (2011).
[40] Pomin, V.H. & Mulloy, B. Current structural biology of the heparin interactome. Curr. Opin. Struct. Biol. 34, 17–25 (2015).
[41] Vanboeckel, C.A.A. & Petitou, M. The unique antithrombin-III binding domain of heparin – a lead to new synthetic antithrombotics. Angew. Chem. Int. Ed. 32, 1671–1690 (1993).
[42] Nugent, M.A. Heparin sequencing brings structure to the function of complex oligosaccharides. Proc. Natl. Acad. Sci. U.S.A. 97, 10301–10303 (2000).

[43] Suflita, M., Fu, L., He, W., Koffas, M. & Linhardt, R.J. Heparin and related polysaccharides: synthesis using recombinant enzymes and metabolic engineering. Appl. Microbiol. Biotechnol. 99, 7465–7479 (2015).
[44] Vaidyanathan, D., Williams, A., Dordick, J.S., Koffas, M.A.G. & Linhardt, R.J. Engineered heparins as new anticoagulant drugs. Bioeng. Transl. Med. 2, 17–30 (2017).
[45] Kang, Z., Zhou, Z., Wang, Y., Huang, H., Du, G. & Chen, J. Bio-based strategies for producing glycosaminoglycans and their oligosaccharides. Trends Biotechnol. 36, 806–818 (2018).
[46] Lord, M.S., Cheng, B., Tang, F., Lyons, J.G., Rnjak-Kovacina, J. & Whitelock, J.M. Bioengineered human heparin with anticoagulant activity. Metab. Eng. 38, 105–114 (2016).
[47] Henriksen, J., Ringborg, L.H. & Roepstorrf, P. On-line size-exclusion chromatography/mass spectrometry of low molecular mass heparin. J. Mass Spectrom. 39, 1305–1312 (2004).
[48] Lesur, D., Duhirwe, G. & Kovensky, J. High resolution MALDI-TOF-MS and MS/MS: application for the structural characterization of sulfated oligosaccharides. Eur. J. Mass Spectrom. 25, 428–436 (2019).
[49] Tissot, B., Gasiunas, N., Powell, A.K., Ahmed, Y., Zhi, Z.L., Haslam, S.M., Morris, H.R., Turnbull, J.E., Gallagher, J.T. & Dell, A. Towards GAG glycomics: analysis of highly sulfated heparins by MALDI-TOF mass spectrometry. Glycobiology 17, 972–982 (2007).
[50] Bultel, L., Landoni, M., Grand, E., Couto, A.S. & Kovensky, J. UV-MALDI-TOF mass spectrometry analysis of heparin oligosaccharides obtained by nitrous acid controlled degradation and high performance anion exchange chromatography. J. Am. Soc. Mass Spectrom. 21, 178–190 (2010).
[51] Henriksen, J., Roepstorff, P. & Ringborg, L.H. Ion-pairing reversed-phased chromatography/mass spectrometry of heparin. Carbohydr. Res. 341, 382–387 (2006).
[52] Zaia, J. On-line separations combined with MS for analysis of glycosaminoglycans. Mass Spectrom. Rev. 28, 254–272 (2009).
[53] Zaia, J., Khatri, K., Klein, J., Shao, C., Sheng, Y. & Viner, R. Complete molecular weight profiling of low-molecular weight heparins using size exclusion chromatography-ion suppressor-high-resolution mass spectrometry. Anal. Chem. 88, 10654–10660 (2016).
[54] Zhang, Q., Chen, X., Zhu, Z., Zhan, X., Wu, Y., Song, L. & Kang, J. Structural analysis of low molecular weight Heparin by ultraperformance size exclusion chromatography/time of flight mass spectrometry and capillary zone electrophoresis. Anal. Chem. 85, 1819–1827 (2013).
[55] Niu, C., Zhao, Y., Bobst, C.E., Savinov, S.N. & Kaltashov, I.A. Identification of protein recognition elements within heparin chains using enzymatic foot-printing in solution and online SEC/MS. Anal. Chem. 92, 7565–7573 (2020).
[56] Brustkern, A.M., Buhse, L.F., Nasr, M., Al-Hakim, A. & Keire, D.A. Characterization of currently marketed heparin products: reversed-phase ion-pairing liquid chromatography mass spectrometry of heparin digests. Anal. Chem. 82, 9865–9870 (2010).
[57] Li, D., Chi, L., Jin, L., Xu, X., Du, X., Ji, S. & Chi, L. Mapping of low molecular weight heparins using reversed phase ion pair liquid chromatography–mass spectrometry. Carbohydr. Polym. 99, 339–344 (2014).
[58] Du, J.Y., Chen, L.R., Liu, S., Lin, J.H., Liang, Q.T., Lyon, M. & Wei, Z. Ion-pairing liquid chromatography with on-line electrospray ion trap mass spectrometry for the structural analysis of N-unsubstituted heparin/heparan sulfate. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 1028, 71–76 (2016).
[59] Miller, R.L., Guimond, S.E., Shivkumar, M., Blocksidge, J., Austin, J.A., Leary, J.A. & Turnbull, J.E. Heparin isomeric oligosaccharide separation using volatile salt strong anion exchange chromatography. Anal. Chem. 88, 11542–11550 (2016).

[60] Li, L., Zhang, F., Zaia, J. & Linhardt, R.J. Top-down approach for the direct characterization of low molecular weight heparins using LC-FT-MS. Anal. Chem. 84, 8822–8829 (2012).
[61] Lin, L., Liu, X., Zhang, F., Chi, L., Amster, I.J., Leach, F.E. 3rd, Xia, Q. & Linhardt, R.J. Analysis of heparin oligosaccharides by capillary electrophoresis-negative-ion electrospray ionization mass spectrometry. Anal. Bioanal. Chem. 409, 411–420 (2017).
[62] Naggar, E.F., Costello, C.E. & Zaia, J. Competing fragmentation processes in tandem mass spectra of heparin-like glycosaminoglycans. J. Am. Soc. Mass Spectrom. 15, 1534–1544 (2004).
[63] Kailemia, M.J., Li, L., Ly, M., Linhardt, R.J. & Amster, I.J. Complete mass spectral characterization of a synthetic ultralow-molecular-weight heparin using collision-induced dissociation. Anal. Chem. 84, 5475–5478 (2013).
[64] Huang, Y., Yu, X., Mao, Y., Costello, C.E., Zaia, J. & Lin, C. De novo sequencing of heparan sulfate oligosaccharides by electron-activated dissociation. Anal. Chem. 85, 11979–11986 (2013).
[65] Leach, F.E. 3rd, Riley, N.M., Westphall, M.S., Coon, J.J. & Amster, I.J. Negative electron transfer dissociation sequencing of increasingly sulfated glycosaminoglycan oligosaccharides on an orbitrap mass spectrometer. J. Am. Soc. Mass Spectrom. 28, 1844–1854 (2017).
[66] Klein, D.R., Leach, F.E. 3rd, Amster, I.J. & Brodbelt, J.S. Structural characterization of glycosaminoglycan carbohydrates using ultraviolet photodissociation. Anal. Chem. 91, 6019–6026 (2019).
[67] Hogan, J.D., Klein, J.A., Wu, J., Chopra, P., Boons, G.J., Carvalho, L., Lin, C. & Zaia, J. Software for peak finding and elemental composition assignment for glycosaminoglycan tandem mass spectra. Mol. Cell. Proteomics 17, 1448–1456 (2018).
[68] Wu, J., Wei, J., Chopra, P., Boons, G.J., Lin, C. & Zaia, J. Sequencing Heparan Sulfate Using HILIC LC-NETD-MS/MS. Anal. Chem. 91, 11738–11746 (2019).
[69] Kailemia, M.J., Park, M., Kaplan, D.A., Venot, A., Boons, G.J., Li, L., Linhardt, R.J. & Amster, I.J. High-field asymmetric-waveform ion mobility spectrometry and electron detachment dissociation of isobaric mixtures of glycosaminoglycans. J. Am. Soc. Mass Spectrom. 25, 258–268 (2014).
[70] Zhao, Y., Abzalimov, R.R. & Kaltashov, I.A. Interactions of Intact unfractionated heparin with its client proteins can be probed directly using native electrospray ionization mass spectrometry. Anal. Chem. 88, 1711–1718 (2016).
[71] Abzalimov, R.R., Dubin, P.L. & Kaltashov, I.A. Glycosaminoglycans as naturally occurring combinatorial libraries: Developing a mass spectrometry-based strategy for characterization of anti-thrombin interaction with low molecular weight heparin and heparin oligomers. Anal. Chem. 79, 6055–6063 (2007).
[72] Singh, A., Kett, W.C., Severin, I.C., Agyekum, I., Duan, J., Amster, I.J., Proudfoot, A.E., Coombe, D.R. & Woods, R.J. The interaction of heparin tetrasaccharides with chemokine CCL5 Is modulated by sulfation pattern and pH. J. Biol. Chem. 290, 15421–15436 (2015).
[73] Minsky, B.B., Dubin, P.L. & Kaltashov, I.A. Electrostatic forces as dominant interactions between proteins and polyanions: an ESI MS study of fibroblast growth factor binding to heparin oligomers. J. Am. Soc. Mass Spectrom. 28, 758–767 (2017).
[74] Yang, Y., Du, Y. & Kaltashov, I.A. The utility of native MS for understanding the mechanism of action of repurposed therapeutics in COVID-19: heparin as a disruptor of the SARS-CoV-2 interaction with its host cell receptor. Anal. Chem. 92, 10930–10934 (2020).
[75] Niu, C., Yang, Y., Huynh, A., Nazy, I. & Kaltashov, I.A. Platelet factor 4 interactions with short heparin oligomers: implications for folding and assembly. Biophys. J. 118, 1371–1379 (2020).
[76] Kotterman, M.A., Chalberg, T.W. & Schaffer, D.V. Viral vectors for gene therapy: translational and clinical outlook. Annu. Rev. Biomed. Eng. 17, 63–89 (2015).

[77] Hanna, E., Rémuzat, C., Auquier, P. & Toumi, M. Gene therapies development: slow progress and promising prospect. J. Mark. Access Health Policy 5, 1265293 (2017).
[78] Shahryari, A., Saghaeian Jazi, M., Mohammadi, S., Razavi Nikoo, H., Nazari, Z., Hosseini, E.S., Burtscher, I., Mowla, S.J. & Lickert, H. Development and clinical translation of approved gene therapy products for genetic disorders. Front. Genet. 10, 868 (2019).
[79] Wang, F., Qin, Z., Lu, H., He, S., Luo, J., Jin, C. & Song, X. Clinical translation of gene medicine. J. Gene Med. 21, e3108 (2019).
[80] Jin, X., Liu, L., Nass, S., O'Riordan, C., Pastor, E. & Zhang, X.K. Direct liquid chromatography/mass spectrometry analysis for complete characterization of recombinant adeno-associated virus capsid proteins. Hum. Gene Ther. Methods 28, 255–267 (2017).
[81] Aloor, A., Zhang, J., Gashash, E.A., Parameswaran, A., Chrzanowski, M., Ma, C., Diao, Y., Wang, P.G. & Xiao, W. Site-specific N-glycosylation on the AAV8 capsid protein. Viruses 10 (2018).
[82] Mary, B., Maurya, S., Arumugam, S., Kumar, V. & Jayandharan, G.R. Post-translational modifications in capsid proteins of recombinant adeno-associated virus (AAV) 1-rh10 serotypes. FEBS J. 286, 4964–4981 (2019).
[83] Kondratov, O., Marsic, D., Crosson, S.M., Mendez-Gomez, H.R., Moskalenko, O., Mietzsch, M., Heilbronn, R., Allison, J.R., Green, K.B., Agbandje-mckenna, M. & Zolotukhin, S. Direct head-to-head evaluation of recombinant adeno-associated viral vectors manufactured in human versus insect cells. Mol. Ther. 25, 2661–2675 (2017).
[84] Liu, A.P., Patel, S.K., Xing, T., Yan, Y., Wang, S. & Li, N. Characterization of adeno-associated virus capsid proteins using hydrophilic interaction chromatography coupled with mass spectrometry. J. Pharm. Biomed. Anal. 189, 113481 (2020).
[85] Benner, W.H. A gated electrostatic ion trap to repetitiously measure the charge and m/z of large electrospray ions. Anal. Chem. 69, 4162–4168 (1997).
[86] Schultz, J.C., Hack, C.A. & Benner, W.H. Mass determination of megadalton-DNA electrospray ions using charge detection mass spectrometry. J. Am. Soc. Mass Spectrom. 9, 305–313 (1998).
[87] Doussineau, T., Paletto, P., Dugourd, P. & Antoine, R. Multiphoton dissociation of electrosprayed megadalton-sized DNA ions in a charge-detection mass spectrometer. J. Am. Soc. Mass Spectrom. 26, 7–13 (2015).
[88] Guttman, M. & Lee, K.K. Isotope Labeling of biomolecules: structural analysis of viruses by HDX-MS. Meth. Enzymol. 566, 405–426 (2016).
[89] Uetrecht, C. & Heck, A.J. Modern biomolecular mass spectrometry and its role in studying virus structure, dynamics, and assembly. Angew. Chem. Int. Ed. 50, 8248–8262 (2011).
[90] Uetrecht, C., Versluis, C., Watts, N.R., Roos, W.H., Wuite, G.J.L., Wingfield, P.T., Steven, A.C. & Heck, A.J.R. High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. Proc. Natl. Acad. Sci. U.S.A. 105, 9216–9220 (2008).
[91] Keifer, D.Z. & Jarrold, M.F. Single-molecule mass spectrometry. Mass Spectrom. Rev. 36, 715–733 (2017).
[92] Pierson, E.E., Keifer, D.Z., Asokan, A. & Jarrold, M.F. Resolving Adeno-Associated Viral particle diversity with charge detection mass spectrometry. Anal. Chem. 88, 6718–6725 (2016).
[93] Kaltashov, I.A. & Mohimen, A. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. Anal. Chem. 77, 5370–5379 (2005).
[94] Kaltashov, I.A. & Abzalimov, R.R. Do ionic charges in ESI MS provide useful information on macromolecular structure? J. Am. Soc. Mass Spectrom. 19, 1239–1246 (2008).

[95] Contino, N.C., Pierson, E.E., Keifer, D.Z. & Jarrold, M.F. Charge detection mass spectrometry with resolved charge states. J. Am. Soc. Mass Spectrom. 24, 101–108 (2013).
[96] Draper, B.E. & Jarrold, M.F. Real-time analysis and signal optimization for charge detection mass spectrometry. J. Am. Soc. Mass Spectrom. 30, 898–904 (2019).
[97] Todd, A.R. & Jarrold, M.F. Dramatic Improvement in sensitivity with pulsed mode charge detection mass spectrometry. Anal. Chem. 91, 14002–14008 (2019).
[98] Todd, A.R. & Jarrold, M.F. Dynamic calibration enables high-accuracy charge measurements on individual ions for charge detection mass spectrometry. J. Am. Soc. Mass Spectrom. 31, 1241–1248 (2020).
[99] Wörner, T.P., Snijder, J., Bennett, A., Agbandje-mckenna, M., Makarov, A.A. & Heck, A.J.R. Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. Nat. Methods 17, 395–398 (2020).
[100] Dunbar, C.A., Callaway, H.M., Parrish, C.R. & Jarrold, M.F. Probing antibody binding to canine parvovirus with charge detection mass spectrometry. J. Am. Chem. Soc. 140, 15701–15711 (2018).
[101] Weiss, V.U., Pogan, R., Zoratto, S., Bond, K.M., Boulanger, P., Jarrold, M.F., Lyktey, N., Pahl, D., Puffler, N., Schelhaas, M., Selivanovitch, E., Uetrecht, C. & Allmaier, G. Virus-like particle size and molecular weight/mass determination applying gas-phase electrophoresis (native nES GEMMA). Anal. Bioanal. Chem. 411, 5951–5962 (2019).
[102] Fernández-García, J., Compton, S., Wick, D. & Fernandez De La Mora, J. Virus size analysis by gas-phase mobility measurements: resolution limits. Anal. Chem. 91, 12962–12970 (2019).
[103] Tseng, Y.H. & Pease, L.F. 3rd. Electrospray differential mobility analysis for nanoscale medicinal and pharmaceutical applications. Nanomed. Nanotechnol. Biol. Med. 10, 1591–1600 (2014).
[104] Kaddis, C., Lomeli, S., Yin, S., Berhane, B., Apostol, M., Kickhoefer, V., Rome, L. & Loo, J. Sizing large proteins and protein complexes by electrospray ionization mass spectrometry and ion mobility. J. Am. Soc. Mass Spectrom. 18, 1206–1216 (2007).
[105] Patel, S., Burga, R.A., Powell, A.B., Chorvinsky, E.A., Hoq, N., McCormack, S.E., Van Pelt, S.N., Hanley, P.J. & Cruz, C.R.Y. Beyond CAR T cells: other cell-based immunotherapeutic strategies against cancer. Front. Oncol. 9, 196 (2019).
[106] Wei, D., Kim, Y., Sugimoto, H., Dong, L. & Qian, M.G. Hybrid LC-MS as a powerful tool for supporting protein bioanalysis in gene and cell therapies. Bioanalysis 12, 977–979 (2020),
[107] Lombard-Banek, C. & Schiel, J.E. Mass spectrometry advances and perspectives for the characterization of emerging adoptive cell therapies. Molecules 25 (2020).
[108] Kalos, M. Biomarkers in T cell therapy clinical trials. J. Transl. Med. 9, 138 (2011).
[109] Meyfour, A., Pahlavan, S., Mirzaei, M., Krijgsveld, J., Baharvand, H. & Salekdeh, G.H. The quest of cell surface markers for stem cell therapy. Cell. Mol. Life Sci. (2020).
[110] Bulte, J.W.M. & Daldrup-Link, H.E. Clinical tracking of cell transfer and cell transplantation: trials and tribulations. Radiology 289, 604–615 (2018).
[111] Perrin, J., Capitao, M., Mougin-Degraef, M., Guérard, F., Faivre-Chauvet, A., Rbah-Vidal, L., Gaschet, J., Guilloux, Y., Kraeber-Bodéré, F., Chérel, M. & Barbet, J. Cell tracking in cancer immunotherapy. Front. Med. 7, 34 (2020).
[112] Walzer, M. & Vizcaíno, J.A. Review of issues and solutions to data analysis reproducibility and data quality in clinical proteomics. In: Mass Spectrometry Data Analysis in Proteomics, Matthiesen, R. (Ed.), 345–371 (Springer New York, New York, NY; 2020).
[113] Van Der Burgt, Y.E.M. & Cobbaert, C.M. Proteoform analysis to fulfill unmet clinical needs and reach global standardization of protein measurands in clinical chemistry proteomics. Clin. Lab. Med. 38, 487–497 (2018).

[114] Lindemann, C., Thomanek, N., Hundt, F., Lerari, T., Meyer, H.E., Wolters, D. & Marcus, K. Strategies in relative and absolute quantitative mass spectrometry based proteomics. Biol. Chem. 398, 687–699 (2017).
[115] Zhang, A., Sun, H. & Wang, X. Serum metabolomics as a novel diagnostic approach for disease: a systematic review. Anal. Bioanal. Chem. 404, 1239–1245 (2012).
[116] Sechi, S.E. Quantitative Proteomics by Mass Spectrometry, Edn. 2nd. p. Xiii, 306 pages (Humana Press: Springer, New York; 2016).
[117] Veenstra, T.D. & Yates, J.R.E. Proteomics for Biological Discovery, Edn. 2nd, 1–335 (Wiley-Blackwell, Hoboken, NJ; 2019).
[118] Brun, V. & Couté, Y.E. Proteomics for Biomarker Discovery: Methods and Protocols. 294 (Springer Science+Business Media, New York, NY; 2019).
[119] Zhang, L. & Vertes, A. Single-cell mass spectrometry approaches to explore cellular heterogeneity. Angew. Chem. Int. Ed. 57, 4466–4477 (2018).
[120] Couvillion, S.P., Zhu, Y., Nagy, G., Adkins, J.N., Ansong, C., Renslow, R.S., Piehowski, P.D., Ibrahim, Y.M., Kelly, R.T. & Metz, T.O. New mass spectrometry technologies contributing towards comprehensive and high throughput omics analyses of single cells. Analyst 144, 794–807 (2019).
[121] Dai, L. & Gao, G.F. Viral targets for vaccines against COVID-19. Nat. Rev. Immunol. 21, 73–82 (2021).
[122] Nabel, G.J. Designing tomorrow's vaccines. N. Engl. J. Med. 368, 551–560 (2013).
[123] Jones, L.H. Recent advances in the molecular design of synthetic vaccines. Nat. Chem. 7, 952–960 (2015).
[124] Knuf, M., Kowalzik, F. & Kieninger, D. Comparative effects of carrier proteins on vaccine-induced immune response. Vaccine 29, 4881–4890 (2011).
[125] Pryde, D.C., Jones, L.H., Gervais, D.P., Stead, D.R., Blakemore, D.C., Selby, M.D., Brown, A.D., Coe, J.W., Badland, M., Beal, D.M., Glen, R., Wharton, Y., Miller, G.J., White, P., Zhang, N., Benoit, M., Robertson, K., Merson, J.R., Davis, H.L. & McCluskie, M.J. Selection of a Novel Anti-Nicotine vaccine: influence of antigen design on antibody function in mice. PloS one 8, e76557 (2013).
[126] Chipinda, I., Hettick, J.M. & Siegel, P.D. Haptenation: chemical reactivity and protein binding. J. Allergy 2011, 839682 (2011).
[127] Sharma, V.K., Sharma, I. & Glick, J. The expanding role of mass spectrometry in the field of vaccine development. Mass Spectrom. Rev. 39, 83–104 (2020).
[128] Phillips, A.J. Developing a predictable regulatory path for nanomedicines by accurate and objective particle measurement. In: Nanomedicines: Design, Delivery and Detection, Braddock, M. (Ed.) 253–280 (Royal Soc Chemistry, Cambridge; 2016).
[129] Brown, B.A., Zeng, X., Todd, A.R., Barnes, L.F., Winstone, J.M.A., Trinidad, J.C., Novotny, M.V., Jarrold, M.F. & Clemmer, D.E. Charge detection mass spectrometry measurements of exosomes and other extracellular particles enriched from Bovine milk. Anal. Chem. 92, 3285–3292 (2020).
[130] Barre, F.P.Y., Heeren, R.M.A. & Potocnik, N.O. Mass spectrometry imaging in nanomedicine: unraveling the potential of MSI for the detection of nanoparticles in neuroscience. Curr. Pharm. Des. 23, 1974–1984 (2017).
[131] De Maar, J.S., Sofias, A.M., Porta Siegel, T., Vreeken, R.J., Moonen, C., Bos, C. & Deckers, R. Spatial heterogeneity of nanomedicine investigated by multiscale imaging of the drug, the nanoparticle and the tumour environment. Theranostics 10, 1884–1909 (2020).

Chapter 9
What is next?

This chapter provides a brief overview of emerging applications of mass spectrometry (MS) in the biopharma and biotechnology sectors, focusing on the growing visibility of this technique in process analytical technology (PAT). Also discussed are longer term prospects of MS in the biopharmaceutical field using as an example the unique challenges put forward by the emergence of personalized medicine.

9.1 The emerging role of mass spectrometry in process analytical technology

The PAT initiative aims at encouraging the drug manufacturers to incorporate modern analytical tools into both production and quality control [1]. FDA considers PAT as "a system for designing, analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality."[1] The FDA's Guidance for Industry on PAT[2] further clarifies that the goal of this initiative is to "enhance understanding and control the manufacturing process, which is consistent with our current drug quality system: quality cannot be tested into products; it should be built-in or should be by design." PAT is a holistic concept, with the term "analytical" being viewed broadly and including chemical, physical, microbiological, mathematical, and risk analyses integrated together. In practice this concept is frequently reduced to specific aspects of the real-time process monitoring (e.g., the use of sensors to monitor pH, dissolved O_2, production of CO_2, and key metabolites) and data processing using the tools of multivariate modeling. However, more complex analyses, such as impurity testing or glycan profiling of recombinant proteins, are typically carried out off-line.

While the appropriateness of a particular analytical method for PAT is determined by the length of the analysis time, not all PAT-compatible analyses have to be carried in-line or even online. Indeed, a typical duration of a number of steps in the biopharmaceutical production/formulation process (Figure 9.1) exceeds 10 h, and the appropriate decision time intervals could exceed 1 h. For example, a typical mammalian cell culture production time frame is 240 h, with a typical decision time of 10 h, while these numbers for the formulation stage are 6 and 0.5 h, respectively [2]. The critically important parameter is the decision time/analysis time (t_d/t_a) ratio [2], which actually determines how useful the results of a particular analysis would be vis-à-vis timely

1 https://www.fda.gov/media/71012/download
2 *ibid.*

https://doi.org/10.1515/9783110546187-009

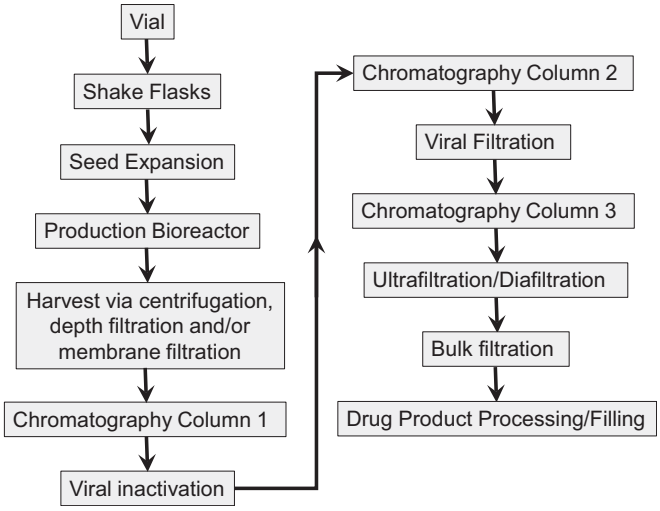


Figure 9.1: A flow diagram for a typical production process of a protein therapeutic (after Rathore et al. [2]).

decision-making. It appears that many MS-based assays would provide favorable t_d/t_a ratios ($\gg$1), even if carried out at-line as opposed to in-line or online formats.

For example, glycosylation profile is one of the quality attributes that is tested during the mammalian cell culture production step, and it is typically accomplished by means of standard oligosaccharide profiling (an hour-long assay resulting in a t_d/t_a ratio of 10). MS analysis of the released glycans is fast and can be carried out in a quantitative format using isotopically labeled derivatizing agents (such as ^{13}C-labeled anthranilic acid), and has been already employed successfully to achieve better process control for production of recombinant proteins in mammalian cell culture (Figure 9.2) [3]. The major bottleneck of this method is the enzymatic deglycosylation step (which is typically accomplished over a period of several hours). Therefore, a streamlined protocol to assess glycosylation profiles based on the intact mass analysis would provide significant time savings and offer better compatibility with PAT requirements. In addition, intact mass measurements of the disulfide-reduced recombinant proteins allow multiple attributes to be monitored, which in the case of mAb products include not only the N-glycosylation profile of the heavy chain, but also identity and glycation status of the light chain, as well as the presence of aglycosylated heavy chains (which are not detected in both standard and MS-based glycan release assays) [4]. The disulfide reduction step prior to the LC-MS intact mass measurements takes only 15 min and has a minimal impact on the t_d/t_a ratio. Further acceleration of the glycan profiling of large disulfide-rich glycoproteins can be achieved using the recently introduced cross-path reactive chromatography (XP-RC)/MS platform [5], where disulfide reduction is carried inside a chromatographic column. This format allows glycosylation profiles to be

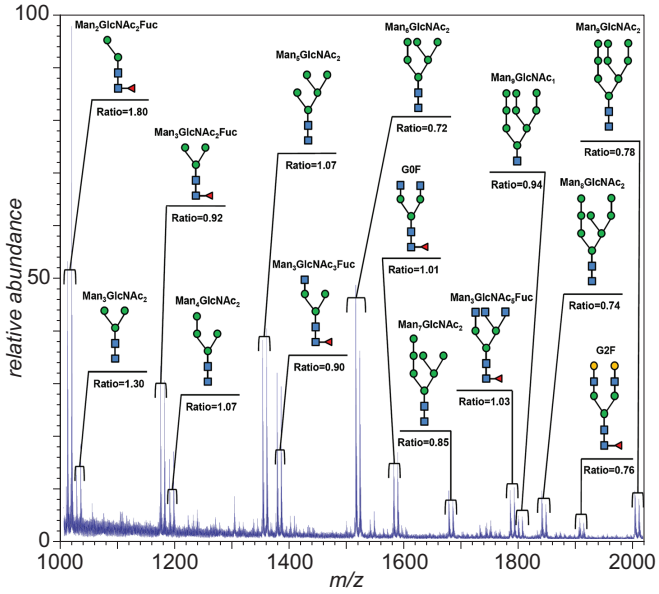


Figure 9.2: A negative ion MALDI-TOF mass spectrum of glycans released from lysates of a cell line at two different time points and mixed at a 1:1 ratio. The two groups of glycans were labeled with $^{12}C_7$ anthranilic acid and $^{13}C_7$ anthranilic acid, respectively, prior to MS analysis. The $^{13}C_7$:$^{12}C_7$ signal ratios are displayed for each of the most abundant glycan types. Adapted with permission from Tep et al. [3].

evaluated within the protein subunits that become metastable following disulfide reduction and would be difficult to analyze using conventional MS-based approaches (Figure 9.3).

Higher order structure of recombinant proteins is another CQA that can be monitored within the PAT format using MS tools [7]. For example, HDX MS-based evaluation of the integrity of the protein conformation discussed in detail in Chapter 4 can be implemented in fast formats that allow the entire measurement cycle to be completed within half-an-hour. This time frame corresponds to a favorable $t_d/t_a = 4$ (a typical decision time for the downstream refolding step is 2 h [2]). Another attribute that is closely related to the higher order structure of cysteine-rich recombinant proteins is the fidelity of their disulfide linkage patterns. While the complete and definitive disulfide mapping is a time-consuming exercise (see Chapter 3 for more detail), there are fast MS-based methods that can be used to assess certain aspects of the disulfide connectivity patterns. For example, the recent introduction of wide-pore UHPLC columns enabled development of UHPLC-MS methods allowing free thiol variants of a recombinant protein (mAb) to be identified. The chromatographic step is completed within five minutes [8], which certainly makes this method compatible with the PAT format. The presence of multiple protein

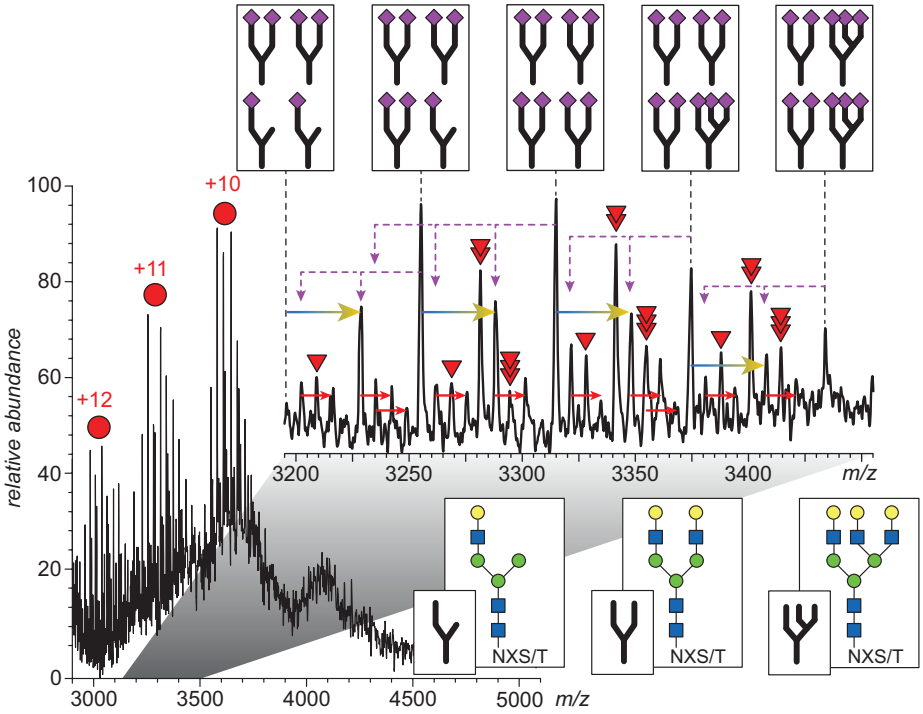


Figure 9.3: The online mass spectrum averaged across a 10.5–11.5 min time window of the XP-RC/MS chromatogram of haptoglobin (elution of the H-chain). The inset shows a zoomed view of the ionic signal at charge state +11. The base peaks in the distribution shown in the inset (identified with the dotted lines) correspond to a-fucosylated glycoforms. The red triangles identify glycoforms that are related to the base peaks by addition of fucose residues (the number of triangles indicates the extent of fucosylation). The dotted purple arrows identify glycoforms related to the base peaks by a loss of sialic acid residues; fucosylation of such species is indicated with red arrows. Adapted with permission from Yang et al. [6].

isomers differing from each other in their disulfide connectivity patterns can be detected by monitoring protein ion charge state distributions in ESI mass spectra and collisional cross sections in IM-MS acquired under denaturing conditions (an approach similar to that described in Chapter 5 in the context of the biosimilarity studies). This approach appears to be sufficiently robust and fast, making it compatible with PAT requirements and enabling its use as a means of monitoring the progress of oxidative refolding of recombinant proteins during the downstream processing [9].

One technique that is poised to become an indispensable tool in PAT armamentarium despite being underutilized at the moment is native SEC-MS [10]. It allows the protein higher order structure integrity to be evaluated based on the appearance of the ionic charge state distributions (see Chapter 4 for a detailed discussion), and

the analysis can be completed within 10–20 min. Utilization of SEC as a front-end separation technique not only solves the ESI solvent compatibility issues, but in fact enables multi-attribute analysis (e.g., detection of size variants and soluble aggregates). Lastly, wider utilization of the MS-based OMICS tools in characterization of both metabolic profiles [11] and proteomes [12] of production cell cultures is likely to be the key for optimization of their performance and achieving superior characteristics of the recombinant proteins.

9.2 Mass spectrometry in personalized medicine

Another emerging and exciting area where MS will undoubtedly play a pivotal role is the design and production of biopharmaceutical products to meet the needs of personalized medicine. MS already lends itself as a powerful tool to facilitate critical advances at the interface of biotechnology and personalized medicine, such as identification of the pericyte secretome components that can be used as templates for the design of neuroregeneration medicines [13]. Personalized peptide vaccines are another area where MS will undoubtedly play a decisive role by allowing the structure of tumor-associated peptides to be determined to enable discovery and development of patient-tailored neoepitopes [14, 15]. Although a detailed discussion of the recent developments in this field goes beyond the scope of this book, this is undoubtedly an area to be watched very closely in the coming years. Lastly, cell-based therapies are another exciting example of personalized medicine, where the patient's own body is used as a reservoir of therapeutic agents. As has been already mentioned in Chapter 8 of this book, MS has a tremendous potential vis-à-vis addressing a range of analytical needs of cell-based therapies, and is likely to enjoy a growing popularity in this field.

In our opinion, nearly all biopharmaceuticals will become an integral part of personalized medicine simply because all medicine will be personalized. We envision a combination of OMICs screens (mostly genomic, but also proteomic and metabolomic) applied to healthy individuals on a regular basis to identify potential risks prior to the disease onset (Figure 9.4). Obviously, this (especially the proteomic and metabolomic screens) cannot be accomplished without using MS; novel approaches to data analysis based on artificial intelligence (AI) are likely to be needed to process these vast data arrays to not only identify credible individual susceptibility risks for specific diseases, but also quantify them. The tools of systems biology would then be used to identify optimal prophylactic agents designed to prevent the disease onset. It might also be possible at this stage to identify theranostic agents to be used as highly specific and reliable reporters of the disease onset. A similar approach could be envisioned for each step of the disease development/progression aiming at achieving the optimal therapeutic outcome with minimal intervention/maximum comfort for the patient (Figure 9.4). Obviously, the specific therapeutic objectives, as well

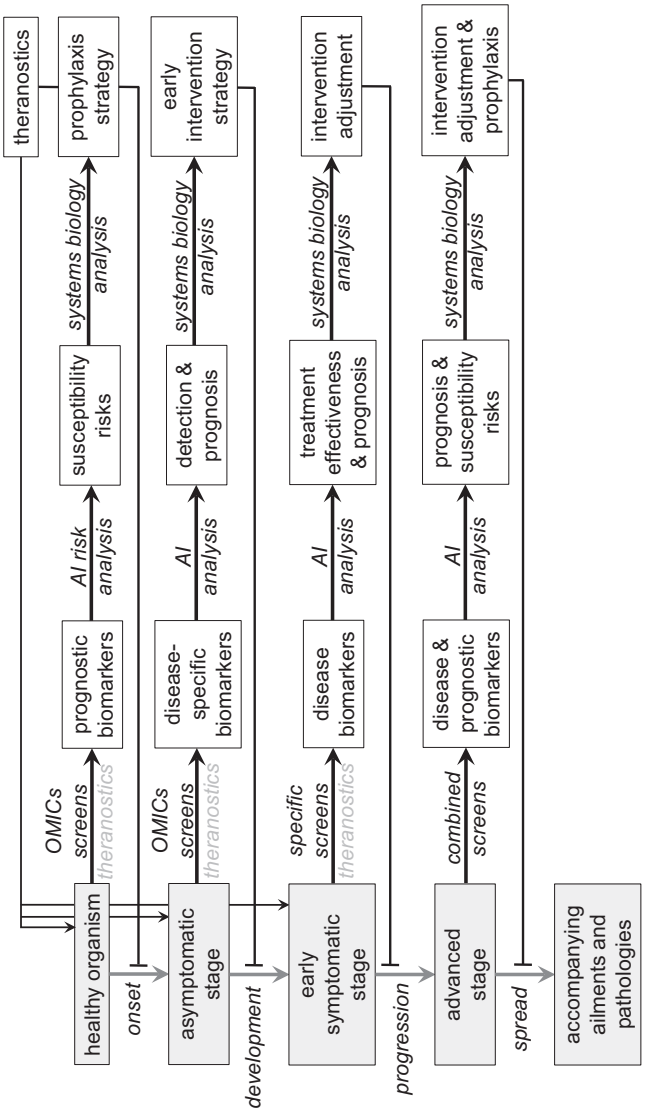


Figure 9.4: A global view flowchart of the information flows in personalized medicine showing analytical tasks specific for every disease stage.

as the biomarker screens applied, would be specific to each stage, but one thing all of them will undoubtedly have in common is the highest level of integration of powerful MS-based methods to obtain comprehensive biomarker datasets and AI tools to process these vast data arrays. Both therapeutics and theranostics are likely to be produced using biotechnological tools, as this route offers the highest flexibility and versatility, can be implemented on a variety of scales to meet the needs of an individual patient, and may be executed within a short time frame. The latter would be particularly important at the advanced disease stages, where the therapeutic intervention window is typically very narrow. MS is bound to become an indispensable part of personalized medicine, acting both as a critical diagnostic tool at the earliest point of the information flow, and as a terminal QC checkpoint allowing rapid yet reliable assessments to be made not only for the quality of the made-on-demand medicines, but indeed for the success of the entire therapeutic strategy.

References

[1] Hinz, D.C. Process analytical technologies in the pharmaceutical industry: the FDA's PAT initiative. Anal. Bioanal. Chem. 384, 1036–1042 (2006).
[2] Rathore, A.S., Bhambure, R. & Ghare, V. Process analytical technology (PAT) for biopharmaceutical products. Anal. Bioanal. Chem. 398, 137–154 (2010).
[3] Tep, S., Hincapie, M. & Hancock, W.S. The characterization and quantitation of glycomic changes in CHO cells during a bioreactor campaign. Biotechnol. Bioeng. 109, 3007–3017 (2012).
[4] Lanter, C., Lev, M., Cao, L. & Loladze, V. Rapid Intact mass based multi-attribute method in support of mAb upstream process development. J. Biotechnol. 314–315, 63–70 (2020).
[5] Pawlowski, J.W., Carrick, I. & Kaltashov, I.A. Integration of on-column chemical reactions in protein characterization by liquid chromatography/mass spectrometry: cross-path reactive chromatography. Anal. Chem. 90, 1348–1355 (2018).
[6] Yang, Y., Pawlowski, J.W., Carrick, I.J. & Kaltashov, I.A. Evaluation of the extent of haptoglobin glycosylation using orthogonal intact-mass MS approaches. ChemRxiv (2021). doi: 10.26434/chemrxiv.13518512.v1.
[7] Guerra, A., Von Stosch, M. & Glassey, J. Toward biotherapeutic product real-time quality monitoring. Crit. Rev. Biotechnol. 39, 289–305 (2019).
[8] Liu, H., Jeong, J., Kao, Y.H. & Zhang, Y.T. Characterization of free thiol variants of an IgG1 by reversed phase ultra high pressure liquid chromatography coupled with mass spectrometry. J. Pharm. Biomed. Anal. 109, 142–149 (2015).
[9] Furuki, K., Toyo'oka, T. & Yamaguchi, H. A novel rapid analysis using mass spectrometry to evaluate downstream refolding of recombinant human insulin-like growth factor-1 (mecasermin). Rapid Commun. Mass Spectrom. 31, 1267–1278 (2017).
[10] Muneeruddin, K., Thomas, J.J., Salinas, P.A. & Kaltashov, I.A. Characterization of small protein aggregates and oligomers using size exclusion chromatography with on-line detection by native electrospray ionization mass spectrometry. Anal. Chem. 86, 10692–10699 (2014).
[11] Chong, W.P. et al. Metabolomics profiling of extracellular metabolites in recombinant Chinese Hamster Ovary fed-batch culture. Rapid Commun. Mass Spectrom. 23, 3763–3771 (2009).

[12] Kang, S. et al. Cell line profiling to improve monoclonal antibody production. Biotechnol. Bioeng. 111, 748–760 (2014).
[13] Gaceb, A., Barbariga, M., Özen, I. & Paul, G. The pericyte secretome: potential impact on regeneration. Biochimie 155, 16–25 (2018).
[14] Supabphol, S., Li, L., Goedegebuure, S.P. & Gillanders, W.E. Neoantigen vaccine platforms in clinical development: understanding the future of personalized immunotherapy. Expert Opin. Investig. 30, 529–541 Drugs in press (2021).
[15] Nelde, A., Rammensee, H.G. & Walz, J.S. The peptide vaccine of the future. Mol. Cell. Proteomics 20, 100022 (2021).

Index

https://doi.org/10.1515/9783110546187-010